



An intermolecular salt bridge linking substrate binding and P1 substrate specificity switch of arterivirus 3C-like proteases



Qian Chen ^{a,b}, Junwei Zhou ^{a,b}, Zhixiang Yang ^{a,b}, Jiahui Guo ^{a,b}, Zimin Liu ^{a,b}, Xinyi Sun ^{a,b}, Qingshi Jiang ^{a,b}, Liurong Fang ^{a,b}, Dang Wang ^{a,b,*}, Shaobo Xiao ^{a,b}

^a State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China

^b Key Laboratory of Preventive Veterinary Medicine in Hubei Province, The Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China

ARTICLE INFO

Article history:

Received 21 February 2022

Received in revised form 26 June 2022

Accepted 27 June 2022

Available online 30 June 2022

Keywords:

3C-like protease

Arterivirus

Salt bridge

Molecular dynamics simulation

Substrate specificity switch

Site-directed mutagenesis

ABSTRACT

Equine arteritis virus (EAV) and porcine reproductive and respiratory syndrome virus (PRRSV) represent two members of the family *Arteriviridae* and pose a major threat to the equine- and swine-breeding industries throughout the world. Previously, we and others demonstrated that PRRSV 3C-like protease (3CL^{PRRSV}) had very high glutamic acid (Glu)-specificity at the P1 position (P1-Glu). Comparably, EAV 3CL^{EAV} exhibited recognition of both Glu and glutamine (Gln) at the P1 position. However, the underlying mechanisms of the P1 substrate specificity shift of arterivirus 3CL^{PRRSV} remain unclear. We systematically screened the specific amino acids in the S1 subsite of arterivirus 3CL^{PRRSV} using a cyclized luciferase-based biosensor and identified Gly116, His133 and Ser136 (using PRRSV 3CL^{PRRSV} numbering) are important for recognition of P1-Glu, whereas Ser136 is nonessential for recognition of P1-Gln. Molecular dynamics simulations and biochemical experiments highlighted that the PRRSV 3CL^{PRRSV} and EAV 3CL^{EAV} formed distinct S1 subsites for the P1 substrate specificity switch. Mechanistically, a specific intermolecular salt bridge between PRRSV 3CL^{PRRSV} and substrate P1-Glu (Lys138/P1-Glu) are invaluable for high Glu-specificity at the P1 position, and the exchange of K138T (salt bridge interruption, from PRRSV to EAV) shifted the specificity of PRRSV 3CL^{PRRSV} toward P1-Gln. In turn, the T139K exchange of EAV 3CL^{EAV} showed a noticeable shift in substrate specificity, such that substrates containing P1-Glu are likely to be recognized more efficiently. These findings identify an evolutionarily accessible mechanism for disrupting or reorganizing salt bridge with only a single mutation of arterivirus 3CL^{PRRSV} to trigger a substrate specificity switch.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The family *Arteriviridae*, which is a member of the order *Nidovirales*, includes six subfamilies *Crocarterivirinae*, *Equarterivirinae*, *Heroarterivirinae*, *Simarterivirinae*, *Variarterivirinae* and *Zealarterivirinae* [1]. Some arteriviruses are known as important veterinary pathogens while others infect certain species of wild rodents or African non-human primates and can only be characterized only by genome sequencing [2,3]. Of these arteriviruses, porcine reproductive and respiratory syndrome virus (PRRSV, belongs to the subfamily *Variarterivirinae*) and equine arteritis virus (EAV, which belongs to the subfamily *Equarterivirinae*) are important etiological agents in the area of veterinary research because of the

economic losses that they cause in the horse- and swine-breeding industries worldwide [4–6]. Arterivirus infections are both cytological and cause outbreaks of either acute or persistent infections in their hosts [7]. Although most EAV infections are sub-clinical, some virulent strains periodically cause significant diseases outbreaks and may be associated with abortions, neonatal mortality, and establishing persistent infection in stallions [8,9]. Pigs of all ages in immunologically naïve herds are vulnerable to being infected with PRRSV [10]. Highly virulent PRRSV strains can cause severe clinical diseases in susceptible herds, which are economically important pathogens of pigs [11–13]. Currently, available drugs and vaccines do not provide adequate protection against these diseases [10,12,14], thus it is important to find an appropriate strategy to prevent arterivirus from proliferating in the host.

Arteriviruses represent a family of positive-sense, single-stranded, enveloped RNA viruses whose viral RNA genomes range in length from about 12 to 16 kb [15,16]. Arteriviruses represent a

* Corresponding author at: State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China.

E-mail address: wangdang@mail.hzau.edu.cn (D. Wang).

family of positive-sense, single-stranded, enveloped RNA viruses whose viral RNA genomes range in length from about 12 to 16 kb [7,17]. Three-fourths of the 5'-proximal length of the genome is taken up by two large open reading frames (ORFs), ORF1a and ORF1b, that encode two long nonstructural replicase polyproteins, pp1a and pp1ab [18,19]. Both polyproteins are autoproteolytically processed by various papain-like cysteine proteases and a 3C-like serine protease (3CL^{pro}) at conserved sites to generate 13–14 mature nonstructural proteins (nsps) for gene transcription and replication [20–24].

Arterivirus 3CL^{pro} plays an indispensable role in the life cycle of arterivirus that cleaves the polyproteins at nine conserved sites [20,21,25]. Considering its essential function in viral replication, arterivirus 3CL^{pro} is a highly prospective target for anti-arterivirus drug design [26–29]. In addition to viral polyprotein processing, we and others demonstrated that arterivirus 3CL^{pro}, specifically EAV 3CL^{pro} and PRRSV 3CL^{pro}, cleave host NF- κ B essential modulator (NEMO) at multiple sites, thereby impairing NF- κ B activation and IFN- β production. Importantly, cleavage of NEMO by PRRSV 3CL^{pro} or EAV 3CL^{pro} led to suppression of the type I IFN signaling pathway [30,31]. Therefore, targeting 3CL^{pro} would hinder viral replication and could also boost host immune response [32].

Many 3CL^{pro} inhibitors for COVID-19 treatment have been proposed recently and some new drug candidates have been successful in preclinical studies [33]. Understanding the substrate recognition mechanism of 3CL^{pro} is a key step in establishing substrate-specific inhibitors of 3CL^{pro}. Meanwhile, substrate envelope-guided design has achieved better potency and resistance profile of 3CL^{pro} inhibitors. The 3CL^{pro} families of arteriviruses are highly variable among themselves and are reasonably conserved only in the active site region. Like most chymotrypsin-like serine proteases, PRRSV 3CL^{pro} employs a canonical Ser-His-Asp triad responsible for recognizing cleavage sites with a glutamic acid (Glu, E) at the P1 position to target viral polyprotein [34,35]. Nevertheless, EAV 3CL^{pro} uses the same catalytic triad but recognizes both Glu and glutamine (Gln, Q) at the P1 position [20,36]. Similar substrate preferences were obtained when arterivirus 3CL^{pro} cleaved host protein, NEMO. Our previous study demonstrated that both EAV 3CL^{pro} and PRRSV 3CL^{pro} induced NEMO cleavage at the conserved residues E166, E171, and E349 at the P1 position, whereas only EAV 3CL^{pro} could target NEMO at Q205 [30]. The substrate preferences of PRRSV 3CL^{pro} and EAV 3CL^{pro} cleaved host proteins might be consistent with the 3CL^{pro} preferences for viral substrates. However, why the cleavage of NEMO at site Q205 is unique to EAV 3CL^{pro} has not been investigated in previous studies. In this study, we characterized the key amino acids in the S1 subsite of arterivirus 3CL^{pro} for P1 substrate specificity by molecular dynamics (MD) simulations and site-directed mutagenesis. The major differences between the S1 subsites manifest at PRRSV 3CL^{pro} Lys138 residue, where EAV 3CL^{pro} contains Thr139 residue. The K138T mutant of PRRSV 3CL^{pro} converts the P1-Glu Substrate Specificity of this protease to that toward P1-Gln seen in the EAV 3CL^{pro}. Moreover, Thr139 in EAV 3CL^{pro} is responsible for the preference for P1-Gln, since the T139K mutant more efficiently recognizes substrates containing P1-Glu than the wild type (WT) EAV 3CL^{pro}. Hence, we identified a potential intermolecular salt bridge linking substrate binding and P1 substrate specificity shift of arterivirus 3CL^{pro}, which might be important when designing specific inhibitors of arterivirus 3CL^{pro}.

2. Materials and methods

2.1. Sequence analysis of the PRRSV and EAV genomes

The complete PRRSV and EAV genomes were selected from the National Center for Biotechnology Information (NCBI, [https://](https://www.ncbi.nlm.nih.gov)

www.ncbi.nlm.nih.gov) for substrate profiling of their 3CL^{pro}. Briefly, we obtained 1,037 complete PRRSV genome and 96 complete EAV genome (obtained 1 November 2021). Nine putative cleavage sites of PRRSV 3CL^{pro} and EAV 3CL^{pro} were inferred from alignments of pp1ab polyprotein amongst PRRSV and EAV isolates. Details on the data set are summarized in [Supporting information Fig. S1](#).

2.2. Plasmids

The eukaryotic expression plasmids for PRRSV 3CL^{pro}, EAV 3CL^{pro}, and NEMO (E166A-E171A-E349A) were described previously [30]. PRRSV 3CL^{pro} and EAV 3CL^{pro} mutants were constructed by overlap extension PCR using specific mutagenic primers and cloned into the C-terminal hemagglutinin (HA) tag-encoding pCAGGS-HA-C plasmid. The luciferase-based biosensor plasmid 233D, which contained oligonucleotides corresponding to ENLYFQ↓YS (cleaved by TEV 3C^{pro}), were used as the biosensor control as described previously [37]. To monitor the activity of PRRSV 3CL^{pro} and EAV 3CL^{pro}, oligonucleotides corresponding to the amino acid sequences of substrate peptides were cloned into the 233D biosensor vector, including PRRSV-nsP3/4, EAV-nsP3/4, EAV-nsP10/11, PRRSV-nsP3/4-Q, and NEMO-Q205. All the constructs were to be validated by DNA sequencing.

2.3. Dual-luciferase assays

Human embryonic kidney (HEK293T) cells were obtained from the China Typical Culture Collection Center (Wuhan, China) and cultured at 37 °C, 5% CO₂ in Dulbecco's modified Eagle medium (DMEM, Invitrogen, Madison, WI, USA), with 10% fetal bovine serum. HEK293T cells in 24-well plates were transfected with indicated luciferase biosensor plasmids and pRL-TK (Promega), along with PRRSV 3CL^{pro} or EAV 3CL^{pro} expression plasmids or the empty control plasmid. The pRL-TK plasmid was used as a control for transfection efficiency. Cells were lysed 28 h post-transfection and firefly luciferase and Renilla luciferase activities were determined using the dual-luciferase reporter assays (Promega, USA). The data represent relative firefly luciferase activity normalized to Renilla luciferase activity and are representative of three independently conducted experiments. All results were presented as means \pm standard deviations. The *p* value of < 0.05 was regarded as significant. All results were analyzed using GraphPad Prism Version 9.3 (www.graphpad.com).

2.4. Western blot

HEK293T cells grown in 6-well plates were transfected with indicated plasmids. After 28 h transfection, cell lysates were prepared in RIPA buffer (Beyotime, China), quantified for protein concentrations, and subject to SDS-PAGE and immunoblotting as previously described [30]. The following antibodies were used to identify the corresponding proteins: anti- β -actin antibody (Antgene, Wuhan, China), anti-Luciferase antibody (Promega), anti-HA antibody (MBL, Nagoya, Japan), and anti-Flag antibody (Macgene, China). Expression of β -actin was detected with an anti- β -actin antibody to confirm loading of equal protein amounts. To confirm the expression levels of HA-tagged WT arterivirus 3CL^{pro} and its mutants, an anti-HA antibody was used for immunoblotting. The expression of luciferase or Flag-tagged NEMO proteins was analyzed using anti-Luciferase antibody or an anti-Flag antibody, respectively. Following incubation with appropriate secondary antibodies, protein bands were imaged with a ChemiDoc Imaging System (Bio-Rad, USA).

2.5. Molecular dynamics simulations

Molecular dynamics simulations were performed as described previously [37]. Briefly, the PRRSV 3CL^{pro} (PDB id: 5Y4L) and EAV 3CL^{pro} (PDB id: 1MBM) complexed with different substrate peptides (nsp3/nsp4 auto-cleavage sequence of PRRSV 3CL^{pro} LGSLL↓GAFRTQ, nsp3/nsp4 auto-cleavage sequence of EAV 3CL^{pro} GGMVFE↓GLFRSP, and nsp10/nsp11 cleavage sequence of EAV 3CL^{pro} CGWEKQ↓SNKISC) were generated utilizing the structure of Glu-SGP in complex with a tetrapeptide (PDB id: 1HPG) as the reference template, because EAV 3CL^{pro} and PRRSV 3CL^{pro} exhibited a high similarity with Glu-SGP structure among all available crystal complex structure. For consistency, these crystal structures were also adopted as the template for the construction of the single mutant complexes using SYBYL-X v.2.0 (v.2.0; <https://omictools.com/sybyl-x-tool>). MD simulations were conducted using the Amber ff14SB force field (with the TIP3P water model) implemented in the GROMACS 2018 software package as formerly described [38–41]. The distance between the edge of the water box and protein was at least 1.2 nm. Molecular systems were neutralized by the addition of a certain number of counterions sodium (Na⁺) or chloride (Cl⁻), and NaCl at 150 mM was used to mimic physiological conditions [42]. For energy minimization and relaxation of the systems, each system was energy minimized by using a 2-step, extensive energy minimization process based on the steepest descent method followed by the conjugate gradient algorithm. Following minimization, each system was gradually heated from 0 to 310 K in 500 ps, then equilibrated at that temperature for another 500 ps. The standard temperature is kept constant at 310 K, then equilibration was performed under constant pressure for 150 ns with no position restrictions on protein. During the MD simulation, the SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms [43]. The temperature was controlled using the modified Berendsen thermostat (V-rescale algorithm) with a collision frequency of 0.2 ps [44]. The pressure was maintained at 1 bar using the Parrinello-Rahman barostat with a compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$ and a coupling constant of 2.0 ps. The non-bonded cutoff distances and van der Waals (vdW) interaction were 15 Å, and the Particle Mesh Ewald (PME) method was utilized to treat long-range electrostatic interactions [45]. A 2 fs step was applied to calculate the motion equations using the Leap-Frog integrator for equilibration steps [43]. The coordinates of the atoms were taken at every 10 ps and used for the final analysis. The backbone root mean square deviations (RMSD) values, root mean square fluctuation (RMSF) values, vdW interactions and electrostatic interactions were calculated using the GROMACS package. In addition, the RMSD values suggest that the different conformations remain stable during 150 ns simulation times (Fig. S2). The 3D-structure of proteins and peptides was visualized using PyMOL software Version 2.5 (Schrodinger, LLC) [46].

2.6. Hydrogen bond calculations

Hydrogen bonds (H-Bonds) were identified based on donor–acceptor distance and donor–acceptor angle (X...Y distance < 3.5 Å and X-H...Y angle < 40°), whereas only the pairs of polar atoms belonging to different hydrophobic groups were involved. The H-Bonds Plugin from Visual Molecular Dynamics was used to perform hydrogen bond analyses for each complex. H-bond occupancy indicated the percentage of indicated H-bonds that are maintained throughout the MD simulation. The data represented all H-bonds with occupancy greater than 10% between P1 residue of the substrate peptides and S1 subsite of WT arterivirus 3CL^{pro}.

2.7. Clustering of MD trajectories using principal component analysis

Principal component analysis (PCA) was used to determine the conformational changes in protein movement by calculation and diagonalization of the covariance matrix of the C-alpha (C α) atoms. The calculated orthogonal vectors or eigenvectors with the highest eigenvalues are named principal components (PCs). PCA was performed using the Bio3d package (v2.4.1; <http://thegrantlab.org/bio3d/>) in R environment (v4.0.4; <https://mirror.fcaglp.unlp.edu.ar/CRAN/>) [47,48]. Trajectories are obtained from dynamic simulation. The 100 ns MD trajectory consists of 10,000 frames after equilibration. One conformation out of ten is sampled. By sampling each conformation, a single long trajectory is obtained, and thus it represents 10% of the 10,000 frames MD trajectory. To determine the internal motions of a protein, alignment is necessary before building the C-matrix using Cartesian coordinates. C α atoms were selected for analysis and in each frame were aligned by removing the overall translation and rotation, and then corresponding PCs were calculated under default parameters using the Bio3d package. The predominant movements during the simulation were plotted using these PCs.

3. Results

3.1. Different substrates preference at the P1 position between PRRSV 3CL^{pro} and EAV 3CL^{pro}

To clarify the substrate specificity of arterivirus 3CL^{pro}, we obtained 1,364 complete genome sequences of 25 different species of arteriviruses from the National Center for Biotechnology Information (NCBI). PRRSV and EAV, as two well-characterized arteriviruses, contributed most of the complete genome sequences of arteriviruses (1037 and 96, respectively). Thus, we selected PRRSV and EAV as representatives to investigate the substrates preference of arterivirus 3CL^{pro}. Based on previous studies [34,35], nine putative cleavage sites of PRRSV 3CL^{pro} and EAV 3CL^{pro} were predicted based on alignments of pp1ab polyproteins amongst all PRRSV and EAV isolates (Fig. 1a-b). The sequence logos of PRRSV 3CL^{pro} and EAV 3CL^{pro} substrates show considerable similarities, but there is a subtle difference at the P1 position. Consistent with previous studies, PRRSV 3CL^{pro} confers an extremely high specificity for the P1 residue of Glu in the substrates. Interestingly, while Glu is frequently found specific to the P1 position in EAV 3CL^{pro}, Gln at P1 appears in one of 9 cleavage sites (nsp10/nsp11) [21]. This cleavage site (nsp10/nsp11) with the preference of Gln at the P1 position is well conserved amongst all EAV isolates (Fig. S1). In addition, the P1' position, on the other hand, is mainly occupied by various small-sized residues such as Gly, Ser and Ala, among which Gly is the most frequent one [21,22,34]. The residue preference at the remaining Pn's residues of PRRSV 3CL^{pro} and EAV 3CL^{pro} is less demanding. These results revealed that the residue specificity of the P1 position varies in the substrates for PRRSV 3CL^{pro} and EAV 3CL^{pro}.

3.2. Structure comparison of S1 subsites in arterivirus 3CL^{pro} using molecular dynamics simulation

Viral 3CL^{pro} is broadly believed to undertake conformational changes after substrate binding. These changes allow them to interact with substrate proteins and to initiate catalytic activity [45]. To further determine whether structural changes in arterivirus 3CL^{pro} in response to substrate binding and hence potentially regulate arterivirus 3CL^{pro} activity, we performed molecular dynamics analysis of apo and substrate-bound arterivirus 3CL^{pro}.

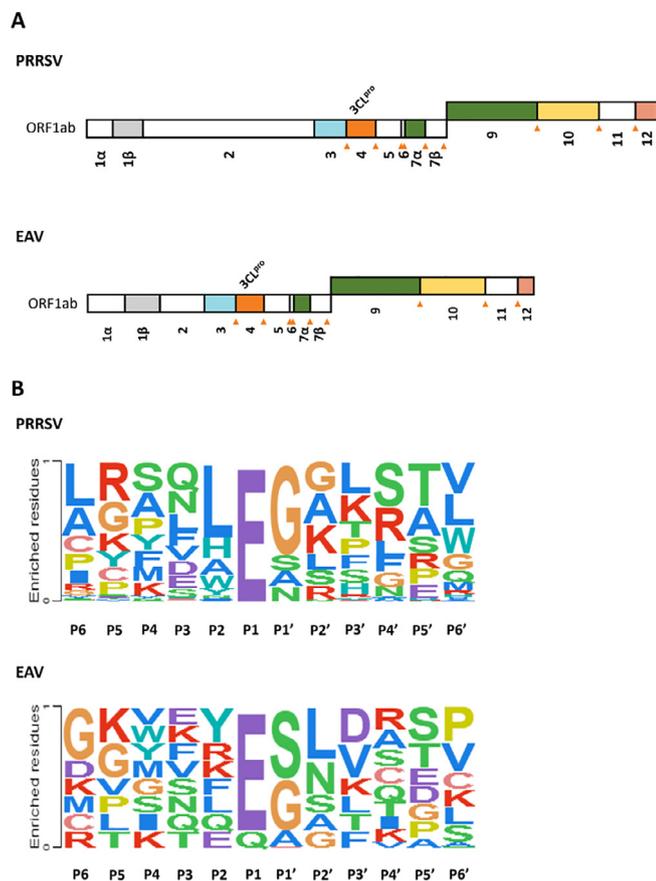


Fig. 1. Sequence logos of the polyprotein junctions cleaved by PRRSV 3CL^{pro} and EAV 3CL^{pro} from different isolates. (a) Schematic diagrams of the PRRSV and EAV polyproteins. The PRRSV 3CL^{pro} and EAV 3CL^{pro} encoded by nsp4 are marked by orange rectangular; all the 3CL^{pro}-mediated cleavage sites in polyproteins are shown in the orange triangle. (b) Conservation of the polyprotein junctions cleaved by PRRSV 3CL^{pro} from 1037 PRRSV strains and EAV 3CL^{pro} from 96 EAV strains. Amino acid sequence logos of the substrates were created using PSSMsearch (<http://slim.icr.ac.uk/pssmsearch/>), and the height of the letters represented the relative frequency of the amino acid. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As shown in Fig. 2a, the 3D structure of PRRSV 3CL^{pro} is remarkably like that of EAV 3CL^{pro}, with the corresponding RMSD of alignment of 1.653 Å, even though PRRSV 3CL^{pro} has low amino acid sequence homology with EAV 3CL^{pro} (about 37%). Both PRRSV 3CL^{pro} and EAV 3CL^{pro} consists of a chymotrypsin-like domains fold (domains I and II) at the N-terminus, an extra C-terminal domain III, and a lengthy loop linking domains II and III [23,34]. A shallow-surfaced gap between domains I and II, to which 6 and 7 antiparallel β sheets from two domains contribute, forms the native substrate-binding site including the catalytic Ser-His-Asp triad and the Gly-Ser oxyanion hole [23]. Surprisingly, despite their high structural similarity, the S1 subsites of PRRSV 3CL^{pro} and EAV 3CL^{pro} showed significantly different amino acid residue compositions. In detail, the S1 subsite in the PRRSV 3CL^{pro} is mainly composed of Thr113/Cys115/Gly116/His133/Ser136/Lys138, while that in EAV 3CL^{pro} is constituted by Thr115/Thr116/Ser117/Gly118/His134/Ser137 (Fig. 2b). These residues are highly conserved in PRRSV 3CL^{pro} and EAV 3CL^{pro}, respectively, suggesting these residues might involve the P1 substrate specificity of arterivirus 3CL^{pro} (Fig. 2c). As expected, the P1 residue, Glu or Gln, engaged in a highly similar interaction network with the S1 sites of the 3CL^{pro} active site in PRRSV and EAV, including a fully conserved network of H-bonds between the P1-Glu/Gln and Thr113, Gly116 and His133 of 3CL^{pro} (all amino acid positions are described using PRRSV 3CL^{pro} num-

bering). However, we observed several significant differences in PRRSV 3CL^{pro} and EAV 3CL^{pro} complexes that highlight active site plasticity. In 3CL^{pro}/P1-Glu complexes, the Ser136 side-chain participates in H-bonds with the P1-Glu. In contrast, the Ser137 side-chain points away from the P1-Gln side-chain in EAV 3CL^{pro}/P1-Gln structure. The second region of variability is at 138 positions. The substrate P1-Glu side chain was stabilized via a salt bridge interaction to the PRRSV 3CL^{pro} Lys138 residue. However, in EAV 3CL^{pro} structures, Thr139 is on the outside of the S1 subsite, regardless of the different P1-Glu and P1-Gln conformations (Fig. 2b).

To further examine the flexibility of the S1 subsites in 3CL^{pro}, we concentrated on the movements of loops enveloping the S1 subsites by RMSF analysis. We found that one of the analyzed loops of apo PRRSV 3CL^{pro}, loop S136–G140, lined a wall of the S1 subsite and was more flexible than the corresponding loops of apo EAV 3CL^{pro} structure, while the neighbouring loops had little flexibility (Fig. 2d). This might be deduced indirectly from the lack of the S136–G140 loop in the PRRSV 3CL^{pro} crystallographic structure in our and others' previous studies [49]. Surprisingly, such flexibility reduced substrate-bound PRRSV 3CL^{pro}, suggesting that the presence of a substrate could stabilize the loops around the active site (Fig. 2d). B-factors analysis of all deposited arterivirus 3CL^{pro} crystal structures completely verified these conclusions. It is worth adding that the residues on this loop differ greatly in 138 and 139 positions between PRRSV 3CL^{pro} and EAV 3CL^{pro} (Fig. S3), suggesting that these two amino acids in this loop are likely to be involved in the P1 substrate specificity switch of arterivirus 3CL^{pro}.

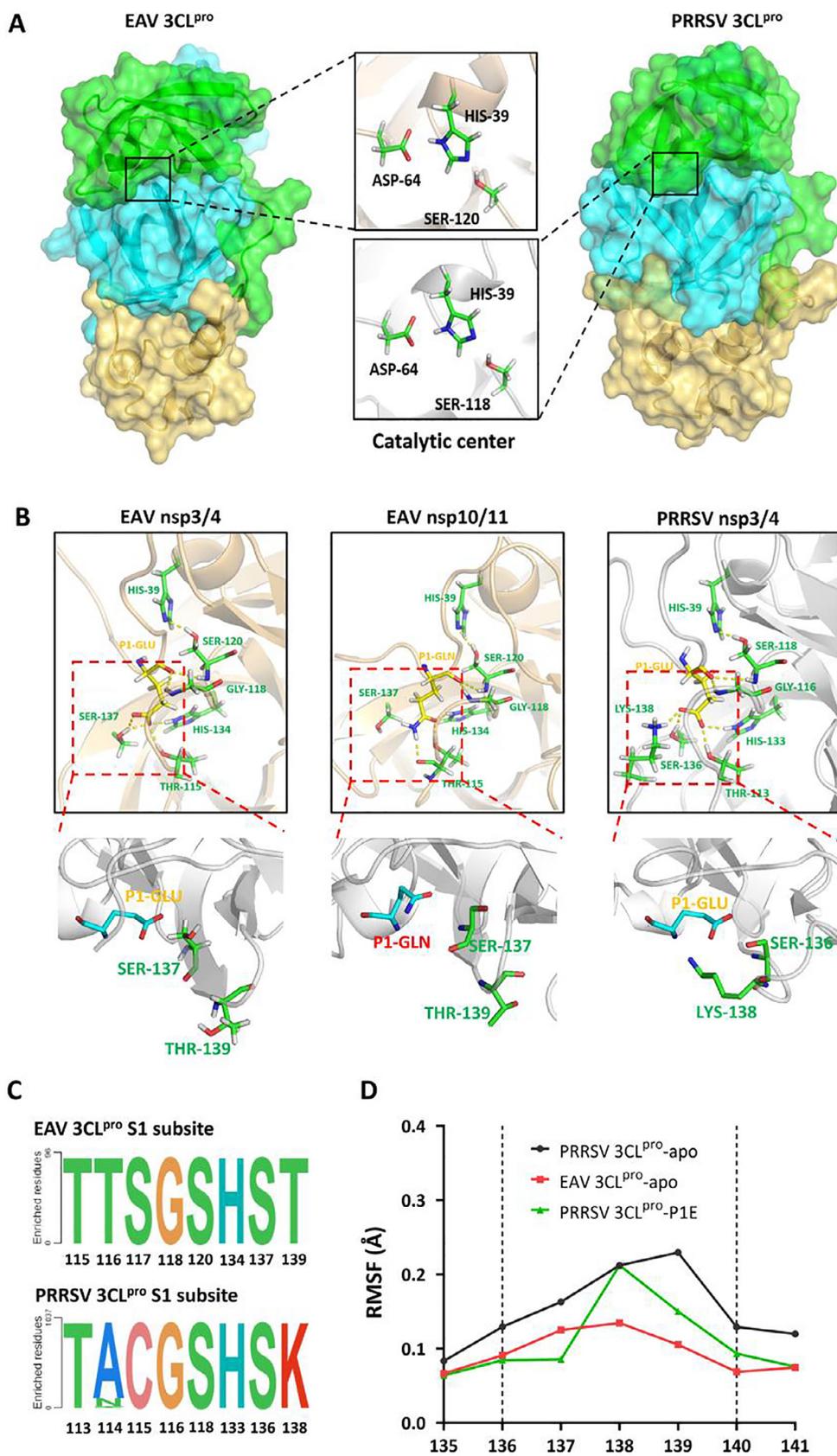
To understand the essential motion changes of apo and substrate-bound conformations of arterivirus 3CL^{pro}, PCA was performed on MD trajectories. For the apo and substrate-bound forms of PRRSV 3CL^{pro}, the first two eigenvectors (EVs) contributed 33.66% and 56.6% of the global motion, respectively (Fig. S4). These results showed that the eigenvalues of the apo conformation were lower than those of the substrate-bound form. Furthermore, the scatter plots generated by the apo and substrate-bound forms of PRRSV 3CL^{pro} indicated a significant difference between the two systems, which also suggested that the apo form of PRRSV 3CL^{pro} showed less dynamic tertiary structure conformation than the substrate-bound form. Similar motion changes were obtained from MD trajectories of the apo and substrate-bound forms of EAV 3CL^{pro}, suggesting that the structural rearrangement by substrate binding might contribute to the increased essential motion in the substrate-bound form of arterivirus 3CL^{pro}.

3.3. Establishment of a biosensor assay for evaluating the protease – substrate specificity of arterivirus 3CL^{pro} in vivo

The traditional approach to studying protease-substrate specificity is usually performed through *in vitro* proteolytic reactions with a purified protease and its predicted peptide or protein substrates in a fluid buffer. *In vitro* characterization is always labour- and time-demanding, and it might not accurately reflect the physiological characteristics of proteases. Here, we have established a fast, reliable, and efficient luciferase-based biosensor to quantify the substrate specificity of arterivirus 3CL^{pro} *in vivo*, which has been used in previous studies to identify cleavage sites recognized by viral protease [37,50–52]. In the biosensor assay, the P6–P6' positions of the candidate recognition regions of arterivirus 3CL^{pro} were inserted into the cycling luciferase biosensor (Fig. 3a). The PRRSV 3CL^{pro} N-terminal auto-cleavage sequence (named PRRSV-nsp3/4, with P1-Glu preference), EAV 3CL^{pro} N-terminal auto-cleavage sequence (named EAV-nsp3/4, with P1-Glu preference) and EAV nsp10/nsp11 cleavage sequence (named EAV-nsp10/11, with P1-Gln preference) were tested, while the Tobacco Etch Virus (TEV) protease recognition sequence ENLYFQYS (named 233D) was

used as a negative control. In principle, arterivirus 3CL^{pro} effectively cleaves its favored substrates between residues P1 and P1', leading to an interaction of the two firefly luciferase domains,

resulting in an active form of the luciferase. Conversely, the unpreferred substrates will not be efficiently cleaved by arterivirus 3CL^{pro}, resulting in a non-active form of the luciferase. As shown



in Fig. 3b, the luciferase activity of PRRSV-nsp3/4 was markedly induced in cells cotransfected with PRRSV 3CL^{pro}, while no activity was detected in negative controls (233D). Moreover, the luciferase activity of both EAV-nsp3/4 and EAV-nsp10/11 was also induced in EAV 3CL^{pro}-cotransfected cells. Interestingly, no activity of EAV-nsp10/11 was detected in PRRSV 3CL^{pro}-cotransfected cells, suggesting that the substrate with P1-Gln was a nonpreferred substrate for PRRSV 3CL^{pro}, which is consistent with previous studies on the P1 substrate specificity of EAV 3CL^{pro} and PRRSV 3CL^{pro} [21]. Western blotting also showed that PRRSV 3CL^{pro}/EAV 3CL^{pro} could split the recombinant firefly luciferase with the preferred cleavage sequence, resulting in a more rapidly migrating protein band (Fig. 3c). Consistency of cleavage and fold induction of the luciferase supports that there is a correspondence between the luciferase activity assay and cleavage of the biosensor construct by arterivirus 3CL^{pro}, pointing to its possible use in evaluating the protease-substrate specificity of arterivirus 3CL^{pro} *in vivo*.

3.4. Molecular determinants of arterivirus 3CL^{pro} substrate specificity

Arterivirus 3CL^{pro} employs a conserved Ser-His-Asp triad as the principal nucleophile and the general acid-base catalyst [23,34,53], which is important for substrate binding and hydrolysis. We considered the possibility that the additional amino acid residues of S1 subsites are also involved in arterivirus 3CL^{pro} activity and P1 substrate preference. To this end, several amino acids in the S1 subsites of PRRSV 3CL^{pro} and EAV 3CL^{pro} were mutated to alanine and assessed the P1 substrate preference of all mutants using three cyclized luciferase-based biosensors (contained Glu or Gln at the P1 position to enable assessment of substrate preference at this position). As shown in Fig. 4a, the putative residues in S1 subsite of PRRSV 3CL^{pro}, Thr113, Cys115, G116, His133, Ser136, and Lys138, were mutated to Ala and tested with a PRRSV-nsp3/4 biosensor (with P1-Glu). In contrast to WT PRRSV 3CL^{pro}, the mutants (T113A, C115A, H133A, K138A) significantly less efficiently cleaved substrate containing Glu at P1, but the mutants (G116A, S136A) almost completely abolished the ability of PRRSV 3CL^{pro} to cleave substrates having P1-Glu. Using an EAV-nsp3/4 biosensor (with P1-Glu), we found that the mutants of EAV 3CL^{pro} retained (T116A, S117A), significantly reduced (T115A, H134) or eliminated (G118A, S137A) the ability to cleave substrate containing Glu at P1 (Fig. 4b). These data suggest that both G116/H133/S136 of PRRSV 3CL^{pro} and G118/H134/S137 of EAV 3CL^{pro} are important for recognition of P1-Glu. Interestingly, the mutants G118A and H134A of EAV 3CL^{pro} lost P1-Gln cleavage activity, but the mutant S137A retained this ability cleaving P1-Gln substrate as did WT EAV 3CL^{pro} (Fig. 4c), suggesting that although Ser137 is important for its ability to cleave substrates having P1-Glu, this residue is nonessential for recognition of P1-Gln.

To further investigate the underlying molecular mechanism of the alanine scan results, we performed MD simulations used by different PRRSV 3CL^{pro} and EAV 3CL^{pro} mutants with substrates. The RMSD values of the C α atoms for each complex converged and remained stable during 150 ns MD simulations. In the last

100 ns of simulations, we calculated the selected H-bond (occupancy greater than 10%) during MD simulations to better capture the electrostatic interactions between the substrate and the protease. Gly116/Gly118 in PRRSV 3CL^{pro}/EAV 3CL^{pro} mainchain amides constitute the oxyanion hole by providing one H-bond to the cleavable P1-Glu and P1-Gln carbonyl oxygen that stabilizes the resulting negative charge during covalent catalysis. As expected, both G116A mutant in PRRSV 3CL^{pro} and G118A mutant in EAV 3CL^{pro} disrupted the engagement of H-bond network around the active sites (Fig. 4d-f), which also eliminated the ability to cleave the substrate (Fig. 4a-c). Moreover, His133/His134 and Thr113/Thr115 in PRRSV 3CL^{pro}/EAV 3CL^{pro} make H-bonds with the invariant side chain of P1-Glu and P1-Gln. Indeed, we found that alanine substitution of each of these two residues significantly reduced the H-bond occupancy between arterivirus 3CL^{pro} and substrates. These conserved H-bond interactions between arterivirus 3CL^{pro} and substrates provide an explanation for the reduced catalytic efficiency of the mutants (H133A/H134A, T113A/T115A in PRRSV 3CL^{pro}/EAV 3CL^{pro}) for both P1-Glu and P1-Gln substrates. By contrast, alanine mutation of other residues that cannot form stable H-bonds with the P1 substrate, such as C115 in PRRSV 3CL^{pro} and T116/S117 in EAV 3CL^{pro}, had less or no significant effect on P1-Glu/P1-Gln cleavage activity. Importantly, Ser136/Ser137 in PRRSV 3CL^{pro}/EAV 3CL^{pro} contacts P1-Glu, but not P1-Gln, through two conserved H-bonds (S136/S137-OG(Side)...P1E-OE1(Side) and S136/S137-OG(Side)...P1E-OE2(Side)). S136A mutant in PRRSV 3CL^{pro} and S137A mutant in EAV 3CL^{pro} caused a loss of H-bonds between side chains of S136/S137 and the P1-Glu (Fig. 4d column 1 vs 6, Fig. 4e column 1 vs 7), whereas S137A mutant in EAV 3CL^{pro} had no significant effect on H-bond interactions toward P1-Gln (Fig. 4f column 1 vs 7). Further confirming this notion, the S137A mutant in EAV 3CL^{pro} lost the ability in cleaving substrate containing P1-Glu but had a similar cleavage activity toward P1-Gln as did WT EAV 3CL^{pro} (Fig. 4b-c). These results suggested that substrate interactions with the Ser136/Ser137 side chains in PRRSV 3CL^{pro}/EAV 3CL^{pro} distinguish P1-Glu and P1-Gln recognition.

Interestingly, in the flexible loop S136-G140 of PRRSV 3CL^{pro}, a salt bridge Lys138-P1-Glu interaction is observed in PRRSV 3CL^{pro}-P1-Glu complex and contributes to binding P1-Glu of the substrate, which might explain why the K138A mutant lost the ability to cleave substrate with P1-Glu. Surprisingly, neither EAV 3CL^{pro}-P1-Glu nor EAV 3CL^{pro}-P1-Gln complexes show the salt bridge observed with Lys138-P1-Glu in the PRRSV 3CL^{pro} substrate complexes. Considering that PRRSV 3CL^{pro} exhibited P1 preference only toward Glu and EAV had both P1-Glu and P1-Gln preference, we speculate that the salt bridge might lead to a P1 substrate specificity shift from P1-Glu/Gln to P1-Glu preference during the evolution of arterivirus 3CL^{pro}.

3.5. Identification of an intermolecular salt bridge linking P1 substrate specificity shift of arterivirus 3CL^{pro}

To further test whether the Lys138-P1-Glu salt bridge links the P1 substrate specificity shift of arterivirus 3CL^{pro}, we attempted to

Fig. 2. Structure comparison of the S1 subsites in PRRSV 3CL^{pro} and EAV 3CL^{pro}. (a) The overall structures of PRRSV 3CL^{pro} (PDB id: 5Y4L) and EAV 3CL^{pro} (PDB id: 1MBM) and details of their catalytic centers. The surface view of PRRSV 3CL^{pro} and EAV 3CL^{pro} comprised three domains (domain I marked in green, domain II marked in blue, and domain III marked in yellow). The canonical Ser-His-Asp triad responsible for recognizing cleavage sites was presented in the middle of the image, which is located in the cleft between domains I and II. (b) The S1 subsites of PRRSV 3CL^{pro} (grey) and EAV 3CL^{pro} (orange) in complex with its substrates are shown as surface and cartoon, respectively. The amino acid residues composited the S1 subsite and the catalytic triad Gly116-His39-Ser118 (PRRSV 3CL^{pro} numbering is used) are displayed as sticks. Enlarged view of residues Ser136 and Lys138 in PRRSV 3CL^{pro} and residues Ser137 and Thr139 in EAV 3CL^{pro}. Based on the structure of arterivirus 3CL^{pro} (PRRSV, PDB id: 5Y4L; EAV, PDB id: 1MBM), the molding structure of the 3CL^{pro}-substrate complex was generated. (c) Sequence alignment for the S1 subsites of the PRRSV 3CL^{pro} and EAV 3CL^{pro} amongst all PRRSV and EAV isolates. Sequence logos were generated using PSSMsearch (<http://slim.icr.ac.uk/pssmsearch/>). (d) RMSF analysis of loop S136-G140 in apo and substrate-bound arterivirus 3CL^{pro}. Per residue RMSF of C α atoms in loop S136-G140 of apo PRRSV 3CL^{pro} (black), apo EAV 3CL^{pro} (red), and substrate-bound PRRSV 3CL^{pro} (green) from MD trajectories. The sequences were aligned, and the PRRSV 3CL^{pro} residue numbering is used for the x-axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

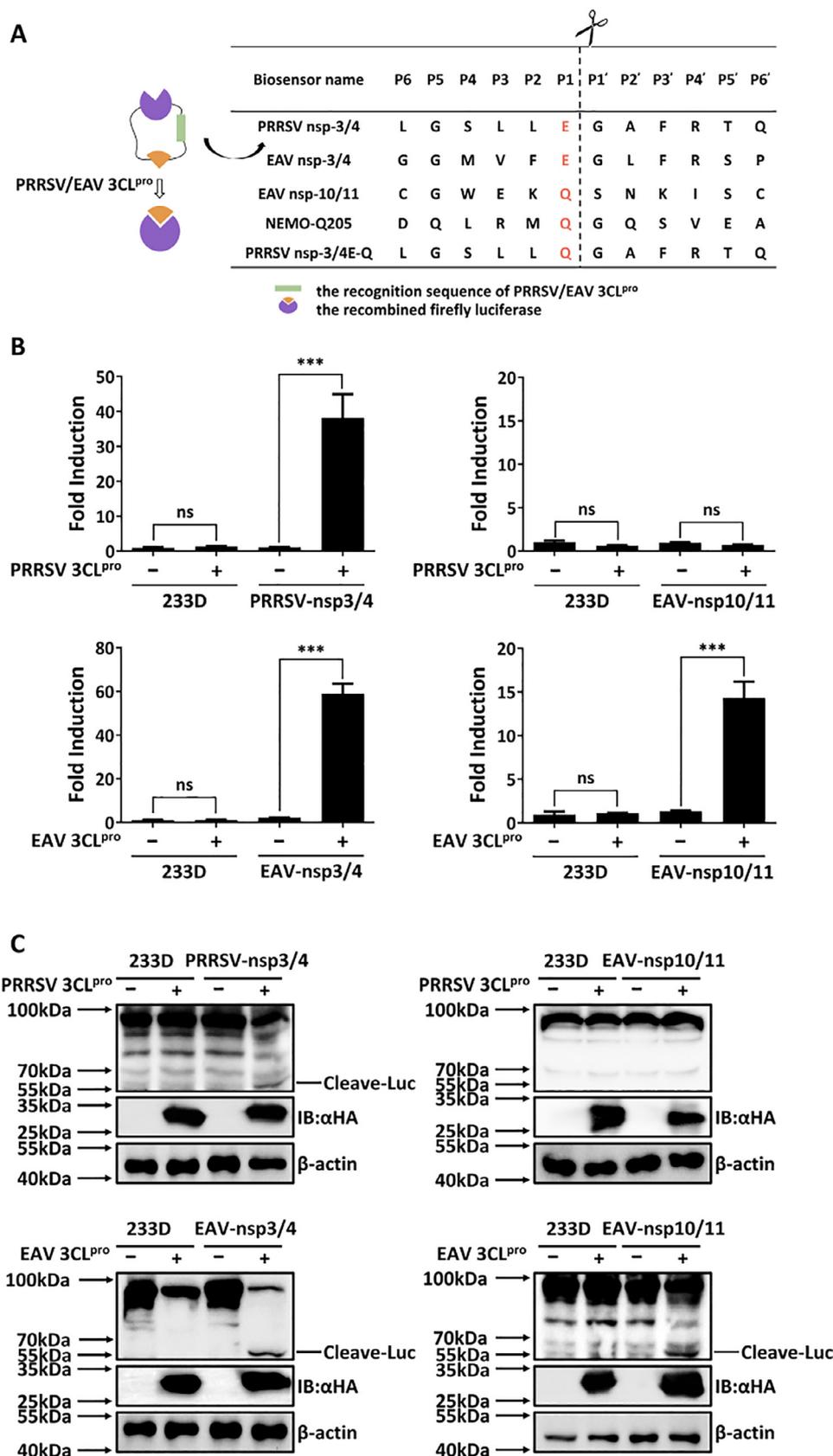


Fig. 3. Exploiting the biosensor assay to evaluate the protease-substrate specificity of arterivirus 3CL^{pro} *in vivo*. (a) Diagram of the generation of luciferase-based biosensors. The purple and the orange structures represent the recombined firefly luciferase. The green rectangle indicates the different recognition sequences that were used to evaluate the activity of PRRSV 3CL^{pro} or EAV 3CL^{pro}. (b and c) HEK293T cells in 24-well plates were transfected with the pRL-TK plasmid, the plasmid encoding PRRSV 3CL^{pro} or EAV 3CL^{pro}, and corresponding luciferase-based biosensors plasmids, or their control 233D. After 28 h transfection, cell lysates were prepared and analyzed by dual-luciferase assays (b) and western blotting (c). ns, not significant. ***p < 0.001. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

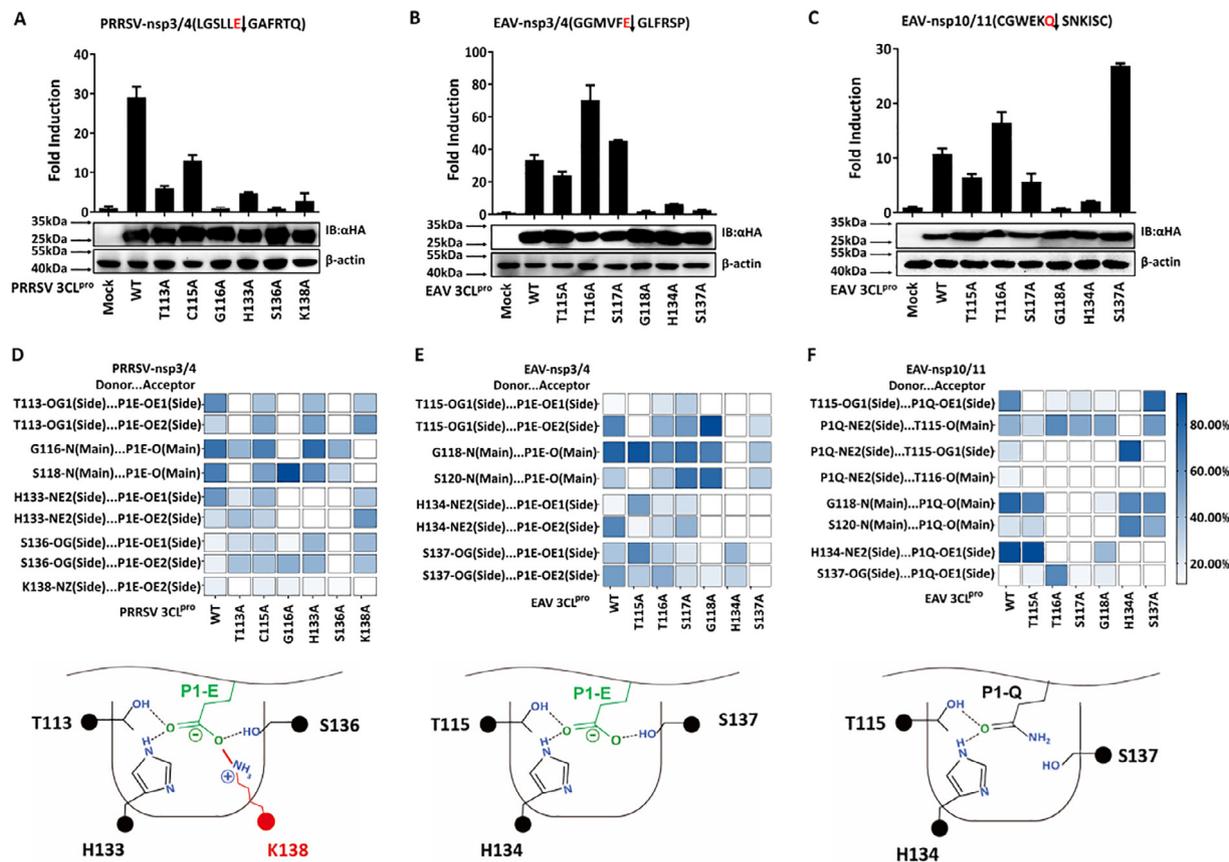


Fig. 4. Identifying the amino acid residues involved in the S1 subsites of arterivirus 3CL^{pro}. (a–c) HEK293T cells were cotransfected with the indicated plasmids encoding WT or mutants of arterivirus 3CL^{pro}, pRL-TK and corresponding biosensor plasmids (a. PRRSV-nsp3/4 biosensor (with P1-Glu); b. EAV-nsp3/4 biosensor (with P1-Glu); c. EAV-nsp10/11 biosensor (with P1-Gln)). Dual-luciferase assays and western blotting were performed at 28 h after the transfection. (d–f) Heatmaps of the variation in H-bond occupancy between P1 residue and S1 subsite of WT or mutants of arterivirus 3CL^{pro}. All H-bonds between S1 subsite and P1 residue occurring with an occupancy rate of at least 10% during MD simulations were investigated. Heatmaps were plotted using GraphPad Prism Version 9.3 (www.graphpad.com). The first one or two characters of the atom name of H-bond comprise the chemical symbol for the atom type. All the atom names beginning with C are carbon atoms, N shows nitrogen and O shows oxygen. The next character is the remoteness indicator code, which follows the Greek alphabet, starting with the alpha carbon (A) and moving on to beta (B), gamma (G), delta (D), epsilon (E), zeta (Z), and eta (H). For example, P1E-OE1 represents the first epsilon oxygen of glutamic acid (E) at the P1 position. Schematic of PRRSV 3CL^{pro} and EAV 3CL^{pro} S1 subsites and their interactions with the preferred P1 residue. PRRSV 3CL^{pro} has a conserved K138 residue at the base of S1 subsite and formed a salt bridge interaction between the substrate P1-E (red). The S136/S137 side-chain participates of PRRSV/EAV 3CL^{pro} in H-bonds with the P1-Glu, whereas the S137 side-chain of EAV 3CL^{pro} points away from the P1-Gln side-chain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

convert P1 substrate specificity of PRRSV 3CL^{pro} to that of EAV 3CL^{pro}. The only amino acid distinction in the binding site region between PRRSV 3CL^{pro} and EAV 3CL^{pro} is in position 138 at the loop near the active site, where Lys is found in PRRSV 3CL^{pro} and Thr in EAV 3CL^{pro}. We generated the T139K mutation in EAV 3CL^{pro}, which we predicted would form a salt bridge linking P1-Glu of the substrate. As shown in Fig. 5a, the T139K mutant of EAV 3CL^{pro} revealed a significant shift in substrate specificity, recognizing substrates containing Glu at P1 more efficiently. Gln-containing substrates were still hydrolyzed, but with much less relative efficiency as compared with the WT EAV 3CL^{pro} (Fig. 5b). Next, we tested the opposite mutations in PRRSV 3CL^{pro} to determine whether mutant PRRSV 3CL^{pro} would behave like EAV 3CL^{pro}. The K138T mutation of PRRSV 3CL^{pro} significantly reduced the P1 substrate specificity toward Glu (Fig. 5c). Simultaneously, the mutant gained the ability to cleave Gln-containing substrates efficiently, an activity that is almost absent in the wild type PRRSV 3CL^{pro}. Interestingly, the K138T mutant exhibits higher expression levels compared to wild-type PRRSV 3CL^{pro} (Fig. 5d). To exclude the possibility that the gained cleavage ability is due to higher expression of the K138T mutant, the experiments of dose–response analysis were performed. Dual-luciferase reporter assay and western blot assay

demonstrated that the gained cleavage ability of K138T to cleave Gln-containing substrates is not due to high expression levels of the mutant (Fig. S5).

Considering that Lys is a large amino acid with a positively charged chain, while Thr is a small amino acid with a polar uncharged side chain, we speculate that electrostatic or vdW might play an important role in P1 substrate specificity shift. To this end, we calculated and compared the electrostatic or vdW energies among arterivirus 3CL^{pro} structures with or without the potential intermolecular salt bridge. As shown in Fig. 6, the dominant binding attractions were determined by several residues in S1 subsites. In contrast with vdW, the electrostatic energy was more favorable for binding (Fig. 6a). As expected, the T139K mutant of EAV 3CL^{pro} formed an intermolecular salt bridge with P1-Glu, producing a strong electrostatic interaction (about -82.83 kJ/mol), whereas K138T mutant of PRRSV 3CL^{pro} significantly reduced the electrostatic force due to the loss of the salt bridge between K138 and substrate P1-Glu (Lys138/P1-Glu) (Fig. 6b). Together, molecular dynamics simulations further show that this salt bridge is particularly critical for P1 substrate specificity of arterivirus 3CL^{pro} via electrostatic forces and that the acquisition of this salt bridge leads to a shift of arterivirus 3CL^{pro} from P1-Gln to P1-Glu preference.

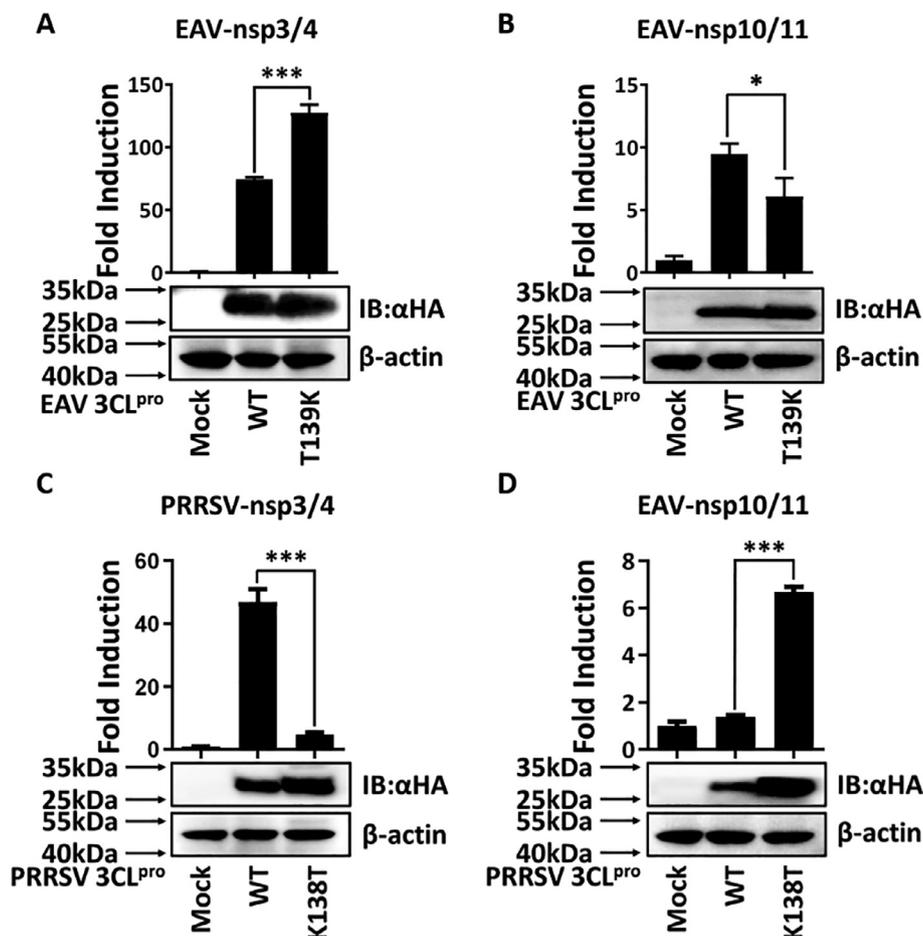


Fig. 5. Identifying an intermolecular salt bridge linking P1 substrate specificity shift of arterivirus 3CL^{pro}. (a–b) HEK293T cells were transfected with plasmid encoding WT or the mutant T139K of EAV 3CL^{pro} (salt bridge formation), pRL-TK and corresponding luciferase-based biosensors plasmids. The cells were harvested after 28 h transfection and analyzed by dual-luciferase assays and western blotting. (c–d) HEK293T cells were cotransfected with a plasmid encoding PRRSV 3CL^{pro} WT or the mutant K138T (salt bridge disruption), pRL-TK and corresponding luciferase-based biosensors plasmids. The cells were harvested after 28 h transfection and analyzed by dual-luciferase assays and western blotting. *p < 0.05; ***p < 0.001

3.6. Evolutionary mechanism of salt-bridge-triggered P1 substrate specificity switch of arterivirus 3CL^{pro}

To exclude the effect of different substrates on the above findings, two different substrates with Gln at P1 position were selected and tested whether the salt bridge disruption of PRRSV 3CL^{pro} leads to the P1 substrate specificity switch. First, we constructed a mutant of PRRSV-nsp3/4 biosensor in which the Glu at P1 position was replaced with Gln and examined their cleavage by PRRSV 3CL^{pro}. As expected, the luciferase activity of P1-E → Q substitution of PRRSV-nsp3/4 biosensor was markedly induced in the K138T mutant of PRRSV 3CL^{pro}-cotransfected cells, whereas no activity was detected in wild type of PRRSV 3CL^{pro}-cotransfected cells (Fig. 7a).

Considering that arterivirus 3CL^{pro} cleaves not only viral polyproteins but also host proteins to facilitate the various steps of the virus infection cycle [30]. Our previous study found that EAV 3CL^{pro}, but not PRRSV 3CL^{pro}, cleaved host protein NEMO at Q205 in the P1 position [30]. We then chose NEMO as a representative host protein to study the effect of salt bridge disruption of PRRSV 3CL^{pro} on the cleavage of the host protein. In the NEMO-Q205 biosensor assay, the P6-P6' positions of EAV 3CL^{pro} recognition regions around NEMO-Q205 were introduced into the cyclized luciferase biosensor (Fig. 3a). Consistent with our previous find-

ings, wild type of PRRSV 3CL^{pro} did not induce the luciferase activity of NEMO-Q205 biosensor, whereas such activity was significantly induced in the K138T mutant of PRRSV 3CL^{pro} cotransfected cells (Fig. 7b). As further evidence supporting this result, western blotting also revealed that the K138T mutant of PRRSV 3CL^{pro} was able to cleave NEMO at Q205, producing an expected faster migrating protein band that was detected by an anti-Flag antibody and had molecular masses of approximately 22 kDa. Presumably, this band is NEMO cleavage product, NEMO-(1–205), mediated by the K138T mutant. As we predicted, no cleavage products were observed for WT PRRSV 3CL^{pro} (Fig. 7c). These data suggested that the salt bridge disruption of PRRSV 3CL^{pro} led to a substrate preference shift of arterivirus 3CL^{pro} from P1-Glu to P1-Gln.

To further investigate the importance of the residue K138 in the salt bridge, two additional mutants (K138L and K138M, mutants with hydrophobic side chain) of PRRSV 3CL^{pro} were constructed. As shown in Fig. 7d–e, these two mutants (K138L and K138M) also gained the ability to cleave the substrate with P1-Gln, while significantly losing the ability to cleave the substrate with P1-Glu, as did the mutants K138A and K138T. Together, these findings identify an evolutionarily accessible mechanism for disrupting or reorganizing salt bridge with only a single mutation of arterivirus 3CL^{pro} to trigger a substrate specificity shift.

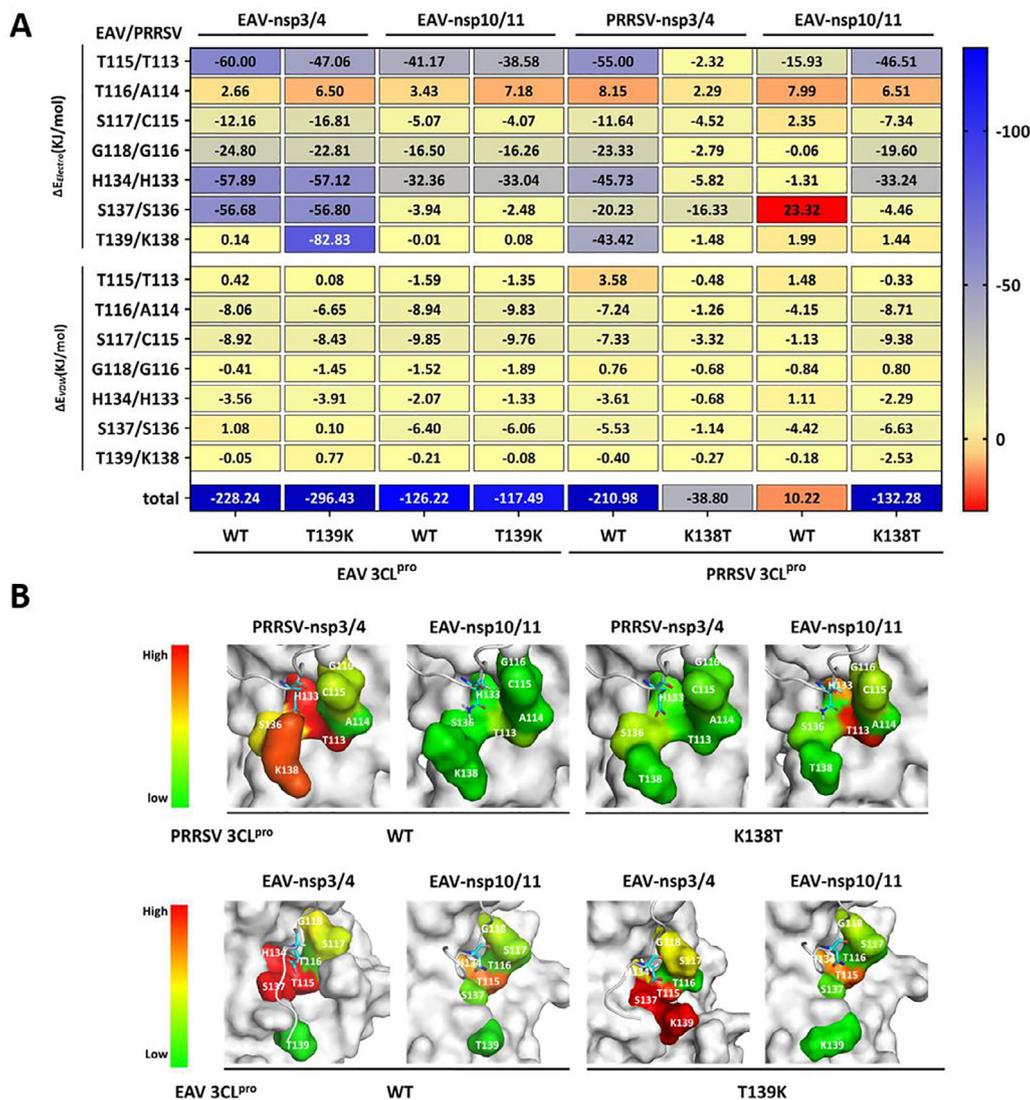


Fig. 6. Electrostatic and vdW interaction energies between P1 residue and S1 subsite of WT arterivirus 3CL^{Pro} or its mutants. (a) Heatmaps represent the different binding free energies between P1 residue and S1 subsite of WT arterivirus 3CL^{Pro} or its mutants, and the contributions from two individual energetic components: electrostatics ($E_{electro}$) and van der Waals (E_{vdW}) energy. Cooler (blue) and warmer (red) colors indicate higher or lower energies contributions to the interaction between P1 residue and S1 subsite, respectively. Heatmaps were plotted using GraphPad Prism Version 9.3 (www.graphpad.com). (b) Comparison of energies surface diagrams of WT arterivirus 3CL^{Pro} and its mutants from MD simulations. The surface representation of the S1 subsite is colored according to the energy of the respective residue positions. Energetically favorable residues are shown as red, neutral as yellow, and unfavorable as green. The P6-P6' positions of substrate are displayed as a white cartoon. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

In most positive-sense, single-stranded RNA viruses, viral 3CL^{Pro} or 3C proteases (3C^{Pro}) cleaved the viral polyprotein at multiple conserved sites and are engaged in translational processing of the viral non-structural proteins [54–59]. The hydrolysis sites of 3C^{Pro}/3CL^{Pro} are broadly resembling and usually include a Gln or Glu residue at the P1 position along with a small amino acid residue downstream [34,60,61]. For arterivirus 3CL^{Pro}, EAV 3CL^{Pro} cleave the bonds formed by both P1-Glu and P1-Gln like some of picornaviruses (e.g., Foot-and-mouth disease virus), while PRRSV 3CL^{Pro} and Glu-specific protease from *S. griseus* (Glu-SGP, the first Glu-specific protease whose spatial structure was determined) prefer P1-Glu [60,62–64]. Here, we established a rapid, sensitive, and efficient luciferase-based biosensor to monitor the activity of arterivirus 3CL^{Pro} *in vivo* and identified key amino acids for P1 substrate specificity in arterivirus 3CL^{Pro} by site-directed mutagenesis and MD simulations. Specifically, a potential intermolecular salt-

bridge in arterivirus 3CL^{Pro} linked substrate binding and P1 substrate switch, this finding could augment the discovery of new specific inhibitors against arterivirus 3CL^{Pro} [65–67].

Previous studies have shown that the structures of the S1 subsites of Glu-SGP, picornavirus 3C^{Pro} and arterivirus 3CL^{Pro} are very close (Fig. S6) [35,63,64,68]. The S1 subsite comprises the three main structural elements: the structures of the S1 subsites have three main structural elements: a conserved histidine residue (His213 in Glu-SGP, His182 in FMDV 3C^{Pro}, His133 in PRRSV 3CL^{Pro} and His134 in EAV 3CL^{Pro}) at the base of the S1 subsite, a conserved Ser/Thr residue lining one “wall” of the S1 subsite (Ser190 in Glu-SGP, Thr159 in FMDV 3C^{Pro}, Thr113 in PRRSV 3CL^{Pro}, Thr115 in EAV 3CL^{Pro}), and a conserved Ser residue (Ser216 in Glu-SGP, Ser136 in PRRSV 3CL^{Pro}, Ser137 in EAV 3CL^{Pro}, corresponding to Gly185 in FMDV 3C^{Pro}). Previously, the significance of residues His134 and Thr115 for EAV 3CL^{Pro} function was supported by site-directed mutagenesis experiments. The mutations His134 also terminated the processing; however, the mutations Thr115 only slightly

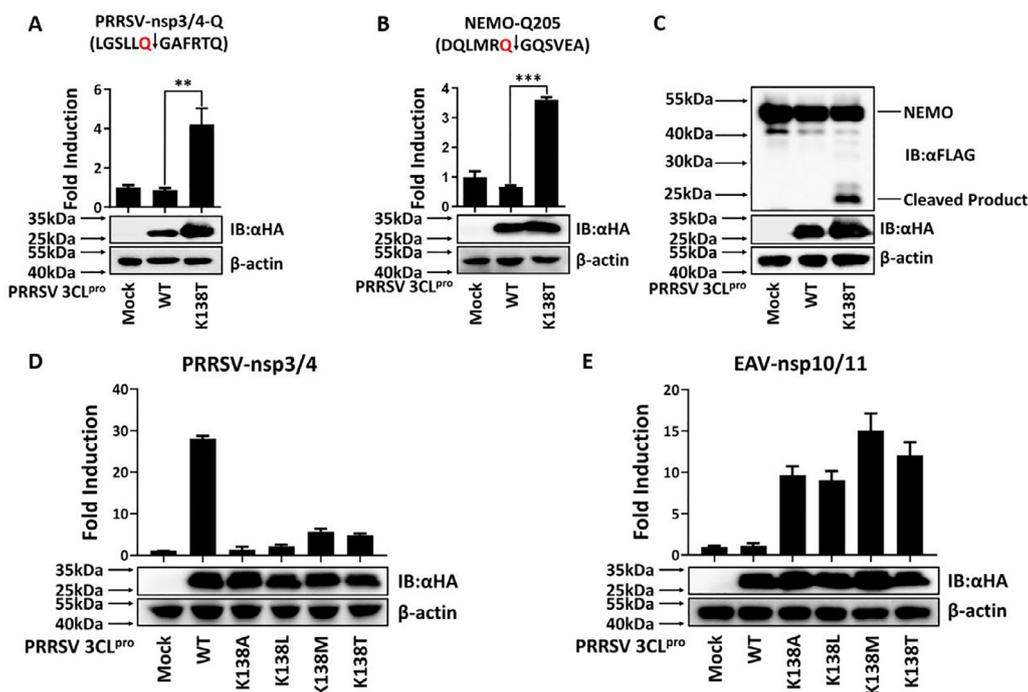


Fig. 7. Evolutionary mechanism of salt-bridge-triggered P1 substrate specificity switch of PRRSV 3CL^{pro}. (a) HEK293T cells were cotransfected with PRRSV-nsp3/4-Q biosensor plasmid, pRL-TK and a plasmid encoding PRRSV 3CL^{pro} WT or the mutant K138T. After 28 h transfection, cell lysates were prepared and analyzed by dual-luciferase assays and western blotting. ***p* < 0.01. (b) HEK293T cells were cotransfected with NEMO-Q205 biosensor plasmid, pRL-TK and a plasmid encoding PRRSV 3CL^{pro} WT or the mutant K138T. Cells lysates were processed as described above for panel A. ****p* < 0.001. (c) HEK293T cells were cotransfected with PRRSV 3CL^{pro} WT or the mutant K138T, along with Flag-tagged NEMO (E166A-E171A-E349A). Cell lysates were prepared 30 h post transfection and analyzed by western blotting. (d-e) HEK293T cells were cotransfected with plasmids encoding WT PRRSV 3CL^{pro} or its mutants, pRL-TK and corresponding biosensor plasmid. After 28 h transfection, cell lysates were prepared and analyzed by dual-luciferase assays and western blotting.

decreased the processing efficiency. Moreover, our findings highlight the essential significance of His134 and the considerably smaller role of the residues Thr115 for the cleavage of substrates with both P1-Glu and P1-Gln by EAV 3CL^{pro}. Meanwhile, our MD analysis also showed that these two residues are essential for the formation of an appropriate geometry of the S1 subsite for recognizing both P1-Glu and P1-Gln (e.g., EAV 3CL^{pro}, FMDV 3C^{pro}). The third element of the S1 site in Glu-SGP is Ser216, which is highly conserved in PRRSV 3CL^{pro} and EAV 3CL^{pro} but replaced with Gly or Ala residues in FMDV 3C^{pro}. It is worth noting that the mutagenesis in the Glu-SGP model indicates that the Ser216Ala/Gly replacement decreased the efficiency in their cleavage activity over P1-Glu substrate but did not make the substrates with P1-Gln the preferred ones [63,69]. Consistent with these results, the mutant S136A of PRRSV 3CL^{pro} lost the ability to cleave substrate with P1-Glu (Fig. 4a) but did not lead to a P1 substrate specificity shift from P1-Glu to P1-Gln preference (data not shown). This suggests that the S136A substitution in PRRSV 3CL^{pro} with that of picornavirus 3C^{pro} may not sufficiently induce the corresponding specificity toward P1-Gln. It should be noted that although the complexes of proteases and short peptide substrates have been widely used for MD simulations to study the interactions between 3CL^{pro} and substrates [70–72], the residues within other region of the substrate have the potential to affect the substrate specificity of arterivirus 3CL^{pro}. Thus, the determination of crystal structures of arterivirus 3CL^{pro} and full-length substrate will provide more details on the molecular determinants of substrate specificities of arterivirus 3CL^{pro}, which deserve further investigation.

As demonstrated in substrate profiling for the Zika virus NS2B-NS3 protease, the preference for a lysine residue in the S2-binding subsite is the result of salt bridge formation between the β-

carboxyl group of the Asp83 cofactor residue and the ε-amino group of the substrate Lys residue [73,74]. A negatively charged S2 binding site accounts for the pronounced preference for lysine and arginine at the P2 position of the substrate, as shown in the P2 round of unfolding. As with the Zika virus NS2B-NS3 protease, the substrate charge compensator is anticipated to be the core structural determinant of the specificity of arterivirus 3CL^{pro} recognizing charged P1-Glu residues [75]. Our data demonstrated that Lys at position 138 of PRRSV 3CL^{pro} might be suggested as candidates for this. Since PRRSV 3CL^{pro} plays an indispensable role in the virus life cycle by cleaving the polyprotein precursors, we attempted to investigate whether the intermolecular salt bridge in PRRSV 3CL^{pro} is important for its replication by a PRRSV reverse genetics approach. However, the salt bridge-breaking mutant of PRRSV 3CL^{pro} (PRRSV 3CL^{pro}-K138T) could not be rescued, and one possible reason is that the salt bridge-breaking mutation almost completely lost the ability to cleave the substrate P1-E in the polyprotein of PRRSV (Fig. 5c). These data suggest that an evolutionary mechanism for the intermolecular salt bridge in PRRSV 3CL^{pro} might be required for its replication.

5. Conclusion

Ensuring P1-Gln substrate specificity in PRRSV 3CL^{pro} required the substitution of Lys138 in the S1 subsite and disruption of a salt bridge between protease and substrate. In turn, the T139K mutant of EAV 3CL^{pro}, building a salt bridge, leads to a shift of EAV 3CL^{pro} from P1-Gln to P1-Glu preference. In conclusion, our findings identified an evolutionarily accessible mechanism for disrupting or reorganizing salt bridge with only a single mutation of viral proteases to trigger a substrate specificity switch.

Author contributions

D.W., Q.C., L.F., S.X. and J.Z. conceptualization; Q.C., J.Z., Z.Y., J.G., Z.L., X.S., and Q.J. methodology; J.Z. and Q.C. data curation; J.Z. and Q.C. software; J.Z. and Q.C. formal analysis; D.W. supervision; Q.C. and D.W. writing—original draft; L.F. and S.X. writing—review & editing; D.W., L.F. and S.X. funding acquisition.

CRedit authorship contribution statement

Qian Chen: Conceptualization, Methodology, Data curation, Software, Formal analysis, Writing – original draft. **Junwei Zhou:** Conceptualization, Methodology, Data curation, Software, Formal analysis. **Zhixiang Yang:** Methodology. **Jiahui Guo:** Methodology. **Zimin Liu:** Methodology. **Xinyi Sun:** Methodology. **Qingshi Jiang:** Methodology. **Liurong Fang:** Conceptualization, Writing – review & editing, Funding acquisition. **Dang Wang:** Conceptualization, Supervision, Writing – original draft, Funding acquisition. **Shaobo Xiao:** Conceptualization, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (32130103, 31872485 and 31941005).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.06.059>.

References

- Brinton MA, Gulyaeva AA, Balasuriya UBR, Dunowska M, Faaberg KS, Goldberg T, et al. ICTV Virus Taxonomy Profile: Arteriviridae 2021. *J Gen Virol* 2021;102. <https://doi.org/10.1099/jgv.0.001632>.
- Balasuriya UB, Carossino M. Reproductive effects of arteriviruses: equine arteritis virus and porcine reproductive and respiratory syndrome virus infections. *Curr Opin Virol* 2017;27:57–70. <https://doi.org/10.1016/j.coviro.2017.11.005>.
- Han J, Wang Y, Faaberg KS. Complete genome analysis of RFLP 184 isolates of porcine reproductive and respiratory syndrome virus. *Virus Res* 2006;122:175–82. <https://doi.org/10.1016/j.virusres.2006.06.003>.
- Neumann EJ, Kliebenstein JB, Johnson CD, Mabry JW, Bush EJ, Seitzinger AH, et al. Assessment of the economic impact of porcine reproductive and respiratory syndrome on swine production in the United States. *J Am Vet Med Assoc* 2005;227:385–92. <https://doi.org/10.2460/javma.2005.227.385>.
- Timoney PJ, McCollum WH. Equine viral arteritis. *Vet Clin North Am Equine Pract* 1993;9:295–309. [https://doi.org/10.1016/s0749-0739\(17\)30397-8](https://doi.org/10.1016/s0749-0739(17)30397-8).
- Wensvoort G, Terpstra C, Pol JM, ter Laak EA, Bloemraad M, de Kluyver EP, et al. Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Vet Q* 1991;13:121–30. <https://doi.org/10.1080/01652176.1991.9694296>.
- Snijder EJ, Meulenbergh JJ. The molecular biology of arteriviruses. *J Gen Virol* 1998;79(Pt 5):961–79. <https://doi.org/10.1099/0022-1317-79-5-961>.
- Balasuriya UB, MacLachlan NJ. The immune response to equine arteritis virus: potential lessons for other arteriviruses. *Vet Immunol Immunopathol* 2004;102:107–29. <https://doi.org/10.1016/j.vetimm.2004.09.003>.
- Bryans JT, Crowe ME, Doll ER, McCollum WH. Isolation of a filterable agent causing arteritis of horses and abortion by mares; its differentiation from the equine abortion (influenza) virus. *Cornell Vet* 1957;47:3–41.
- Botner A, Strandbygaard B, Sorensen KJ, Have P, Madsen KG, Madsen ES, et al. Appearance of acute PRRS-like symptoms in sow herds after vaccination with a modified live PRRS vaccine. *Vet Rec* 1997;141:497–9. <https://doi.org/10.1136/vr.141.19.497>.
- Tian K, Yu X, Zhao T, Feng Y, Cao Z, Wang C, et al. Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. *PLoS ONE* 2007;2:e526.
- Kiss I, Sami L, Kecskemeti S, Hanada K. Genetic variation of the prevailing porcine respiratory and reproductive syndrome viruses occurring on a pig farm upon vaccination. *Arch Virol* 2006;151:2269–76. <https://doi.org/10.1007/s00705-006-0805-0>.
- Stadejek T, Stankevicius A, Murtaugh MP, Oleksiewicz MB. Molecular evolution of PRRSV in Europe: current state of play. *Vet Microbiol* 2013;165:21–8. <https://doi.org/10.1016/j.vetmic.2013.02.029>.
- Holyoak GR, Balasuriya UB, Broaddus CC, Timoney PJ. Equine viral arteritis: current status and prevention. *Theriogenology* 2008;70:403–14. <https://doi.org/10.1016/j.theriogenology.2008.04.020>.
- Snijder EJ, Kikkert M, Fang Y. Arterivirus molecular biology and pathogenesis. *J Gen Virol* 2013;94:2141–63. <https://doi.org/10.1099/vir.0.056341-0>.
- den Boon JA, Snijder EJ, Chirnside ED, de Vries AA, Horzinek MC, Spaan WJ. Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J Virol* 1991;65:2910–20. <https://doi.org/10.1128/JVI.65.6.2910-2920.1991>.
- Li Y, Tas A, Sun Z, Snijder EJ, Fang Y. Proteolytic processing of the porcine reproductive and respiratory syndrome virus replicase. *Virus Res* 2015;202:48–59. <https://doi.org/10.1016/j.virusres.2014.12.027>.
- Conzelmann KK, Visser N, van Woensel P, Thiel HJ. Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group. *Virology* 1993;193:329–39. <https://doi.org/10.1006/viro.1993.1129>.
- Li Y, Tas A, Snijder EJ, Fang Y. Identification of porcine reproductive and respiratory syndrome virus ORF1a-encoded non-structural proteins in virus-infected cells. *J Gen Virol* 2012;93:829–39. <https://doi.org/10.1099/vir.0.039289-0>.
- van Dinten LC, Wassenaar AL, Gorbalenya AE, Spaan WJ, Snijder EJ. Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains. *J Virol* 1996;70:6625–33. <https://doi.org/10.1128/JVI.70.10.6625-6633.1996>.
- van Dinten LC, Rensen S, Gorbalenya AE, Snijder EJ. Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and is essential for virus replication. *J Virol* 1999;73:2027–37. <https://doi.org/10.1128/JVI.73.3.2027-2037.1999>.
- Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol* 2000;81:853–79. <https://doi.org/10.1099/0022-1317-81-4-853>.
- Barrette-Ng IH, Ng KK, Mark BL, van Aken D, Cherney MM, Garenc C, et al. Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. *J Biol Chem* 2002;277:39960–6. <https://doi.org/10.1074/jbc.M206978200>.
- Blanck S, Ziebuhr J. Proteolytic processing of mesonivirus replicase polyproteins by the viral 3C-like protease. *J Gen Virol* 2016;97:1439–45. <https://doi.org/10.1099/jgv.0.000458>.
- Manolaridis I, Gaudin C, Posthuma CC, Zevenhoven-Dobbe JC, Imbert I, Canard B, et al. Structure and genetic analysis of the arterivirus nonstructural protein 7alpha. *J Virol* 2011;85:7449–53. <https://doi.org/10.1128/JVI.00255-11>.
- Wassenaar AL, Spaan WJ, Gorbalenya AE, Snijder EJ. Alternative proteolytic processing of the arterivirus replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease. *J Virol* 1997;71:9313–22. <https://doi.org/10.1128/JVI.71.12.9313-9322.1997>.
- Soumana DI, Kurt Yilmaz N, Ali A, Prachanronarong KL, Schiffer CA. Molecular and Dynamic Mechanism Underlying Drug Resistance in Genotype 3 Hepatitis C NS3/4A Protease. *J Am Chem Soc* 2016;138:11850–9. <https://doi.org/10.1021/jacs.6b06454>.
- Ragland DA, Nalivaika EA, Nalam MN, Prachanronarong KL, Cao H, Bandaranayake RM, et al. 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* 2014;136:11956–63. <https://doi.org/10.1021/ja504096m>.
- Heinz DW, Hyberts SG, Peng JW, Priestle JP, Wagner G, Grutter MG. Changing the inhibitory specificity and function of the proteinase inhibitor eglin c by site-directed mutagenesis: functional and structural investigation. *Biochemistry* 1992;31:8755–66. <https://doi.org/10.1021/bi00152a011>.
- Chen J, Wang D, Sun Z, Gao L, Zhu X, Guo J, et al. Arterivirus nsp4 Antagonizes Interferon Beta Production by Proteolytically Cleaving NEMO at Multiple Sites. *J Virol* 2019;93. <https://doi.org/10.1128/JVI.00385-19>.
- Huang C, Zhang Q, Guo X, Yu Z, Xu A-T, Tang J, et al. Porcine Reproductive and Respiratory Syndrome Virus Nonstructural Protein 4 Antagonizes Beta Interferon Expression by Targeting the NF-κB Essential Modulator. *J Virol* 2014;88. <https://doi.org/10.1128/jvi.01396-14>.
- Ng CS, Stobart CC, Luo H. Innate immune evasion mediated by picornaviral 3C protease: Possible lessons for coronaviral 3C-like protease? *Rev Med Virol* 2021;31. <https://doi.org/10.1002/rmv.2206>.
- Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 2020;582. <https://doi.org/10.1038/s41586-020-2223-y>.
- Tian X, Lu G, Gao F, Peng H, Feng Y, Ma G, et al. Structure and cleavage specificity of the chymotrypsin-like serine protease (3CLSP/nsp4) of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV). *J Mol Biol* 2009;392:977–93. <https://doi.org/10.1016/j.jmb.2009.07.062>.
- Nienaber VL, Breddam K, Birktoft JJ. A glutamic acid specific serine protease utilizes a novel histidine triad in substrate binding. *Biochemistry* 1993;32:11469–75. <https://doi.org/10.1021/bi00094a001>.

- [36] van Aken D, Zevenhoven-Dobbe J, Gorbalenya AE, Snijder EJ. Proteolytic maturation of replicase polyprotein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication and includes internal cleavage of nsp7. *J Gen Virol* 2006;87:3473–82. <https://doi.org/10.1099/vir.0.82269-0>.
- [37] Zhou J, Fang L, Yang Z, Xu S, Lv M, Sun Z, et al. Identification of novel proteolytically inactive mutations in coronavirus 3C-like protease using a combined approach. *FASEB J* 2019;33:14575–87. <https://doi.org/10.1096/fj.201901624RR>.
- [38] Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, et al. Amber 10. University of California; 2008.
- [39] Case DA, Cheatham 3rd TE, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. *J Comput Chem* 2005;26:1668–88. <https://doi.org/10.1002/jcc.20290>.
- [40] Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Miguez AN, et al. Ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* 2020;16. <https://doi.org/10.1021/acs.jctc.9b00591>.
- [41] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 2015;11. <https://doi.org/10.1021/acs.jctc.5b00255>.
- [42] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35.
- [43] Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* 1977;23:327–41.
- [44] Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD. Langevin stabilization of molecular dynamics. *J Chem Phys* 2001;114:2090–8.
- [45] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys* 1995;103:8577–93.
- [46] Schrodinger LLC. The PyMOL Molecular Graphics System. Version 2021;2:5.
- [47] Yao X-Q, Scarabelli G, Skjærven L, Grant BJ. The Bio3D Package: New Interactive Tools for Structural Bioinformatics. *Biophys J* 2014;106. <https://doi.org/10.1016/j.bpj.2013.11.2288>.
- [48] Grant BJ, Skjærven L, Yao XQ. The Bio3D packages for structural bioinformatics. *Protein Sci* 2021;30. <https://doi.org/10.1002/pro.3923>.
- [49] Goettig P, Brandstetter H, Magdolen V. Surface loops of trypsin-like serine proteases as determinants of function. *Biochimie* 2019;166:52–76. <https://doi.org/10.1016/j.biochi.2019.09.004>.
- [50] Shyu YJ, Liu H, Deng X, Hu CD. Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. *Biotechniques* 2006;40:61–6. <https://doi.org/10.2144/000112036>.
- [51] Matayoshi ED, Wang GT, Krafft GA, Erickson J. Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer. *Science* 1979;1990(247):954–8. <https://doi.org/10.1126/science.2106161>.
- [52] Zettler J, Schutz V, Mootz HD. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* 2009;583:909–14. <https://doi.org/10.1016/j.febslet.2009.02.003>.
- [53] Snijder EJ, Wassenaar AL, van Dinten LC, Spaan WJ, Gorbalenya AE. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J Biol Chem* 1996;271:4864–71. <https://doi.org/10.1074/jbc.271.9.4864>.
- [54] Kraut J. Serine proteases: structure and mechanism of catalysis. *Annu Rev Biochem* 1977;46:331–58. <https://doi.org/10.1146/annurev.bi.46.070177.001555>.
- [55] Hartley BS. Homologies in serine proteinases. *Philos Trans R Soc Lond B Biol Sci* 1970;257:77–87. <https://doi.org/10.1098/rstb.1970.0010>.
- [56] Harper E, Berger A. On the size of the active site in proteases: pronase. *Biochem Biophys Res Commun* 1972;46:1956–60. [https://doi.org/10.1016/0006-291x\(72\)90076-9](https://doi.org/10.1016/0006-291x(72)90076-9).
- [57] Kitamura N, Semler BL, Rothberg PG, Larsen GR, Adler CJ, Dorner AJ, et al. Primary structure, gene organization and polypeptide expression of poliovirus RNA. *Nature* 1981;291:547–53. <https://doi.org/10.1038/291547a0>.
- [58] Matthews DA, Smith WW, Ferre RA, Condon B, Budahazi G, Sisson W, et al. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* 1994;77:761–71. [https://doi.org/10.1016/0092-8674\(94\)90059-0](https://doi.org/10.1016/0092-8674(94)90059-0).
- [59] Selina PI, Karaseva MA, Komissarov AA, Safina DR, Lunina NA, Roschina MP, et al. Embryotoxic activity of 3C protease of human hepatitis A virus in developing Danio rerio embryos. *Sci Rep* 2021;11:18196. <https://doi.org/10.1038/s41598-021-97641-5>.
- [60] Zunszain PA, Knox SR, Sweeney TR, Yang J, Roque-Rosell N, Belsham GJ, et al. Insights into cleavage specificity from the crystal structure of foot-and-mouth disease virus 3C protease complexed with a peptide substrate. *J Mol Biol* 2010;395:375–89. <https://doi.org/10.1016/j.jmb.2009.10.048>.
- [61] den Boon JA, Faaberg KS, Meulenberg JJ, Wassenaar AL, Plagemann PG, Gorbalenya AE, et al. Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases. *J Virol* 1995;69:4500–5. <https://doi.org/10.1128/JVI.69.7.4500-4505.1995>.
- [62] Parks GD, Palmenberg AC. Site-specific mutations at a picornavirus VP3/VP1 cleavage site disrupt in vitro processing and assembly of capsid precursors. *J Virol* 1987;61:3680–7. <https://doi.org/10.1128/JVI.61.12.3680-3687.1987>.
- [63] Stennicke HR, Birktoft JJ, Breddam K. Characterization of the S1 binding site of the glutamic acid-specific protease from *Streptomyces griseus*. *Protein Sci* 1996;5:2266–75. <https://doi.org/10.1002/pro.5560051113>.
- [64] Birtley JR, Knox SR, Jaulent AM, Brick P, Leatherbarrow RJ, Curry S. Crystal structure of foot-and-mouth disease virus 3C protease. New insights into catalytic mechanism and cleavage specificity. *J Biol Chem* 2005;280:11520–7. <https://doi.org/10.1074/jbc.M413254200>.
- [65] Gouhar SA, Elshahid ZA. Molecular docking and simulation studies of synthetic protease inhibitors against COVID-19: a computational study. *J Biomol Struct Dyn* 2021. <https://doi.org/10.1080/07391102.2021.1997822>.
- [66] Jin Z, Mantri Y, Retout M, Cheng Y, Zhou J, Jorns A, et al. A Charge-Switchable Zwitterionic Peptide for Rapid Detection of SARS-CoV-2 Main Protease. *Angewandte Chemie - International Edition* 2022;61. <https://doi.org/10.1002/anie.202112995>.
- [67] Drayman N, DeMarco JK, Jones KA, Azizi SA, Froggatt HM, Tan K, et al. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. *Science* 1979;2021:373. <https://doi.org/10.1126/science.abg5827>.
- [68] Kim Y, Lovell S, Tiew KC, Mandadapu SR, Alliston KR, Battaile KP, et al. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J Virol* 2012;86:11754–62. <https://doi.org/10.1128/JVI.01348-12>.
- [69] Demidyuk IV, Chukhontseva KN, Kostrov SV. Glutamyl Endopeptidases: The Puzzle of Substrate Specificity. *Acta Naturae* 2017;9:17–33.
- [70] Suárez D, Díaz N. SARS-CoV-2 Main Protease: A Molecular Dynamics Study. *J Chem Inf Model* 2020;60. <https://doi.org/10.1021/acs.jcim.0c00575>.
- [71] Ferreira JC, Fadl S, Rabeh WM. Key dimer interface residues impact the catalytic activity of 3CLpro, the main protease of SARS-CoV-2. *J Biol Chem* 2022;298. <https://doi.org/10.1016/j.jbc.2022.102023>.
- [72] Ramos-Guzmán CA, Ruiz-Pernía JJ, Tuñón I. Multiscale Simulations of SARS-CoV-2 3CL Protease Inhibition with Aldehyde Derivatives. Role of Protein and Inhibitor Conformational Changes in the Reaction Mechanism. *ACS Catal* 2021;11. <https://doi.org/10.1021/acscatal.0c05522>.
- [73] Gruba N, Rodríguez Martínez JJ, Grzywa R, Wysocka M, Skorenski M, Burmistrz M, et al. Substrate profiling of Zika virus NS2B-NS3 protease. *FEBS Lett* 2016;590:3459–68. <https://doi.org/10.1002/1873-3468.12443>.
- [74] Meijers R, Blagova EV, Levdkov VM, Rudenskaya GN, Chestukhina GG, Akimkina TV, et al. The crystal structure of glutamyl endopeptidase from *Bacillus intermedius* reveals a structural link between zymogen activation and charge compensation. *Biochemistry* 2004;43:2784–91. <https://doi.org/10.1021/bi035354s>.
- [75] Cortes A, Emery DC, Halsall DJ, Jackson RM, Clarke AR, Holbrook JJ. Charge balance in the alpha-hydroxyacid dehydrogenase vacuole: an acid test. *Protein Sci* 1992;1:892–901. <https://doi.org/10.1002/pro.5560010707>.