ACS Catalysis

pubs.acs.org/acscatalysis

Research Article

Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease

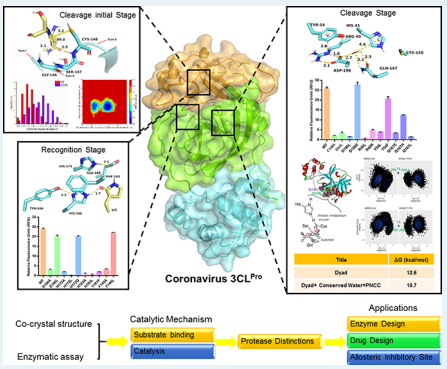
Hao Wang, Shuai He, Weilong Deng, Ying Zhang, Guobang Li, Jixue Sun, Wei Zhao, Yu Guo, Zheng Yin, Dongmei Li,* and Luqing Shang*

Cite This: *ACS Catal.* 2020, 10, 5871−5890

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Coronavirus 3C-like protease ($3CL^{Pro}$) is a highly conserved cysteine protease employing a catalytic dyad for its functions. $3CL^{Pro}$ is essential to the viral life cycle and, therefore, is an attractive target for developing antiviral agents. However, the detailed catalytic mechanism of coronavirus $3CL^{Pro}$ remains largely unknown. We took an integrated approach of employing X-ray crystallography, mutational studies, enzyme kinetics study, and inhibitors to gain insights into the mechanism. Such experimental work is supplemented by computational studies, including the prereaction state analysis, the ab initio calculation of the critical catalytic step, and the molecular dynamic simulation of the wild-type and mutant enzymes. Taken together, such studies allowed us to identify a residue pair (Glu-His) and a conserved His as critical for binding; a conserved GSCGS motif as important for the start of catalysis, a partial negative charge cluster (PNCC) formed by Arg-Tyr-Asp as essential for catalysis, and a conserved water molecule mediating the remote interaction between PNCC and catalytic dyad. The data collected and our insights into the detailed mechanism have allowed us to achieve a good understanding of the difference in catalytic efficiency between $3CL^{Pro}$ from SARS and MERS, conduct mutational studies to improve the catalytic activity by 8-fold, optimize existing inhibitors to improve the potency by 4-fold, and identify a potential allosteric site for inhibitor design. All such results reinforce each other to support the overall catalytic mechanism proposed herein.



KEYWORDS: *MERS-CoV $3CL^{Pro}$, MERS-CoV $3CL^{Pro}$, catalytic mechanism, GSCGS motif, partial negative charge cluster, conserved water*

INTRODUCTION

The COVID-19 (Coronavirus Disease-2019) outbreak is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and has become a public health emergency worldwide.[1,2] Relevant to the COVID-19 pandemic, the Severe Acute Respiratory Syndrome (SARS) in 2003 and the Middle East Respiratory Syndrome (MERS) in 2015 also caused tremendous social panic due to their high fatality rates at 10% and 30%, respectively.[3−6] All three syndromes are caused by coronaviruses belonging to the same beta-coronavirus family. 3C-Like protease ($3CL^{Pro}$) is a cysteine protease critical to the life cycles of all three types of coronaviruses. Thus, it is an attractive target for drug design, especially for designing pan-coronavirus antivirals.[7−9] We are interested in examining the mechanistic details of $3CL^{Pro}$ with the hope of aiding in drug design efforts.

Specifically, the MERS coronavirus (MERS-CoV) possesses a single-stranded positive-sense RNA genome with 2 open reading frames (ORFs) and encodes two polyprotein precursors,[10−13] which are cleaved by $3CL^{Pro}$ and a papain-like cysteine protease (PL^{Pro}) to generate 16 nonstructural proteins (NSP1−16).[14−17] Among them, $3CL^{Pro}$, as the main protease, is critical for most of the cleavage events during polyprotein processing.[18,19] The catalytic mechanism of MERS-CoV $3CL^{Pro}$ is largely unknown. For SARS-CoV $3CL^{Pro}$, a Gln in the substrate P1 position and a small amino acid residue, such as Gly, Ser, or Ala, in the P1′ position are proposed to be essential based on an analysis of the active sites sequence.[20−23] In addition, a Cys···His catalytic dyad is reported to complete the proteolytic task through a common nucleophilic-type reaction in SARS-CoV $3CL^{pro}$ (Figure 1).[24,25] Although the substrate specificity and nucleophilic-attack model of the SARS-CoV $3CL^{pro}$ dyad have been illustrated, the detailed catalytic mechanism of coronavirus $3CL^{Pro}$ is still unclear. This inspires us to work on deciphering the comprehensive mechanism of coronavirus $3CL^{Pro}$.

Received: January 9, 2020
Revised: April 28, 2020
Published: April 28, 2020

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

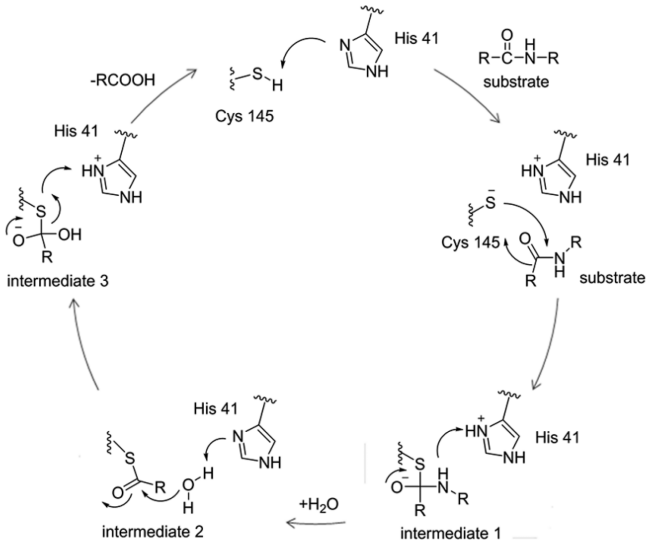


Figure 1. General overall scheme of the SARS-CoV 3CL^{Pro} catalytic mechanism.

Here, we report the comprehensive molecular catalytic mechanism (binding and catalysis) of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. For substrate binding, the protease adopted a residue pair (Glu-His) and a stable hydrogen bond formed by a conserved His and the substrate Gln to recognize the conserved Gln in the P1 position of the substrate. For catalysis, a conserved GSCGS motif was identified and shown to stabilize the active site of the substrate and maintain the mobility of the catalytic cysteine side chain. In addition, a partial negative charge cluster (PNCC) formed by Arg-Tyr-Asp was proven to be essential for catalysis, via a remote interaction mediated by a conserved water molecule. The distinctions between proteases from SARS and MERS were explored in terms of their catalytic efficiency. Furthermore, mutation studies were conducted to improve the enzymatic activity (8-fold), the inhibitor of MERS-CoV 3CL^{Pro} was optimized to improve the potency via introduction of a powerful hydrogen bond interaction between the inhibitor and the critical Q195 of MERS-CoV 3CL^{Pro}. A potential allosteric inhibitory site was identified. Hence, the results will help in the understanding of the enzymology, as well as de novo protein design and novel inhibitor development.

MATERIALS AND METHODS

Enzyme Preparation. The wild type (WT) MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} genes were synthesized by Genewiz, Inc. Subsequently, MERS-CoV 3CL^{Pro} was cloned into the modified pET-28bs vector (Novagen), and the SARS-CoV 3CL^{Pro} was constructed in the pEGX-6P vector (Novagen). The detailed primer information was presented in Table S1 of the Supporting Information (SI). Then, the constructed plasmids were transformed into *Escherichia coli* BL21 (DE3) cells (TransGen Biotech, Beijing, China), and target protein was induced by 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 16 °C for 18 h. The harvested cells were resuspended into lysis buffer containing 20 mM Tris-HCl (pH 8), 150 mM NaCl, 4 mM $MgCl_2$, 5% glycerol, and homogenized with ultrasonic cell disintegration at low temperature. Following centrifugation at 12 000 rpm for 40 min at 4 °C to remove cell debris, the supernatant was loaded onto the Ni-nitrilotriacetic acid (Ni-NTA) column (GE Healthcare). After washing the resin with the washing buffer containing 20 mM imidazole (pH 8), SUMO protease was added to generate MERS-CoV 3CL^{Pro} or washed with buffer containing 200 mM imidazole (pH 8) to separate SARS-CoV 3CL^{Pro}. Crude protein was purified by Superdex 75 gel filtration chromatography (GE Healthcare) or Superdex 200 gel filtration chromatography (GE Healthcare) and verified by SDS-PAGE analysis (Figure S1). Finally, the target protein was concentrated into 30 mg/mL and stored at −80 °C.

Protein Mutation. The mutational protein was prepared by using the Fast Mutagenesis System Kit (Transgen Biotech Co. LTD) following the manufacturer's instructions. The primers of mutates were presented in Table S2. Following mutagenesis, the mutational recombinant plasmid was verified via gene sequencing, and the mutational proteases were expressed in the enzyme preparation method.

Activity Measurements. The FRET-based peptide NMA-TSAVLQSGFRK(DNP)M was synthesized via a solid-phase method and used as a substrate, which turned fluorescent upon cleavage of the Gln-Ser bond by 3CL^{pro}. In brief, 2.0 μM MERS-CoV 3CL^{Pro} was incubated with seven different concentrations of the inhibitor (2-fold dilution) including DMSO only as blank control in 50 μL assay buffer (pH = 8.0, 20 mM Tris-HCl, 150 mM NaCl) at 37 °C for 30 min. Subsequently, the reaction was initiated following the addition of 30 μM solution of the substrate (50 μL). The change of relative fluorescence units was obtained by a microplate reader (Thermo Varioskan Flash, U.S.A.) at λ_{ex} of 340 nm and λ_{em} of 440 nm. As a consequence, the IC_{50} value of the inhibitor was calculated based on the inhibitory curve fitting by GraphPad Prism 7.0. To determine the kinetic parameters of the cleavage reaction, the prerequisite was to calculate the relationship between the relative fluorescence units and the substrate concentration via precalibrating the instrument with the free fluorescent moiety NMA standard. The kinetic parameters (K_m and k_{cat}) of enzyme and mutations were calculated via kinetics curve fitting by the GraphPad Prism 7.0.[26]

Chemistry. To study the identifying process of protease, several known inhibitors were synthesized as probe molecules.[27] The synthesis routes were presented in Schemes S1 and S2. In brief, intermediates **4a** and **4b** were obtained from L-glutamate acid following protection of the amino acid, S_N2 substitution reaction, and reduction reaction. Following subsequent condensation reaction, reductive reaction, and oxidative reaction, aldehydes **8a** and **8b** were obtained.[27] Meanwhile, aldehydes **12a** and **12b** were synthesized via similar methods.

Crystallization, Data Collection, Structure Determination, and Refinement. Purified SARS-CoV 3CL^{Pro} was concentrated to 4 mg/mL in the buffer, which contains 20 mM Tris-HCl, 150 mM NaCl, and a pH of 8.0. The inhibitors were preincubated with SARS-CoV 3CL^{Pro} in 5:1 stoichiometric ratio at 4 °C overnight. After iterative rounds of optimization of the crystallization conditions, the crystals of SARS-CoV 3CL^{Pro} were suitable to grow in 0.1 M MES, pH 6.0, 2%−10% PEG 8000 buffer condition, and in a hanging-drop, vapor diffusion method at 16 °C. For collecting X-ray diffraction data, the crystals were flash-cooled in liquid nitrogen followed by dragging the crystals through the crystallization solution supplemented with 20% glycerol. The X-ray diffraction data were collected at the Rigaku RU200 X-ray generator at Tsinghua University. The data were processed using the HKL2000 package. The space group was identified as C2, and one molecule was detected per asymmetric unit. The crystal structure of SARS-CoV 3CL^{Pro} (PDB code: 1UJ1) was used as

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

the initial searching model to determine the complex structure we obtained. Subsequently, the manual model was refined by performing COOT and PHENIX software through rigid-body refinement, energy minimization, and individual b-factor refinement. Finally, the quality of the final refined model was verified by the program PHENIX validation module and the statistics information were summarized in Table 1 and 2.

Table 1. Data Collection and Refinement Statistics

PDB accession no.	6LNQ
data collection statistics	
X-ray Source	Rigaku RU200 X-ray generator
wavelength (Å)	1.5418
space group	C 121
unit cell parameters (Å; °)	a = 109.068, b = 80.529, c = 53.354; α = 90.00, β = 104.36, γ = 90.0
resolution range* (Å)	50.00−2.28 (2.28−2.24)*
unique reflections	21337 (1033)
completeness (%)	99.9 (97.8)
redundancy	6.5 (5.4)
$I/\sigma(I)$	16.88 (3.36)
R_{merge} (%)	12.2 (53.9)
refinement statistics	
resolution range (Å)	43.55−2.244 (2.325−2.244)
reflections used in refinement	21323 (2010)
reflections used for R-free	1094 (113)
R_{work} (%) *	19.92 (30.51)
R_{free} (%) *	23.90 (34.75)
number of non-hydrogen atoms	2508
protein	2371
ligands	37
solvent	100
average B-factors	30.24
protein	29.88
ligands	47.04
solvent	32.41
r.m.s. deviations	
bond lengths (Å)	0.008
bond angles (deg)	0.87
Ramachandran	
favored (%)	97.04
allowed (%)	2.96
outliers (%)	0.00

MD Simulations. The Amber FF14SB force field was utilized and both of the systems were minimized for 10 000 steps, containing 5000 steps of steepest descent minimization and 5000 steps of conjugate gradient minimization.[28−30] Then the systems were heated from 0 to 300 K in 1 ns constant volume MD simulation. In the heating stage, a force constant of 20 kcal/mol was employed to constrain the complex and the Langevin thermostat was utilized for temperature control.[31] After that, a 100 ns MD simulation was performed for each system without any constraints. In the MD simulation, the cutoff value of the van der Waals interactions was set to 10 Å. The Particle Mesh Ewald (PME) method was applied to calculate the long-range electrostatic interactions, and the SHAKE algorithm was applied to restrain all of the bond lengths that involved hydrogen atoms. Snapshots of the system were saved every 10 ps.[32,33]

CPPTRAJ, Pocket Volume Measurer (POVME) and Molecular Mechanics (MM) Calculation Methods. The CPPTRAJ module was applied to calculate the Root Mean Square Deviations (RMSD), distance, dihedral, angle, and solvent accessible surface area (SASA) of each system in AmberTools15.[34] The redundant volume was calculated by the Pocket Volume Measurer (POVME) script.[35] For calculating the binding free energy between two significant residues, Molecular Mechanics (MM) calculation was introduced. To execute the MM calculation, in brief, a total of 500 snapshots were extracted from the last 5 ns trajectory of each system, and all parameters were used in default values in the calculation.[36−38]

Potential of Mean Force (PMF) Calculation. For generating an energy landscape to further explore the energetic change, the PMF was used and the energy landscape was obtained via the equation:

$$\Delta G(x, y) = k_B T \ln g(x, y) \quad (1)$$

In the equation, k_B represents the Boltzmann constant, T is the simulation temperature, and the $g(x, y)$ is the normalized probability distribution. The explicit relative energy bar is presented near the energy landscape.[39]

Prereaction State (PRS) Analysis. The initial protein structure used in the Prereaction State (PRS) analysis was constructed with the thiolate-imidazolium ion pair model. Basically, the first step of nucleophilic reaction was assumed to be critical in the cascaded mechanism.[40,41] Accordingly, the two complexations of peptide substrate and the protease were constructed for MERS-CoV Wild type (exp. K_m:23.1 ± 2.1 μM, k_{cat}:0.38 ± 0.02 min^{-1}; k_{cat}/K_m: ∼16.4 mM^{-1} min^{-1}) and mutant M168L/T174 V (exp. K_m: 9.2 ± 1.1 μM, k_{cat}: 1.27 ± 0.06 min^{-1}; k_{cat}/K_m: ∼137.2 mM^{-1} min^{-1}) with the mutate module in Discovery Studio software package, using the QM-calculated transition state information. Water molecules were assigned with the TIP3P model, and the ff14SB force field was applied for the classical molecular dynamics simulation. The complexes were placed in a truncated octahedral box of water molecules, extending 10.0 Å along each dimension. A certain number of counterions Na^+ were added to neutralize the calculated system. The MD systems were first minimized by the steepest descent minimization of 1000 steps followed the conjugate gradient minimization of 9000 steps, heated up from 0 to 300 K at constant volume in 50 ps, and equilibrated for another 50 ps without any restraints. In the MD simulations, the Particle Mesh Ewald (PME) method was employed for long-range electrostatic interactions. Finally, multiple 10 ns of trajectories (100 000 frames) were collected for the further PRS analysis, similar to our previous studies.[42−44]

Ab Initio Calculations. For each reaction system, a snapshot that was close to the average simulated structure was extracted from the stable MD trajectory as the starting structure for the ab initio calculations. Specifically, we tracked the distance between the sulfur atom of C148 and the carbonyl carbon atom of substrate active site in the attacking state which conformed Burgi−Dunitz criteria of near-attach-conformation parameters (the S---C distance <3.5 Å and attacking angle alpha (100−110°) and SCOCa dihedral (85−95°), i.e. distance D1 in Scheme S3. Within stable MD trajectories, we obtained an average D1 value and chose the snapshot with the D1 value being close to the corresponding average D1 value. Each system was truncated as a model (Scheme S3) to mimic the reaction pathway in the protein environment. During the geometry optimization, the boundary atoms were fixed at the position they were in in the protein environment. This ensured that each reaction moiety stayed in the same orientation as that in the

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

Table 2. Data Collection and Refinement Statistics

PDB accession no.	6LNY	6LO0
data collection statistics		
X-ray source	Rigaku RU200	Rigaku RU200
wavelength (Å)	1.540	1.540
space group	*C* 121	*C* 121
unit cell parameters (Å; °)	a = 109.049, b = 80.67, c = 53.407; α = 90.00, β = 104.491, γ = 90.00	a = 108.669, b = 81.544, c = 53.382; α = 90.00, β = 104.349, γ = 90.00
resolution range* (Å)	50.00−2.29 (2.29−2.25)	50.00−2.28 (2.28−2.24)
unique reflections	21307 (1001)	21495 (930)
completeness (%)	99.5 (95.2)	98.7 (87.3)
redundancy	6.2 (4.7)	6.5(5.5)
$I/\sigma(I)$	19.21 (2.65)	21.68 (6.27)
R_{merge} (%) *	10.6 (67.3)	9 (31.8)
refinement statistics		
resolution range (Å)	43.71−1.94 (2.325−2.245)*	43.71−1.939 (2.008−1.939)*
reflections used in refinement	21274 (1995)	27854 (867)
reflections used for R-free	1997 (187)	1412 (66)
R_{work} (%) *	20.41 (27.72)	22.72 (40.90)
R_{free} (%) *	25.28 (34.41)	26.83 (46.91)
number of non-hydrogen atoms	2510	2589
Protein	2371	2371
Ligands	30	29
solvent	109	189
average B-factors	32.98	29.46
protein	32.72	29.06
ligands	40.06	37.04
solvent	36.72	33.32
r.m.s. deviations		
bond lengths (Å)	0.009	0.007
bond angles (deg)	0.97	0.92
Ramachandran		
favored (%)	97.70	97.70
allowed (%)	2.30	2.30
outliers (%)	0.00	0.00

protein environment. Geometry optimization was conducted using the B3LYP functional and a uniform 6-31+G* basis set.[45−47] Frequency analysis was used to calculate the Gibbs free energies at 298.15 K and 1 atm. Single-point energy calculation was carried out on the optimized geometry with the MP2 method and the 6-31+G* basis set. The SMD solvation method with a dielectric constant of ε = 5.6968 (chlorobenzene) was used to model the weak polarization effect of the protein environment. The ab initio calculation was carried out using the Gaussian 09 program.[48]

RESULTS

Overall Investigation of the Structures and the Enzymatic Activity of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}, as well as the Catalytic Center in MERS-CoV 3CL^{Pro}. With the aim of understanding the molecular catalytic pattern of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}, it is important to examine the experimentally determined structure of the protease with a bound ligand. For MERS-CoV 3CL^{Pro}, we started with a reported complex structure (PDB code: 4RSP) for further investigation. To check the catalytic mechanism and unravel the general coronavirus 3CL^{Pro} catalytic mechanism, we sought to determine the complex structure of SARS-CoV 3CL^{Pro} with a peptidomimetic aldehyde inhibitor (PDB code: 6LNQ; for detailed information, see Table 1). Generally, MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} appear as a quasi-ellipsoid and possess three separate portions (domain I, domain II, and domain III), which are connected by flexible loops (Figure 2A). More narrowly, the structures of the two proteases chiefly comprise 10 helices, α1−α10, and 13 β-sheets, β1−β13 (Figure S2). Therein, domains I and II are reported to execute catalytic activity, while domain III is responsible for protease dimerization.[49] By superimposing the two structures, the root-mean-square deviation (RMSD) value indicated that the catalytic zone (domain I and domain II) presented higher similarity than the dimerization domain (domain III) (Figure 2A). In addition, similar results emerged among four other coronavirus 3CL^{Pro} proteins (human coronavirus HKU1 (HCoV-HKU1), bat coronavirus HKU4 (BCoV-HKU4), human coronavirus NL63 (HCoV-NL63), and human coronavirus 229E (HCoV-229E)), which hinted that various coronaviruses 3CL^{Pro} have a conserved proteolysis mechanism (Figure S3).

For establishment of the enzymatic assay, the active sites of the polypeptide precursors were presented via bioinformatics analysis, and a FRET-based dodecapeptide as an experimental substrate was constructed following analysis of these active sites (Figures 2B and S4). For validating the availability of the

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

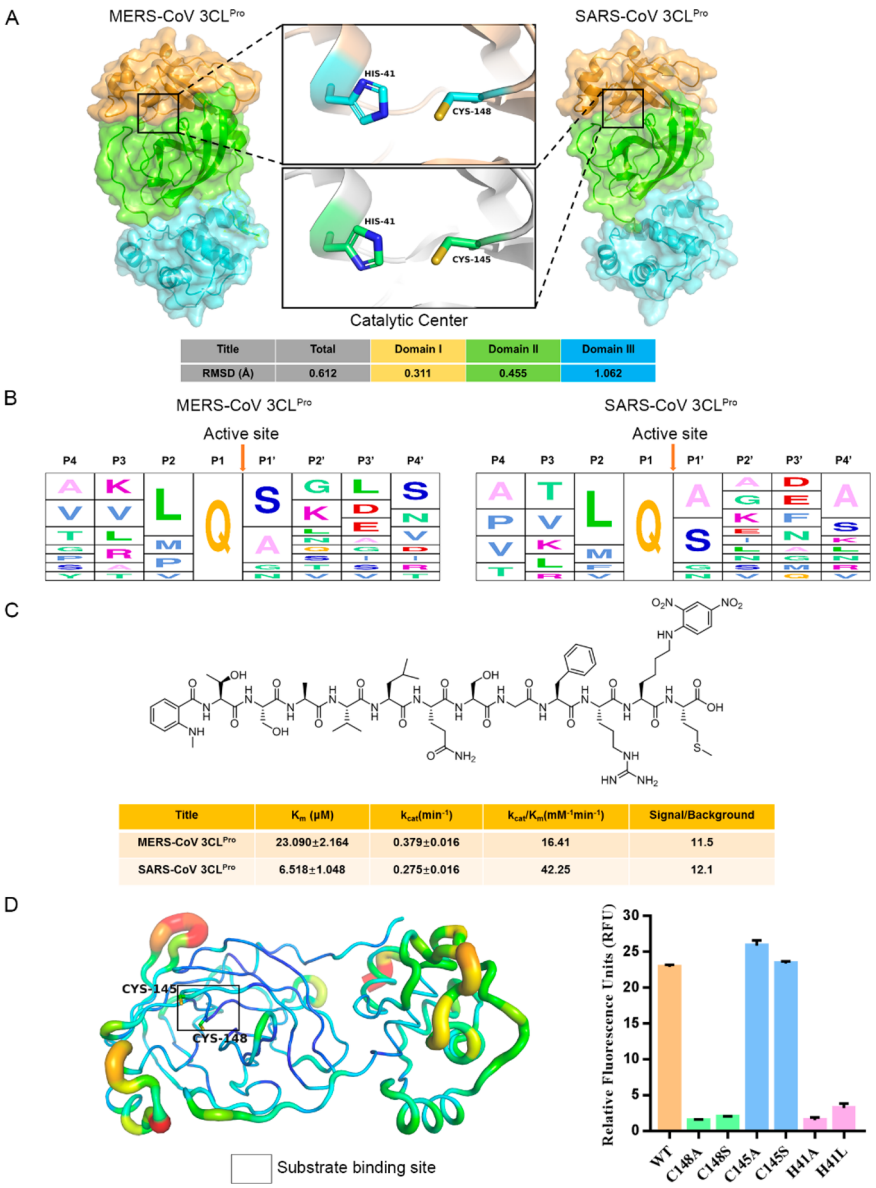


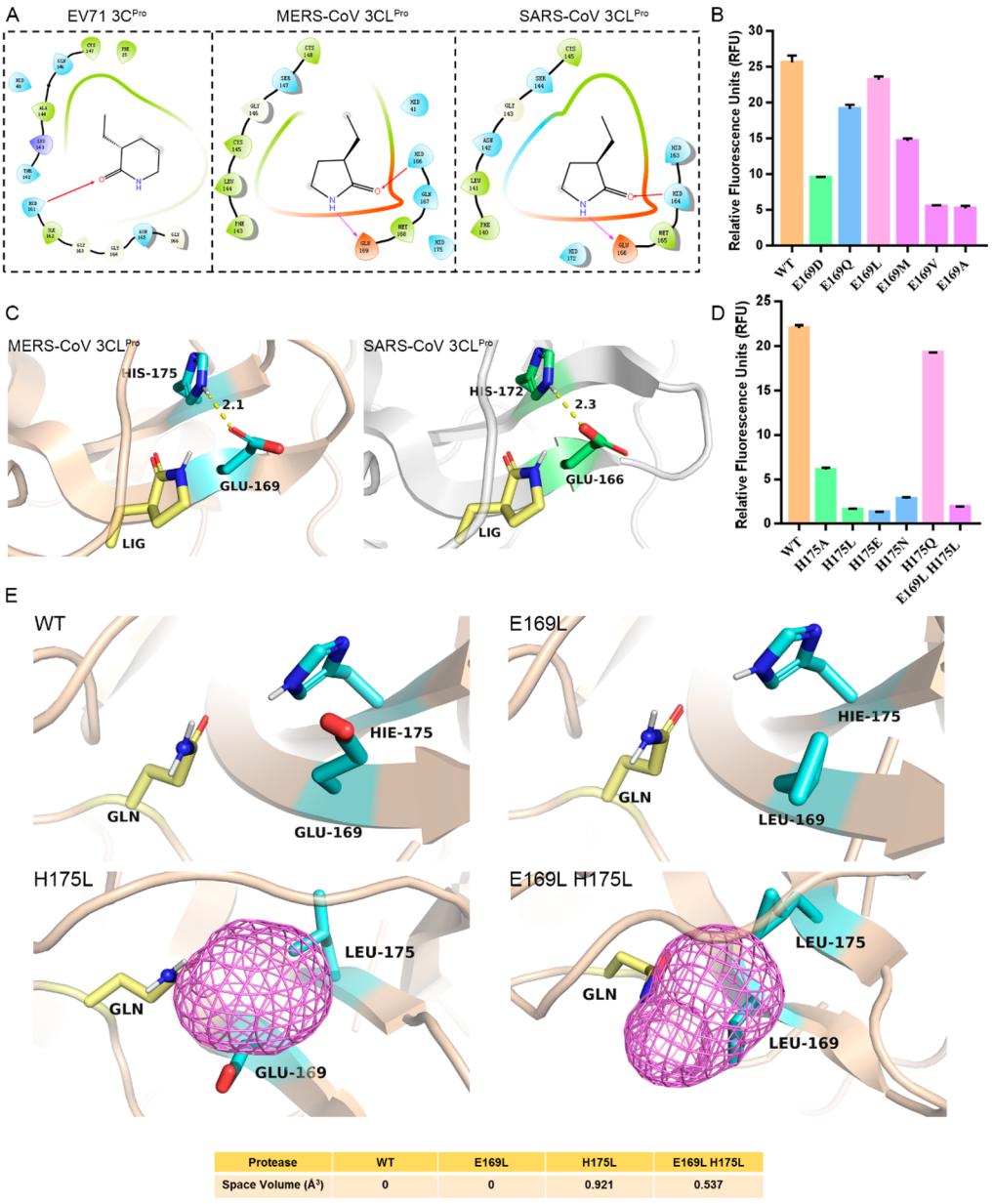
Figure 2. Structure of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}, enzymatic assay development and verification of the MERS-CoV 3CL^{Pro} catalytic center. (A) The overall structure of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. Domain I is presented in orange, Domain II is presented in green, and Domain III is presented in cyan. The RMSD distinction between MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} is shown in the table below. The catalytic centers of the two proteases are shown in the center of the image. The 3CL^{Pro} catalytic dyad is displayed as sticks (MERS-CoV 3CL^{Pro} in cyan and SARS-CoV 3CL^{Pro} in green) and the protease is displayed as a cartoon (MERS-CoV 3CL^{Pro} in wheat color and SARS-CoV 3CL^{Pro} in white). (B) Bioinformatics analysis of the substrate active site sequence. (C) The structure of the designed substrate and the kinetic parameters of the substrate. (D) Catalytic center verification of MERS-CoV 3CL^{Pro}. The structure of MERS-CoV 3CL^{Pro} is shown as a b-factor to indicate that the two cysteines are stable in the structure. The two cysteines required to investigate are emphasized, and the frame represents the binding site of the substrate on the left side. Enzymatic assays of the wild type and mutant proteases are shown on the right side. The data presented are the mean values from experiments in triplicate and the error bars indicate the standard deviations.

substrate, MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} were tested, and the kinetic dynamic parameters of the proteases were determined (Figures 2C and S5).

Identification of the catalytic center is essential to provide a foundation for decoding the catalytic mechanism. In MERS-CoV 3CL^{Pro}, two cysteines (C145 and C148) are located in the catalytic center (Figure 2D), whereas SARS-CoV 3CL^{Pro} exploits only C145 to execute catalysis. Nevertheless, C148 has been reported to be essential to the protease, although the significance of C145 in MERS-CoV 3CL^{Pro} is still unknown. To examine the influence of C145 in MERS-CoV 3CL^{Pro}, we mutated both Cys to Ser and Ala. Enzymatic assays using the purified variants C145S and C145A showed that these mutations manifested comparable efficiency to WT MERS-CoV 3CL^{Pro}, while the variants C148S and C148A almost eliminated the catalytic efficiency (Figure 2D). Meanwhile, the mutants H41A and H41L were also incapable of cutting the substrate, which suggested the catalytic dyad H41−C148 in MERS-CoV 3CL^{Pro}.

Conserved Residue Pair (E−H) Utilize Steric Effect for 3CL^{Pro} Recognition of Glutamine Substrates. For substrate binding, the highly conserved native substrate P1 site residue attracted our attention (Figure S4). We previously developed inhibitors of EV71 3C protease (EV71 3C^{Pro}) and determined

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890



Protease	WT	E169L	H175L	E169L H175L
Space Volume (Å³)	0	0	0.921	0.537

Figure 3. Comprehensive investigation of the effect of the conserved Glu-His in MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. (A) Structural analysis of the S1 pocket from three proteases interacting with the inhibitor P1 site. The results are shown as a two-dimensional structure using the Maestro software. Hydrogen bond interactions are presented as purple arrows (inhibitor P1 site acts as a hydrogen bond donor) and red arrows (inhibitor P1 site act as a hydrogen bond acceptor). (B) Biochemical assay of the wild type and E169 mutant proteases for proteolytic activity. The data presented are the mean values from experiments in triplicate, and the error bars indicate the standard deviations. (C) Structural analysis of the surroundings of E169 in MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. The proteases are shown as cartoons (MERS-CoV 3CL^{Pro} in wheat color and SARS-CoV 3CL^{Pro} in white), and the inhibitor is presented as yellow sticks. (D) Catalytic activity analysis of the wild type and H175 mutant proteases. The data presented are the mean values from experiments in triplicate, and the error bars indicate the standard deviations. (E) POVME script calculation for the spare volume enveloping the substrate glutamine, 169 site residue, and 175 site residue. The average structure of the WT and mutant proteases following MD simulation are shown (the protease is presented as a cartoon, the key residue in the protease is shown as cyan sticks, and the substrate glutamine is shown as sticks) and the spare volume is displayed as a violet mesh. Meanwhile, the value of the redundant volume is shown at the bottom.

the complex structure of EV71 3C^{Pro} with a peptidomimetic inhibitor (PDB code: 5BPE).[50−53] By comparing the S1 pocket structures of EV71 3C^{Pro} and coronavirus 3CL^{Pro}, we noticed an extra hydrogen bond interaction between the conserved protease glutamine (E169 of MERS-CoV 3CL^{Pro} and E166 of SARS-CoV 3CL^{Pro}) and the lactam ring of the inhibitor, which presumably led to the enhanced inhibitor affinity (Figure 3A). Therefore, to verify this assumption, mutagenesis of MERS-CoV 3CL^{Pro} E169 was performed.

In principle, we began mutagenesis via replacement of E169 with Asp that also contains a carboxyl group with negative charge. Following evaluation, compared with the WT protease, the mutant E169D exhibited inferior catalytic efficiency, which suggested that the electrostatic force of E169 was too laborious to affect the substrate binding process (Figure 3B). Subsequently, enzymatic assays were conducted using purified variants with five neutral amino acid residues (Ala, Gln, Leu, Met, and Val) instead of Glu (Figure 3B). Comparable

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

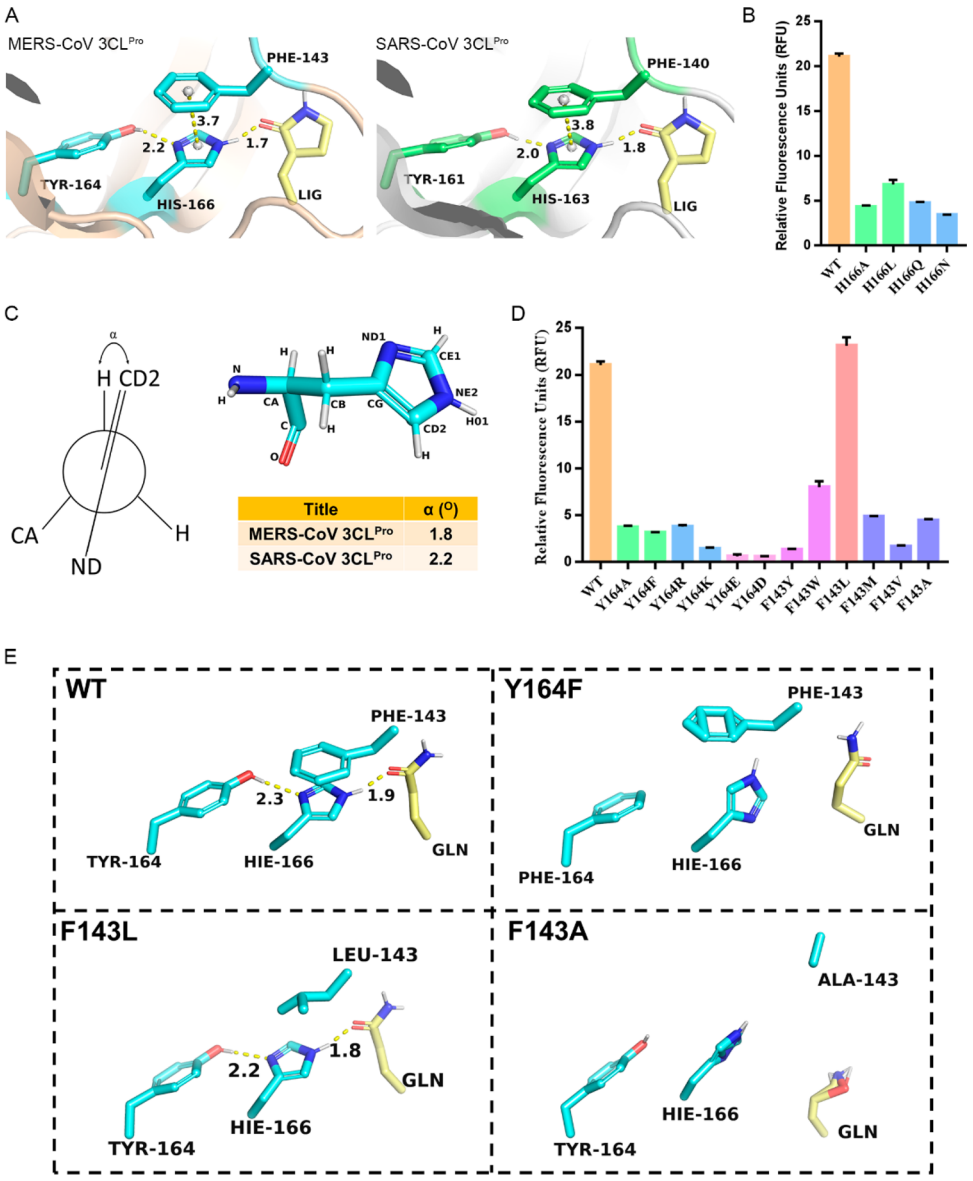
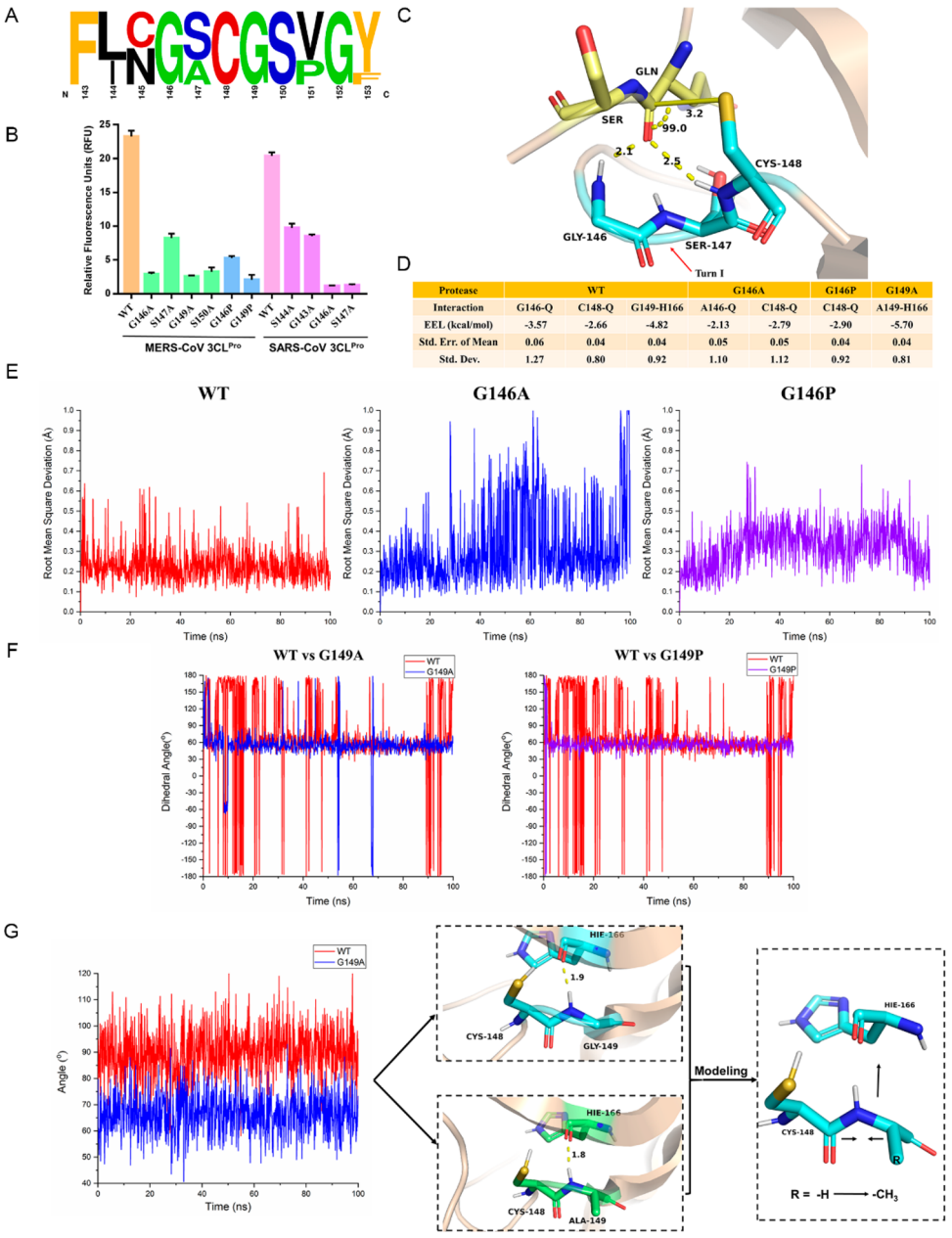


Figure 4. A stable histidine is responsible for the 3CL^{Pro} recognition stage for the substrate glutamine. (A) The structural analysis of the area surrounding the substrate glutamine and H166. (B) The catalytic activity evaluation of WT and H166 mutants. The data presented are the mean values from experiments in triplicate, and the error bars indicate the standard deviations. (C) The sketch map of the dihedral (CD2-CG-CB-H) from the conserved histidine (left). The general symbol for atom in the histidine and the value of the dihedral (right). (D) The proteolytic activity analysis of the Y164 and F143 mutants in MERS-CoV 3CL^{Pro}. The data presented are the mean values from experiments in triplicate and the error bars indicate the standard deviations. (E) The average structure of the WT and mutant (Y164F, F143L, and F143A) protease following MD simulation. In the structures, the protease is presented as a wheat colored cartoon, and the significant residue is shown as sticks (the protease residue is in cyan, and the substrate glutamine is in yellow).

enzymatic activity was observed for the E169L and E169Q mutants, which possessed side chains of similar bulk to Glu, while the E169A, E169 V, and E169 M mutants, which had side chains with various volumes at the 169th site, showed decreased catalytic activity compared with WT protease, which indicated that steric effects at E169 were required to fit the substrate (Figure 3B). However, the flexible side-chain of E169 meant that this amino acid alone could not achieve this effect.

When scanning the structure, conserved residue H175 was conventionally ignored owing to its weak interaction with the inhibitor. However, the inextricable interaction between H175 and E169 made it reasonable to speculate that H175 was indispensable for proteolytic bioactivity (Figure 3C). Subsequently, crippling the impregnable interactions by replacing H175 with Ala, Leu, Glu, and even Asn could remarkably reduce the enzymatic activity, though the mutation of H175 to Gln, which shared a similar hydrogen bond donor feature to His, could retain the protease bioactivity (Figure 3D). The results indicated that the hydrogen bond interaction between H175 and E169 was essential to the protease, which was in contrast to the previous E169L mutant result (Figure 3B). Then, a multipoint mutant (E169L/H175L) was created, and this mutant expressed inferior catalytic activity (Figure 3D). To explain these unexpected results, molecular dynamic (MD) simulations were applied to imitate the binding mode of the protease with the natural substrate, which was construed form the peptidomimetic aldehyde structure and optimized by MD simulation. As consequence, compared with WT, the distance

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890



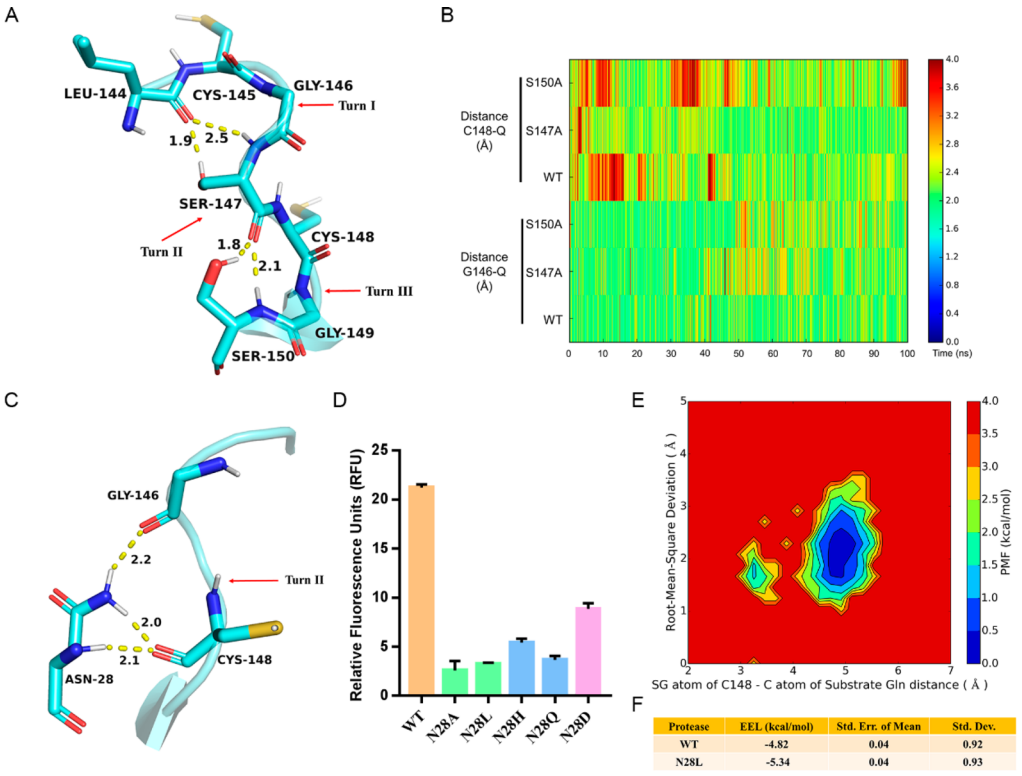
Protease	WT			G146A		G146P	G149A
Interaction	G146-Q	C148-Q	G149-H166	A146-Q	C148-Q	C148-Q	A149-H166
EEL (kcal/mol)	-3.57	-2.66	-4.82	-2.13	-2.79	-2.90	-5.70
Std. Err. of Mean	0.06	0.04	0.04	0.05	0.05	0.04	0.04
Std. Dev.	1.27	0.80	0.92	1.10	1.12	0.92	0.81

Figure 5. Comprehensive research on G146 and G149 in the conserved GSCGS motif. (A) Conservation analysis of six common coronavirus 3C-like proteases. (B) Catalytic activity assays of the WT and mutant proteases in MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (C) The explicit structure of the WT protease revealing that the protease GSCGS motif compactly interacts with the substrate active site after MD simulation. The protease is presented as a wheat colored cartoon and the significant residue is highlighted as sticks. (D) MM calculation of the significant interaction in WT and mutant proteases. (E) The time evolution of the substrate glutamine RMSD value in the WT and mutants. (F) The time evolution of the dihedral of C148 (SG-CB-CA-C) in WT protease (in red) and mutations (G149A in blue and G149P in violet). (G) The model to explain the underlying reason for the restraining effect of G149A toward the free rotation of the C148 thiol. The left part is the time evolution of the φ dihedral (148C-149N-149CA-149C) in the WT and G149A mutant proteases. Following mutagenesis of G149 to A149, the distinction in the φ dihedral between the WT and G149A mutant proteases led to more compact binding between C148-G149 and H166, which hampered the rotation of the C148 side chain in the G149A mutant (right).

between residues 169 and 175 was remarkably enlarged in the mutant H175L and mutant E169L/H175L, while the mutant E169L displayed the same distance as the WT protease (Figure S6), which hinted that redundant space may strikingly affect protease catalytic activity in the H175L and mutant E169L/H175L. Following POVME script calculation, the results demonstrated that the redundant space interfered with the fit between the protease and substrate, which indicated that E169 mainly interacted with the substrate through steric effects rather than electrostatic interactions (Figure 3E).

Stable Conserved Histidine Participates in the Recognition of Glutamine Substrates by 3CL^{Pro}. Due to the essential fitting model of E169-H175, it led us to question whether a residue with a suitable side chain volume might act similar role with Gln to fill the S1 pocket and start the cleavage event. However, coronavirus 3CL^{Pro} preferred Gln to His, which exhibits a similar side chain volume as Gln in the substrate P1

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890



Protease	EEL (kcal/mol)	Std. Err. of Mean	Std. Dev.
WT	-4.82	0.04	0.92
N28L	-5.34	0.04	0.93

Figure 6. Comprehensive research on special structure of the conserved GSCGS motif. (A) Typical structure of the WT MERS-CoV 3CL^{Pro} protease following MD simulation to reveal the consecutive three turns constituting the GSCGS motif. The protease is presented as a cartoon, and the significant residue is highlighted as sticks. (B) The time evolution of the distance (C148-substrate glutamine and G146-substrate glutamine) in the WT and mutants. (C) Typical structure of the WT MERS-CoV 3CL^{Pro} protease following MD simulation to reveal the formation of turn II in the GSCGS motif. (D) Catalytic activity assays of the WT and mutant proteases in MERS-CoV 3CL^{Pro}.The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (E) PMF calculation for the SG atom of C148-substrate glutamine carbonyl group distance vs the RMSD of total system in N28L MERS-CoV 3CL^{Pro}. (F) MM calculation of the interaction between G149 and H166 in WT and mutant N28L.

site, invalidating this assumption. To explore the underlying reason, the structures of the two proteases were scrutinized, and a conserved histidine (H166 in MERS-CoV 3CL^{Pro} and H164 in SARS-CoV 3CL^{Pro}) caught our attention, as it formed robust hydrogen bond interactions with the inhibitor (Figure 4A). In an effort to validate the effect of the robust hydrogen, mutagenesis study on H166 of MERS-CoV 3CL^{Pro} was performed. Mutants H166A and H166L could reduce the proteolytic activity, which clearly demonstrated the significance of the intense hydrogen bond (Figure 4B). As Gln and Asn have similar hydrogen bond donor and acceptor abilities as His, both of them were suitable replacements for histidine. But mutants H166N and H166Q had significantly decreased activity, implying that other characteristics might define the conserved histidine (Figure 4B). After examining the structures, due to the inappropriate dihedral (CD2-CG-CB-H) presented by H166, conserved H166 exhibited an eclipsed conformation, leading to the torsional tension and a rotating trend (Figure 4C). Thus, there were restrictive ingredients that limited the rotation of the H166 side chain and maintained H166 in its energized conformation, which made H166 interact with the substrate Gln in an appropriate angle and distance.

By scrutinizing the two structures, conserved Y164 was detected as exploiting its phenolic hydroxyl group to interact with H166 (Figure 4A). In theory, the acidic phenolic hydroxyl group of Y164 can provide a proton or form a stronger hydrogen bond with H166. To explore the significance and acting mode of Y164, we chose mutants Y164A, Y164F, Y164R, Y164 K, Y164E, and Y164D, considering electrical and steric factors. Compared with WT, all of the variants showed depressing catalytic activity. This result suggested that the phenolic hydroxyl group of Y164 interacted with H166 via the hydrogen bond to restrain H166 in a suitable orientation to promote preferable substrate recognition (Figure 4D). To understand the role of Y164, MD simulations were performed, and the results showed that replacement of Y164 led to the free rotation of H166 and leaving of the substrate Gln from the S1 pocket (Figures 4E and S7).

Meanwhile, F143 is located near H166 and appears to interact with H166 via $\pi-\pi$ stacking interactions to hamper the rotation of H166 (Figure 4A). Nevertheless, two variants containing an aromatic ring, F143Y and F143W, exhibited significantly reduced activities, indicating that the $\pi-\pi$ stacking may be not the major factor. In the case of variants F143L, F143M, F143 V, and F143A, F143L unexpectedly displayed excellent performance equal to WT (Figure 4D). Compared with that of WT and F143L mutant, MD simulation results of the mutant F143A revealed a deficiency in steric restriction, leading to an unstable connection between H166 and Y164 and the substrate glutamine (Figures 4E, S8, and S9). Therefore, F143 was proven to employ steric effects to restrain the rotation of H166. Similar results involving the S1 pocket were exhibited in SARS-CoV 3CL^{Pro} during the substrate binding stage (Figure S10).

Special GSCGS Motif Plays a Significant Role in the Start of Catalysis. With regard to the catalysis, it was inevitable to explore the sequence nearby the active Cys. On the basis of

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

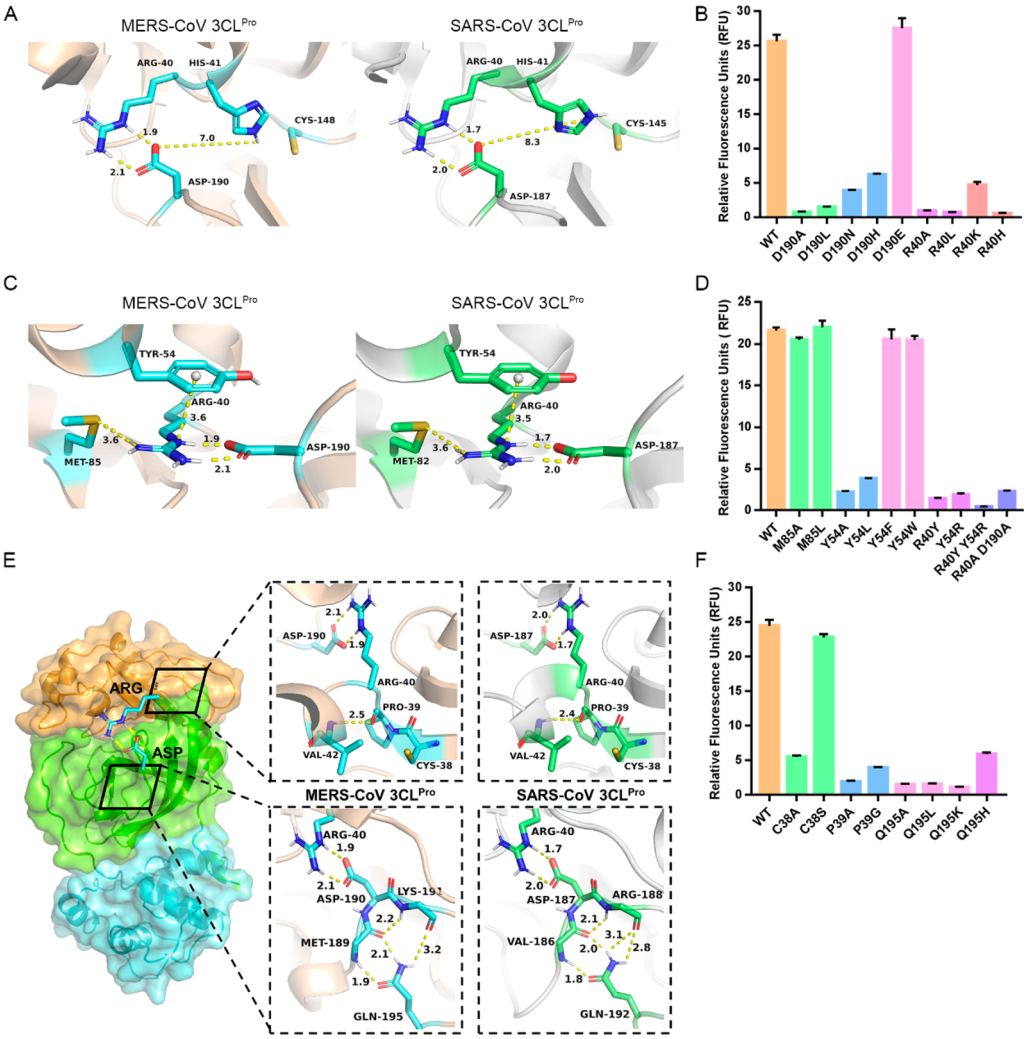


Figure 7. Importance of the partial negative charge cluster (PNCC) constituted by Arg-Tyr-Asp in MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. (A) The structural analysis of the area surrounding the catalytic center in MERS-CoV 3CL^{Pro} (left) and SARS-CoV 3CL^{Pro} (right). (B) Catalytic activity assay of the D190 and R40 mutants in MERS-CoV 3CL^{Pro}. The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (C) The structural analysis of the area surrounding the key Arg-Asp seeking an additional residue participating in the PNCC. (D) Catalytic activity analysis on the Y54 mutants, M85 mutants, and multipoint mutants in MERS-CoV 3CL^{Pro}. The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (E) Structural scrutiny to understand on the typical structural characteristics for maintaining the compact connection between conserved Arg and Asp in MERS-CoV 3CL^{Pro} (left) and SARS-CoV 3CL^{Pro} (right). (F) Catalytic activity evaluation of the C38, P39, and Q195 mutants in MERS-CoV 3CL^{Pro}. The data present the average results from three independent experiments, and the error bars indicate the standard deviations.

the alignment of the common coronavirus 3CL^{Pro} sequence, a conserved GSCGS sequence caught our attention (Figure 5A). As a familiar protein linker, the GS sequence was frequently applied for protein engineering. However, the GS sequence, especially a double GS sequence, infrequently emerges near the viral cysteine protease catalytic center. Therefore, to verify the significance of the motif, the residues in the GSCGS motif of MERS-CoV 3CL^{Pro} were sequentially replaced with Ala and all of the mutants represented reduced activities, which manifested that the GSCGS motif was essential for protease activity (Figures 5B and S11). Further exploring the structure of the protease, the GSCGS motif exhibited three consecutive turns via obvious hydrogen bonding (Figure 6A). As Pro is an appropriate residue to form a turn in the protein structure, the mutation of Gly to Pro was performed. The decreased bioactivity of variants G146P and G149P indicated the irreplaceability of the GSCGS motif and the significant role of Gly in the GSCGS motif (Figure 5B).

For investigating the effect of G146 in the motif, MD simulations were performed, and the results suggested that G146 and C148 could form firm hydrogen bonds with the substrate active site in the WT protease conformation (Figure 5C). However, the conformation of the G146A and G146P mutants following MD simulation demonstrated that the 146th residue failed to have an optimal distance or interactive angle, and even a deficiency in hydrogen bonding interaction, to bind the substrate active site (Figures 5D, S12, and S13). As a consequence, compared with WT protease, a modest swing of the substrate Gln residue emerged in the G146A and G146P mutants. This result indicated that G146 could fix the carbonyl group at the substrate active site, decrease the swing at the active site, and facilitate the assault of C148 (Figure 5E).

With analysis on the MD simulation results of variants G149A and G149P, the dihedral constituted from SG-CB-CA-C of C148 presented extreme stability compared with the WT protease, which suggested that the rotation frequency of the

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

C148 side chain had been remarkably decreased (Figure 5F). In parallel, the PMF calculation result for WT protease manifested two low energy conformations appearing in the trajectory of the protease. However, the results of variants G149A and G149P suggested that only one low energy conformation, where the thiol was far from the active site of the substrate, emerged in the motion of the protease (Figure S14) and manifested the rotation of C148 side chain relied on G149. To unveil the underlying reason, the rotation of the C148 thiol in the WT protease was further investigated and two typical states (resting state and attacking state) were noticed in the MD simulations (Figure S15). In the resting state, the C148 thiol preferred to deviate from the substrate and interact with H166. Conversely, the C148 thiol presented a suitable conformation, close to the substrate in the attacking state. By superimposing the attacking and resting state conformations, the C148-G149 main chain exhibited obvious deflection, while the dihedral φ and ψ angles of C148 were stable (Figures S16 and S17). This result indicated that the translation of the C148-G149 main chain other than rotation might contribute to the rotation of the C148 thiol. In the G149A mutant, the smaller A149 φ dihedral led to a compact conformation of A149 and enhanced the rigid hydrogen bond between A149 and H166, which indicated that the translation of the C148-A149 main chain would demand more energy to break the restriction of H166 to A149 in the G149A mutant (Figure 5D and 5G). Additionally, due to the deficiency of the hydrogen bond interaction in G149P, the C148 thiol formed a more robust interaction with H166 in compensation, interfering with the rotation of the C148 thiol (Figure S18).

Special Three Consecutive Turns in GSCGS Motif. To smoothly execute the task discussed above, the GSCGS motif must maintain three special and consecutive turns. Turns I and III are guaranteed via two interior hydrogen bonds of the motif mediated by S147 and S150 (Figure 6A). Mutagenesis of Ser to Ala would seriously enlarge the distance (G146 carbonyl at the active site and C148 carbonyl at the active site) and hamper the anchoring effect of G146 and C148 toward the active site in the MD simulation (Figure 6B). Interestingly, considering that turn III is dominated by a hydrogen bond between S150 and S147, the substation of S150 with Ala would interfere not only with turn III but also turn I. This result explained the more drastic effect of variant S150A than that of S147A on the distance.

Alternatively, turn II is maintained by an exterior hydrogen bond formed by N28 (Figure 6C). Mutagenesis of N28 to Ala, Leu, Asp, and Gln, as well as His, resulted in the reduced activities of the protease (Figure 6D). To investigate the concrete effect of N28, MD simulations were performed on the mutant N28L. As a consequence, similar to that in the mutant G149A, the free rotation of the C148 thiol was restricted in mutant N28L, which further proved that the mutagenesis of N28 would deprive the pulling force of N28 to the carbonyl group of C148 and lead to more robust hydrogen bond interaction between the main chain of C148-G149 and H166. Hence, dominated by N28, the formation of turn II contributed to the rotation of the C148 thiol (Figure 6E and 6F).

Partial Negative Charge Cluster Constituted of Asp190, Arg40, and Tyr54 Exhibited Indispensable Functions during the Catalytic Process. With the development of the catalysis, bearing an analogous chymotrypsin-like structure of $3CL^{Pro}$, the EV71 3C protease cuts the polyprotein precursors with its catalytic triad (Glu, His, and Cys) (Figure S19). Nevertheless, there is a catalytic dyad (His and Cys) in coronavirus $3CL^{Pro}$ due to the Glu deficiency. After aligning the six coronavirus $3CL^{Pro}$ sequences (Figure S20), the conserved acidic amino acid D190 caught our attention. Interestingly, D190 is located almost 6.7 Å away from H41, which seemed problematic for promoting firm interaction between D190 and the catalytic dyad due to the long distance (Figure 7A). To discover the significance of D190, D190 was mutated to Leu, Ala, His, and Asn. As a result, the D190A, D190L, D190H, and D190N mutants failed to effectively cleave the substrate, while the D190E mutant retained protease catalytic activity, which implied that the acidic residue was indispensable, despite the far distance from the catalytic dyad (Figure 7B).

In addition to D190, R40 formed a strong electric interaction with D190 in the coronavirus $3CL^{Pro}$, which might neutralize the negative charge of D190 (Figure 7A). Following mutagenesis of R40, the catalytic activity of mutants exhibited a remarkable decline (Figure 7B). Interestingly, when R40 was mutated to an aliphatic residue (Leu or Ala), the density of the D190 negative charge would be enhanced and the proteolytic activity might in theory be improved. However, the mutants R40L and R40A manifested decreased proteolytic activity. Therefore, a partial negative charge cluster constituted from R40-D190 was supposed to play a significant role in the catalysis activity. Meanwhile, the value of the positive charge of the Arg side chain was usually superior to the value of the negative charge manifested by Asp side chain in physical condition. This implied the existence of an extra residue that could alleviate the positive charge of R40 for maintaining the partial negative charge of R40-D190. Investigation of the structures of the two proteases showed that Y54 and M85 were located near R40 and might interact with R40 via a π-cation interaction and an electron atmosphere−cation interaction, respectively (Figure 7C). Following mutagenesis of Y54 and M85, the results suggested that Y54, rather than M85, could interact with R40 via the π−cation interaction, which potentially supported that Y54 could interact with R40 via the π-cation interaction to alleviate the electrical interference of R40 to D190 and further caused D190 to present partial negative charge (Figure 7D). Additionally, multipoint mutations were introduced, and the loss of catalytic activities of the variants reiterated the importance of the partial negative charge cluster (PNCC) constituted by D190-R40-Y54 (Figure 7D).

Since the D190-R40-Y54 synergistically took effect on the catalysis, the interior close connection of PNCC required a special structure for support. Investigation of the protease sequence near R40 identified a remarkably conserved sequence containing C38 and P39 (Figure S21). There was a typical β-turn formed by P39−V42 hydrogen bond interactions, which led to R40 protruding from the loop and potentially enhanced the interior interaction of PNCC (Figure 7E). However, the deficiency of apparent interaction involving C38 suggested that the C38 side chain volume might be responsible for shaping the remarkably conserved short loop. To confirm this hypothesis, we mutated C38 and P39, and the results verified the assumption (Figure 7F).

From the perspective of D190, an atypical turn was formed, and conserved residue Q195 interacted with the main chain carbonyl group of M189 and K191, which caused a protrusion of D190 from the loop and promoted the R40-D190 interaction (Figures S21 and 7E). The decreased catalytic activities of the mutagenesis proved the significance of Q195 (Figure 7F).

Fixed by Gln, a Conserved Water Mediates the Remote Interaction between PNCC and the Dyad. Owing to the deficiency of intuitive binding, remote interactions

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

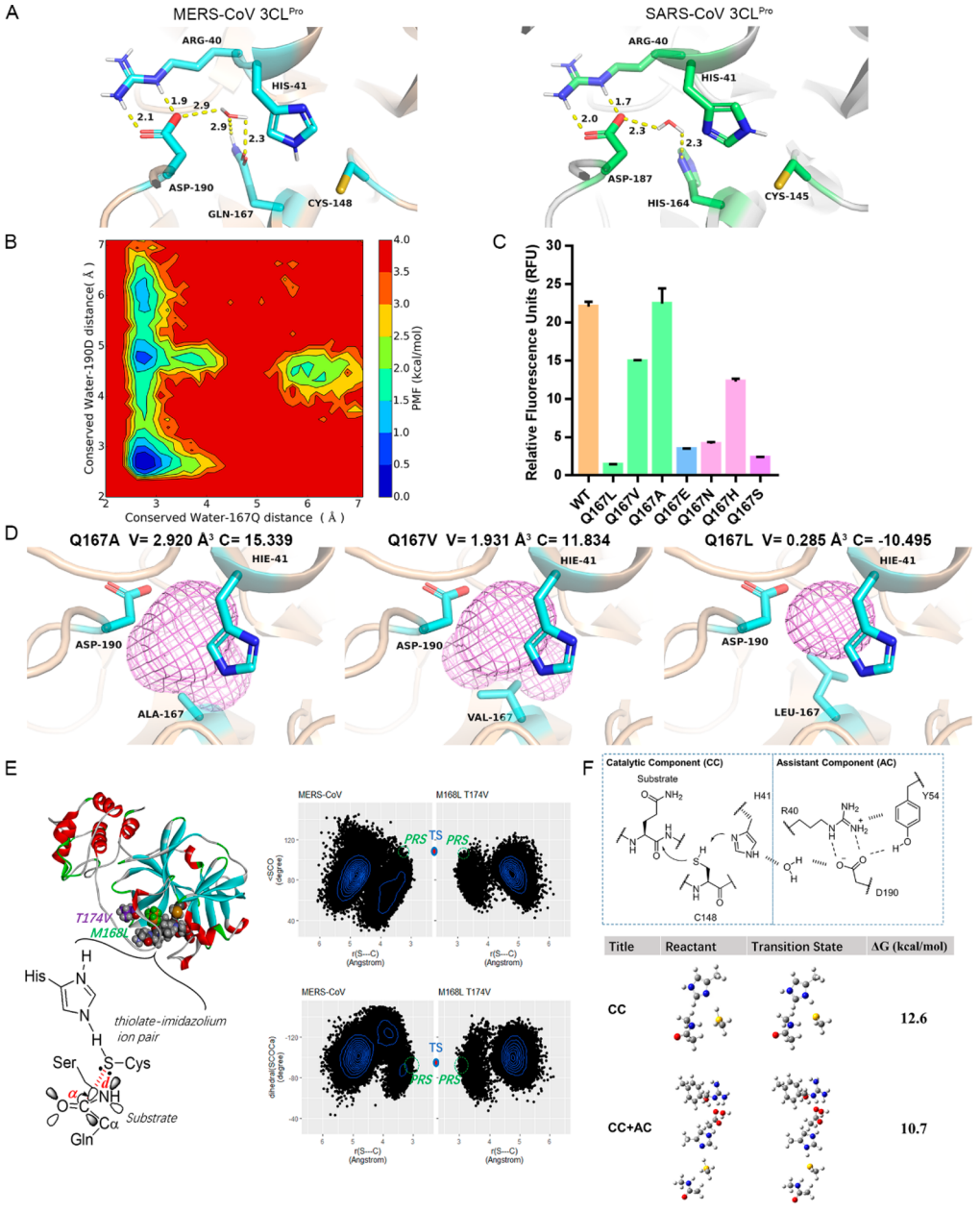
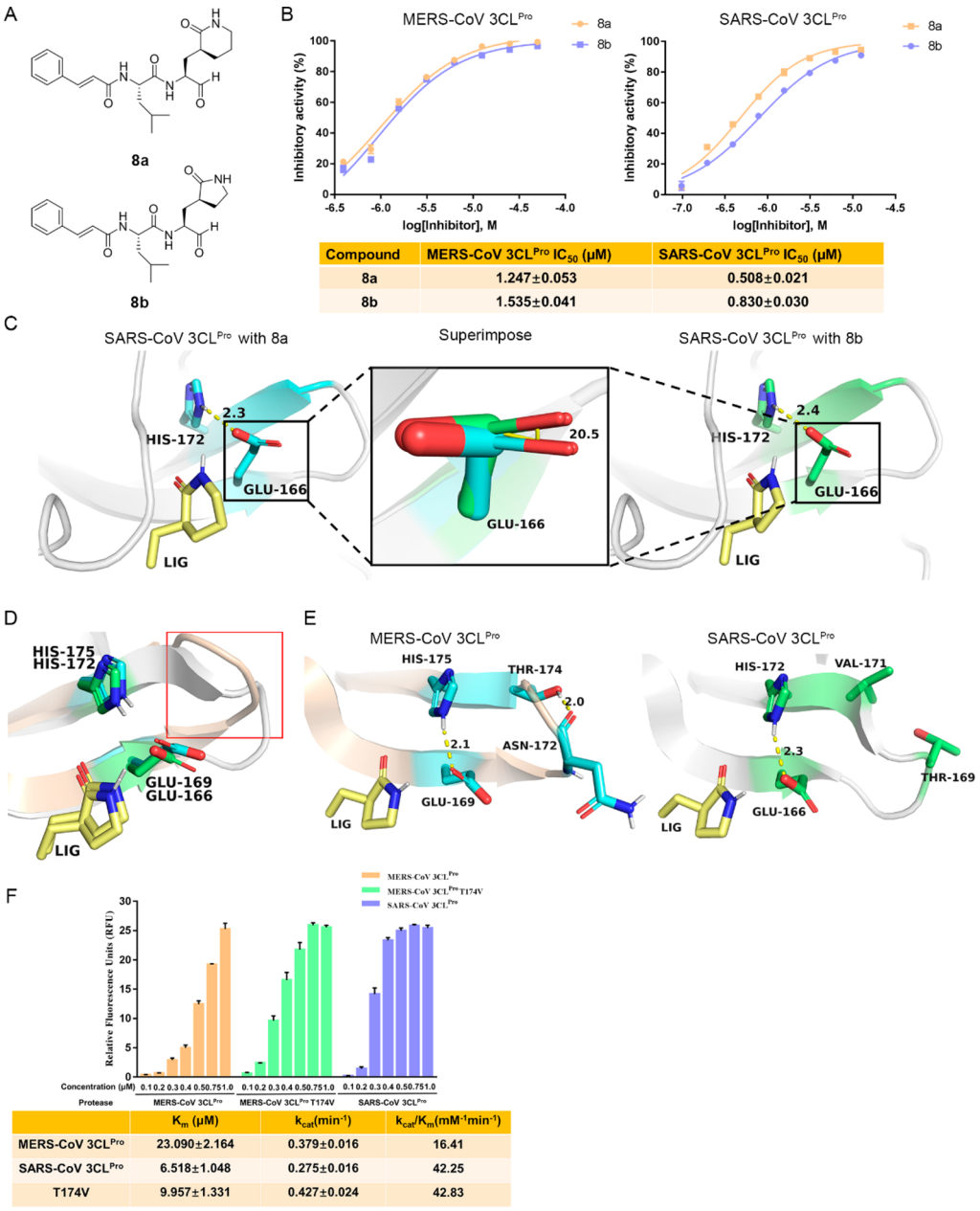


Figure 8. Indispensable roles of the remote interaction mediated by conserved water in catalysis. (A) Structural scrutiny of the space between H41 and the PNCC in MERS-CoV 3CL^{Pro} (left) and SARS-CoV 3CL^{Pro} (right). (B) PMF calculation for the conserved water−Q167 distance vs conserved water−D190 distance. (C) Catalytic activity assays of the Q167 mutants in MERS-CoV 3CL^{Pro}. The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (D) POVME calculation of the space volume enveloping 41H, Q167, M168, and D190 in the average structure of Q167A (left), Q167 V (center), and Q167L (right) following MD simulation. The proteases are shown as wheat colored cartoons, the key residues are presented as sticks and the redundant volume is exhibited as a violet mesh. The calculated space volume value (refer to the symbol V) and the contribution to the polar surface area of residue (refer to the symbol C) are shown above. (E) The PRS analysis for the WT and M168VT174 V mutant of MERS-CoV 3CL proteases. (F) Optimized structures and free energy barriers for the first step of catalysis stage.

are generally neglected. In coronavirus 3CL^{Pro}, the location of PNCC is approximately 6−7 Å from the catalytic dyad, which makes it difficult to assume powerful interactions between the dyad and PNCC. However, the investigation of the structures of the two proteases identified a conserved water, which might mediate the remote interaction between these two essential components (Figure 8A). To verify the necessity of the conserved water, MD simulations were employed to imitate the motion of the conserved water that was extracted from its original location in MERS-CoV 3CL^{Pro}. Following PMF calculation, the lower system energy demand compelled the water to penetrate into the protein and locate itself at a suitable site, which proved the essentiality of conserved water (Figure 8B).

Meanwhile, the conserved water was noticed to interact with a key residue (Q167 in MERS-CoV 3CL^{Pro} and H164 in SARS-CoV 3CL^{Pro}), which was considered to hold the conserved water in an appropriate location (Figure 8A). Thus, the Q167 in MERS-CoV 3CL^{Pro} was mutated to aliphatic residues (Figure 8C). As a consequence, biochemical results and POVME

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890



Compound	MERS-CoV 3CL^{Pro} IC_{50} (μM)	SARS-CoV 3CL^{Pro} IC_{50} (μM)
8a	1.247±0.053	0.508±0.021
8b	1.535±0.041	0.830±0.030

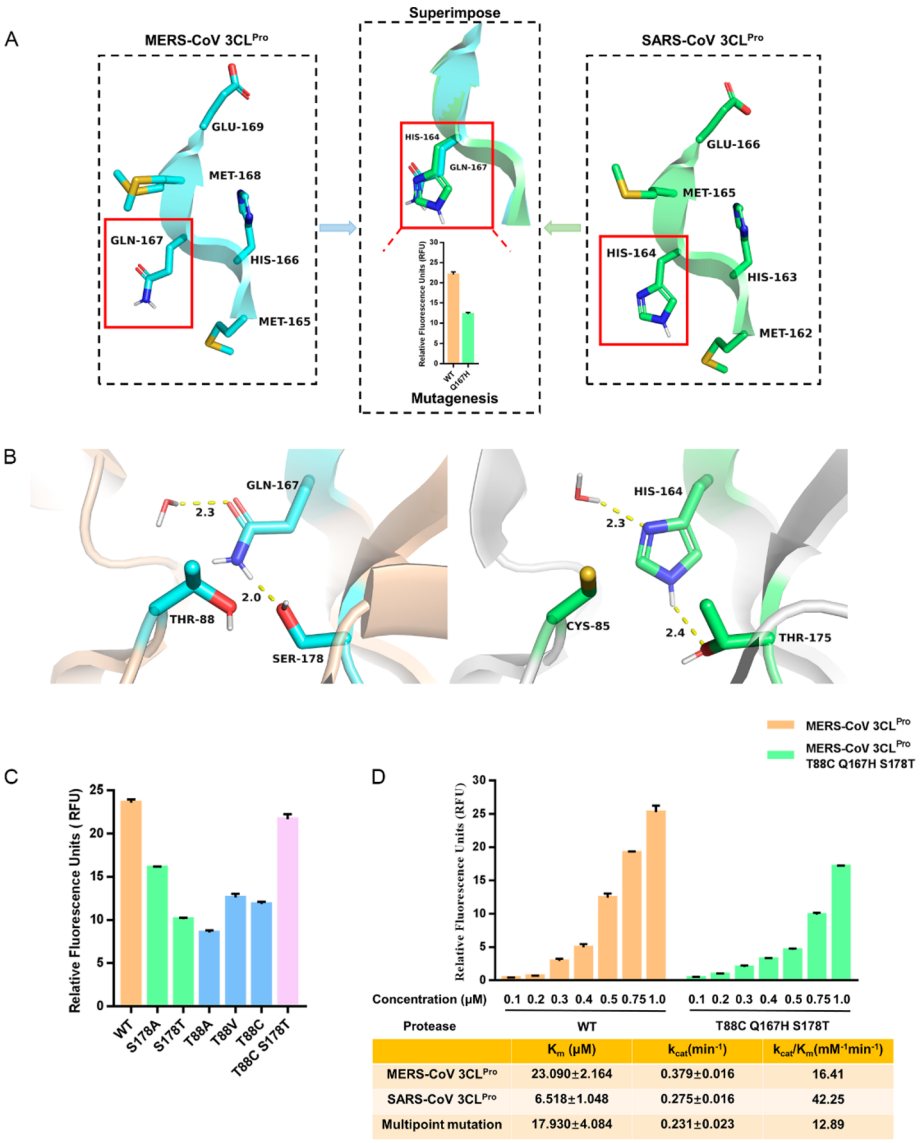
	K_m (μM)	k_{cat}(min^{-1})	k_{cat}/K_m(mM^{-1}min^{-1})
MERS-CoV 3CL^{Pro}	23.090±2.164	0.379±0.016	16.41
SARS-CoV 3CL^{Pro}	6.518±1.048	0.275±0.016	42.25
T174V	9.957±1.331	0.427±0.024	42.83

Figure 9. A significant distinction between MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} during recognition. (A) The structure of aldehydes **8a** and **8b**. (B) Inhibitory activity evaluation of the two aldehyde inhibitors against MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. The IC_{50} values were calculated and are presented in the table at the bottom. The data presented are the mean values from experiments in triplicate, and the error bars indicate the standard deviations. (C) The structural analysis from the binding model of inhibitor **8a** (left) and **8b** (right) in the cocrystal structure. The superimposed structure is applied at the center to indicate the major distinction of the two inhibitors' binding modes with respect to E166. (D) Superimposition of the MERS-CoV 3CL^{Pro} (in white) and SARS-CoV 3CL^{Pro} (in wheat color) structures. The main structural distinction of two proteases in the S1 pocket is highlighted with a black frame. (E) The detailed distinction in the highlighted secondary structure between SARS-CoV 3CL^{Pro} (displayed in gray cartoon and green sticks) and MERS-CoV 3CL^{Pro} (displayed in wheat colored cartoon and cyan sticks). (F) Catalytic activity analysis and enzyme kinetic parameters of the T174 V mutant. The data present the average results from three independent experiments, and the error bars indicate the standard deviations. The significant residue is shown as sticks (MERS-CoV 3CL^{Pro} residue in cyan, SARS-CoV 3CL^{Pro} residue in green, and the substrate glutamine in yellow).

calculation results suggested that the volume and hydrophilicity of the pocket created by H41, Q167, M168, and D190 were critical to the catalytic activity of protease (Figure 8D). With further investigation, the lower catalytic activity emerged in the variants (Q167S, Q167N, and Q167H) with hydrophilic side chains at 167th residue, which additionally proved that the long hydrophilic side chain of Q167 anchored the conserved water in the suitable location. These results verified the significance of the conserved water. Furthermore, the mutation of Q167 to Glu, with an additional negative center, would decrease the enzymatic activity and further verified the importance of the partial negative charge of PNCC (Figure 8C). A highly similar result was observed for SARS-CoV 3CL^{Pro} (Figure S22).

As discussed above, to further explore the concrete effects of the remote interaction composed of PNCC and the conserved water, Prereaction State (PRS) analysis and quantitative

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890



	K_m (μM)	k_{cat}(min^{-1})	k_{cat}/K_m(mM^{-1}min^{-1})
MERS-CoV 3CL^{Pro}	23.090±2.164	0.379±0.016	16.41
SARS-CoV 3CL^{Pro}	6.518±1.048	0.275±0.016	42.25
Multipoint mutation	17.930±4.084	0.231±0.023	12.89

Figure 10. Distinction between MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} during the catalysis stage. (A) The distinct structure to fix the conserved water between MERS-CoV 3CL^{Pro} (left) and SARS-CoV 3CL^{Pro} (right), and the catalytic activity analysis of the Q167H mutant in MERS-CoV 3CL^{Pro}. (B) Structural scrutiny of the area surrounding Q167 in MERS-CoV 3CL^{Pro} (left) and H164 in SARS-CoV 3CL^{Pro} (right). (C) Catalytic activity assays of the T88 and S178 mutants in MERS-CoV 3CL^{Pro}. The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (D). Catalytic activity assays of the T88C, Q167H, and S178T individual and multipoint mutants from MERS-CoV 3CL^{Pro} in a dose-dependent manner (upper figure) and the kinetic parameters for multipoint mutant catalysis (table below). The data present the average results from three independent experiments and the error bars indicate the standard deviations.

calculations were performed. According to the previous QM/MM calculation,[39,40] the first step of MERS-CoV 3CL^{Pro}-catalyzed reaction was believed to be the nucleophilic attack of the cysteine residue, and the resulting thioester would be rapidly hydrolyzed by a water molecule under the general base catalysis of His41. Figure 8E showed the accessibility of the Gln-Ser peptide bond of the substrate to the Cys148 thiolate of the MERS-CoV 3CL^{Pro}. It is highly essential for the nucleophilic thiolate and the acceptor carbonyl group to align up in the sp2-sp3 conversion to maximize the orbital interaction between the peptide π^* (LUMO) and sulfur lone pair (HOMO) in the nucleophilic addition.[54] Using the Burgi−Dunitz criteria of near-attach-conformation parameters (the S---C distance <3.5 Å and attacking angle alpha (100−110°) and SCOCa dihedral (85−95°)), the calculated PRS indicated a significant population belong to the resting state, where the S−C distance is as long as ∼4.8 Å. It is likely that the 3CL^{Pro} is yet poorly evolutionary for this substrate. Therefore, we chose an attacking state conformation on the basis of the criteria described at methods and performed ab initio calculation. Compared with the state excluding the remote interaction, the integral state presented a lower catalytic energy barrier, which suggested that the existence of the remote interaction could extremely accelerate the first step of the nucleophilic reaction (Figure 8E). Taken together with the partial negative charge of PNCC, it was considered to temporarily stabilize the protonated histidine with conserved water mediating rather than fastening the protonated H41, therefore interfering with the subsequent catalytic process.

Distinction of Catalytic Efficiency between MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} during Substrate Binding Stage. According to the investigations described above, the

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

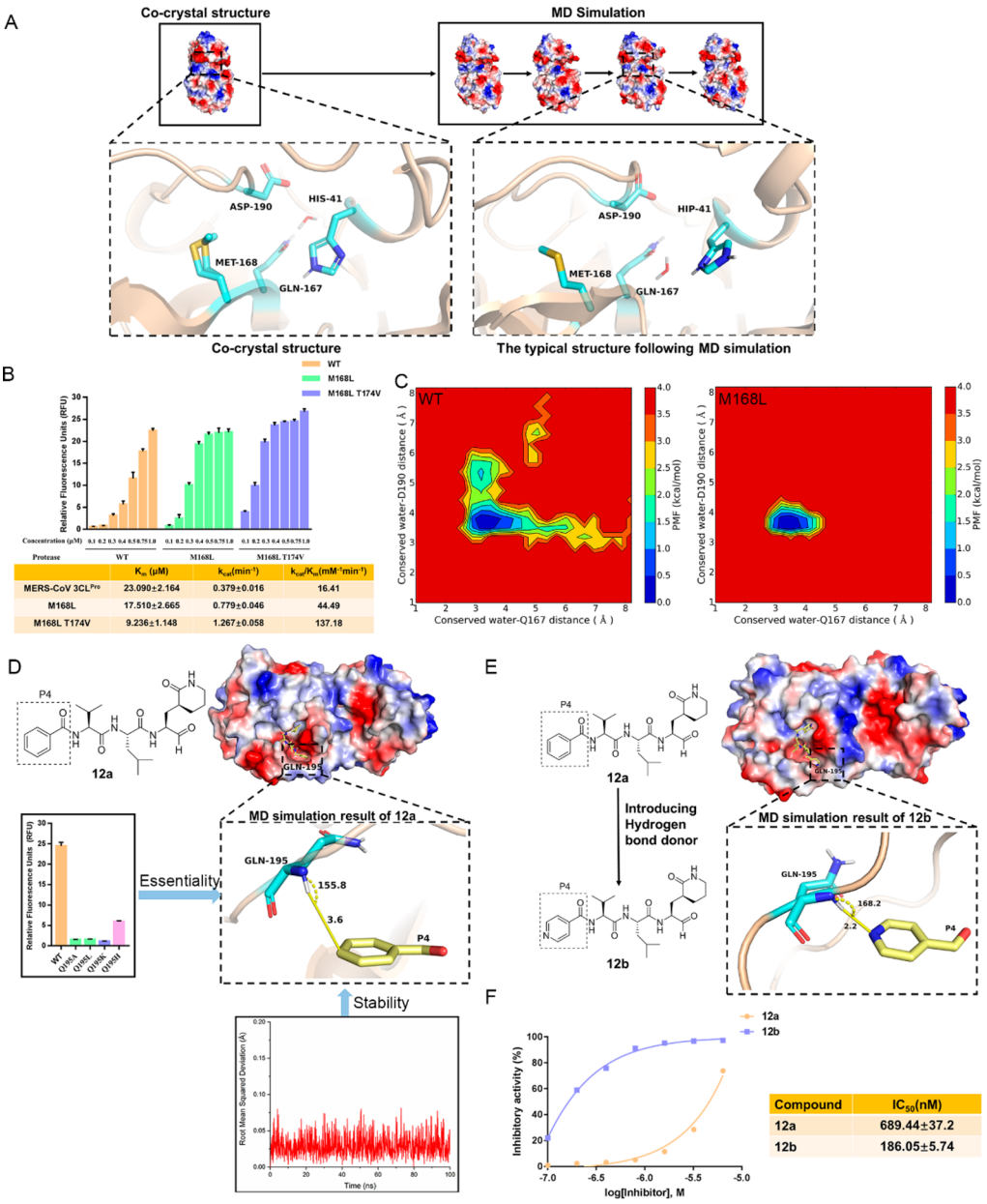


Figure 11. Two effective applications based on the investigated mechanism. (A) Structural analysis of the area surrounding of the conserved water in the cocrystal structure of MERS-CoV 3CL^{Pro} (left) and the typical structure following MD simulation (right). (B) Catalytic activity assays of the M168L mutant in a dose-dependent manner (upper figure) and the kinetic parameters for the mutant during catalysis (table below). The data present the average results from three independent experiments, and the error bars indicate the standard deviations. (C) PMF calculation for the conserved water—Q167 distance vs conserved water—D190 distance in WT (left) and M168L mutant (right) proteases. (D) The target to optimize the inhibitor. The importance and the stability of Q195 was reflected by the stable RMSD value, supporting that it become the target to optimize the inhibitor. Meanwhile, the typical structure of MERS-CoV 3CL^{Pro} binding with compound **12a** following MD simulation showed that the para-position of the aldehyde **12a** P4 site was a potential site to be optimized. (E) The strategy to optimize the aldehyde **12a**. Introducing a hydrogen bond acceptor to generate aldehyde **12b** might promote superior affinity toward MERS-CoV 3CL^{Pro} through MD simulation, and the average structure of MERS-CoV 3CL^{Pro} binding to **12b** following MD simulation is exhibited. (F) The inhibitory activity investigation for **12a** and **12b**. The data present the average results from three independent experiments, and the error bars indicate the standard deviations.

comprehensive catalytic mechanism of coronavirus 3CL^{Pro} was proposed. Insights into the catalytic mechanism would contribute to penetrate into the individuality of coronavirus 3CL^{Pro} and decipher the distinction of the catalytic efficiency between MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}.

When designing inhibitors against 3CL^{Pro}, cyclization of the glutamine side-chain to a (S)-γ-lactam and (S)-δ-lactam was an efficient application of the configuration restriction strategy. Therefore, peptidomimetic aldehydes **8a** and **8b** were synthesized, and their inhibitory activities were evaluated (Figures 9A and 9B). Compound **8a** showed more preferable inhibitory activities than **8b** against the two proteases. We determined two cocrystal structures of each compound bound to SARS-CoV 3CL^{Pro} (PDB code: 6LNY, 6LO0; see the detailed information referred to in Table 2). Analysis of the two structures detected a slight deflection of E166, which demonstrated the obvious extrusion force existed between the lactam ring and a rigid β-sheet segment (consisting of G170,

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

V171, and H172) and resulted in the enhancement of the fitting interaction between the protease and **8a** (Figure 9C). Interestingly, a similar section presented a turn dominated by the hydrogen bond (N172-T174) and a loose flexible loop in MERS-CoV 3CL^{Pro} (Figure 9D).

In addition, SARS-CoV 3CL^{Pro} displayed more efficient substrate affinity than MERS-CoV 3CL^{Pro} (lower K_m value) (Figure 2C). Therefore, we hypothesized that the presence of the turn might transform the secondary structure of the segment and cause the catalytic distinction of the two proteases. The mutant T174 V mutation could increase catalytic activity, which suggested that T174 might dominate the secondary structure of the segment and lead to the significant catalytic distinction of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} (Figures 9E and 9F).

Distinction between MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} during the Catalysis. In the catalysis, H164 of SARS-CoV 3CL^{Pro} was noticed to replace the corresponding Q167 in MERS-CoV 3CL^{Pro}. However, the Q167H variant in MERS-CoV 3CL^{Pro} exhibited inferior enzymatic activity (Figure 10A). It was first considered that the amino acid residues surrounding the Q167 for MERS-CoV 3CL^{Pro} and H164 for SARS-CoV 3CL^{Pro} might determine the selectivity and further interfere with protease activity. In MERS-CoV 3CL^{Pro}, the S178 side chain acted as a hydrogen bond acceptor to fix Q167, while T88 might be responsible for maintaining the suitable space volume among T88, Q167, and S178 (Figure 10B). With the change of Q167 in MERS-CoV 3CL^{Pro} to H164 in SARS-CoV 3CL^{Pro}, the T175 side chain acted as a hydrogen bond acceptor to fix H164. In addition, T175 and C85 were responsible to maintain the appropriate space volume among C85, H164, and T175 (Figure 10B).

To verify the analysis on the environment of the residues for anchoring the conserved water, mutations of T88 and S178 in MERS-CoV 3CL^{Pro} were performed. The enzymatic results showed that the single point mutation resulted in decreased protease activities, but the multipoint mutation S178T/T88C almost remained in the catalytic activity (Figure 10C). Meanwhile, the MD simulations results indicated the violent swing of Q167 emerged in mutants S178A and T88A, while stable Q167 appeared in mutant S178T/T88C (Figures S23 and S24). These results verified the similar environment of MERS-CoV 3CL^{Pro} Q167 and SARS-CoV 3CL^{Pro} H164, and suggested that the key residue to anchor the conserved water rather than the environment might be the main ingredient to cause the distinction of two proteases. Therefore, on the basis of the elimination of environmental impacts, a multipoint mutant (T88C/Q167H/S178T) of MERS-CoV 3CL^{Pro} was created. Compared with WT MERS-CoV 3CL^{Pro}, the decreased proteolytic ability and remarkably decreased k_{cat} value of the multipoint mutation T88C/Q167H/S178T indicated that the residue, anchoring the conserved water, allowed for the distinction between the catalytic capabilities of the two proteases in the cleavage stage (Figure 10D).

Two Effective Strategies for Enzyme and Drug Design Based on the Catalytic Mechanism. The high specificity of 3CL^{pro} prompts us to improve the catalytic efficiency for engineering enzyme like Tobacco Etch Virus protease (TEV-P) in intracellular regulation and Trypsin in protein sequence analysis.[55,56] In addition, the significance of 3CL^{pro} prompts us to develop effective antiviral agents. Therefore, the profound investigation of the catalytic mechanism contributes to enzyme design for expanding enzyme applications and antiviral drug design.

For the MERS-CoV 3CL^{Pro}, as a result of the conserved water being located at the zone enveloped by H41, Q167, M168, and D190 to mediate the remote interaction, restricting the motion of conserved water might be helpful to maintain the remote interaction between the dyad and PNCC. Moreover, in the MD simulations on the catalytic process (H41 formed protonated histidine-HIP and C148 became thiolate following an abstract proton processing of H41), a typical conformation of the conserved water approaching M168 was captured (Figure 11A). Therefore, mutagenesis of M168 to hydrophobic Leu was investigated to impede the conserved water from approaching L168 and maintain the tight interaction between PNCC and protonated histidine. The enhanced enzymatic activity proved the validity of the mutagenesis strategy, and the underlying mechanism was shown to restrict the motion of the conserved water in compact area via MD simulation, which further proved the temporary stabilizing effect of PNCC via the conserved water (Figures 11B, 11C, and S25). A highly similar mechanism was also proven for SARS-CoV 3CL^{Pro} (Figure S26).

In combination with the previous study on the distinction of two protease in recognition stage, multipoint mutagenesis was performed. The enzymatic results suggested the multipoint mutation M168L/T174 V of MERS-CoV 3CL^{Pro} exhibited an 8-fold increase in catalytic efficiency compared to WT protease. On the basis of the PRS analysis, it was computationally observed that the Cys148 SG atom in the highest-efficient mutant (M168L/T174 V) points to the carbonyl C atom statistically and dynamically better than those in the WT protease. On the basis of the Burgi−Dunitz criteria of near-attach-conformation parameters, the calculated PRS ratio was about 1:1.8 between the WT and mutant, which was conclusive that the distal mutations at the 168th and 174th sites can promote the reactive population of PRS via synergetic effect. Meanwhile, the PRS analysis on the multiple mutant (M168L/T174 V) reiterate the validity of molecular mechanism investigation (Figure 8E).

Insights into the catalytic mechanism are meaningful to guide inhibitor design. According to the explicit investigation, essential Q195 was located at the surface of the substrate binding pocket and presented extreme stability in the MD simulations. This implied that Q195 could become a significant target to optimize inhibitors (Figure 11D). In a previous investigation, **12a** was designed to inhibit MERS-CoV 3CL^{Pro} and expressed a specific inhibitory activity. In the study focused on the binding model of **12a**, MD simulation results showed that the P4 site of **12a** approached the Q195. Therefore, a pyridine group as a hydrogen acceptor was introduced into the inhibitor at P4 to generate **12b**, which might smoothly interact with Q195. The results of enzymatic analysis and MD simulations revealed that **12b** exhibited an almost 4-fold increase in inhibitory activity against MERS-CoV 3CL^{Pro} compared to **12a** and formed a tight interaction with Q195 (Figures 11E, 11F, and S27).

DISCUSSION AND SUMMARY

By exploring MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}, the experimental results clearly present a comprehensive molecular catalytic mechanism for both MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. Taking MERS-CoV 3CL^{Pro} as an example, for substrate binding, E169-H175 identifies the substrate Gln via steric effect rather than through direct electrostatic interaction of E169. Meanwhile, the protease utilizes a hydrogen bond formed

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

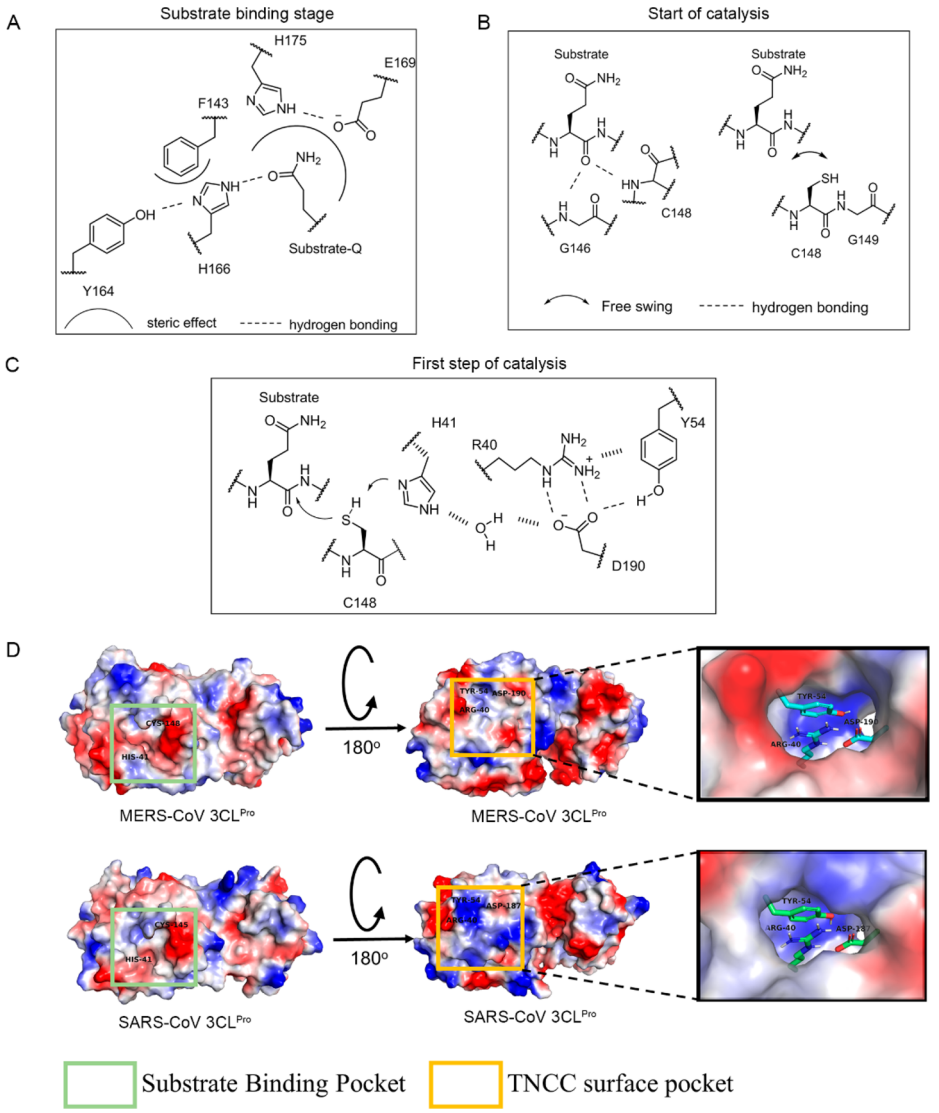


Figure 12. Summary of the catalytic mechanism in MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}, and the allosteric inhibitory site in coronavirus 3CL^{Pro}. MERS-CoV 3CL^{Pro} was applied as an example to illustrate the mechanism. (A) The detailed recognition mechanism with the S1 pocket. (B) The explicit mechanism during the start of catalysis stage. (C) The explicit mechanism during the first step of the catalysis stage. (D) The protein surface of MERS-CoV 3CL^{Pro} (upper) and SARS-CoV 3CL^{Pro} (bottom), and the allosteric inhibitory site in the two proteases. The zone enveloped in the green frame refers to the substrate binding sites, and the zone surrounded with the orange frame refers to the PNCC. Following structural scrutiny, the PNCC is located at the surface of the protein and expressed as a pocket shape, which infers that the partial negative charge cluster may become the target to inhibit the coronavirus 3CL^{Pro}.

by Y164 and the steric effect of F143 to stabilize H166 in an eclipsed conformation which is advantageous for H166 to establish the powerful hydrogen bond interaction with the Gln of substrate and enhance the recognition of the substrate Gln by the protease (Figure 12A). Throughout the investigation, the conserved carbonyl group rather than the side chain N-terminal amide bond in the Gln is considered essential for recognition of 3CL^{Pro}. This implies that replacement of the Gln side chain N-terminal amide bond with a suitable volume group, which should be fit for the conserved residue pair (E169-H175). This strategy may extend the variety of the P1 fragment of 3CL^{Pro} inhibitor.

During the catalysis, a conserved GSCGS motif is identified to form consecutive three turns and plays a crucial role for 3CL^{Pro}. In detail, guaranteed by S147 and S150, Turn I and Turn III fix the substrate active site in an appropriate location and abate the vibrating frequency of active site, which contribute to the attack of active C148. Additionally, maintained by an external hydrogen bond interaction between the motif and N28, stable Turn II and G149 are conducive to the free rotation of C148 side chain such that the C148 thiol approached the substrate active site (Figure 12B). Alternatively, mediated by the conserved water, a remote interaction between dyad and PNCC (R40-Y54-D190) is confirmed to smoothly accelerate the catalysis, as the PNCC can temporarily stabilize the protonated H41 (Figure 12C). Since the PNCC is located at the surface of the opposite side of the active center, it may be a fascinating allosteric site to design 3CL^{Pro} inhibitors via interfering with the partial charge of the PNCC (Figure 12D).

To ensure the accuracy of the comprehensive catalytic mechanism we determined, an analog catalytic mechanism is checked with SARS-CoV 3CL^{Pro}. Importantly, the critical residues are highly conserved among diverse coronaviruses. The mechanisms reported here may be common among coronavirus 3CL^{Pro} and contribute to provide the foundation

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

for the further investigation of SARS-CoV-2 3CL^{Pro} (Figure S20). More significantly, in view of the fact that coronavirus is effortless to evolve, investigation on the general catalytic mechanism of coronavirus 3CL^{Pro} is meaningful to face further virus variation.

According to the catalytic mechanism, the distinctions of two protease on the catalytic characteristics are explicitly investigated. In brief, the essential secondary structure dominated by T174 of MERS-CoV 3CL^{Pro} is confirmed to be responsible for the inferior substrate affinity of MERS-CoV 3CL^{Pro} than SARS-CoV 3CL^{Pro}. In catalysis, the more stably conserved water anchored by Q167 causes more tight remote interaction and leads to the more efficient excision on the substrate in MERS-CoV 3CL^{Pro}. Moreover, the insight into the catalytic mechanism of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro} can effectively guide mutation studies to improve the catalytic potency of the protease. It also meaningfully directs the establishment of powerful hydrogen bond interaction between inhibitor and significant Q195 of MERS-CoV 3CL^{Pro} to improve the inhibitory activity.

In summary, the cumulative experimental results clearly reveal the comprehensive molecular catalytic mechanism of MERS-CoV 3CL^{Pro} and SARS-CoV 3CL^{Pro}. On the basis of the comprehensive mechanism, the distinction of two proteases on the catalytic efficiency is investigated and effective applications are energetically explored. The results presented should provide a solid foundation for understanding the enzymatic mechanism, and help efforts in rational de novo protein design and antiviral drug design.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscatal.0c00110.

Figures S1−S27; Table S1, sequence of the designed primers for construction of plasmid; Table S2, sequence of the designed primers for protein mutagenesis; chemical synthesis methods in detail; NMR data and figures; ab initio model; and figures of the time evolution of RMSD in each system (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Dongmei Li − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin, Nankai University, Tianjin 300350, People's Republic of China*; Email: dongmeili@nankai.edu.cn

Luqing Shang − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*; orcid.org/0000-0002-8052-1249; Email: shanglq@nankai.edu.cn

Authors

Hao Wang − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*

Shuai He − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*

Weilong Deng − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*

Ying Zhang − *Laboratory of Structural Biological & Ministry of Education and Laboratory of Protein Science, School of Medicine and Life Sciences, Tsinghua University, Beijing 100084, People's Republic of China*

Guobang Li − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*

Jixue Sun − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin, Nankai University, Tianjin 300350, People's Republic of China*

Wei Zhao − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*

Yu Guo − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China*; orcid.org/0000-0002-8109-7515

Zheng Yin − *State Key Laboratory of Medicinal Chemical Biology, College of Pharmacy and KLMDASR of Tianjin and Drug Discovery Center for Infectious Disease, Nankai University, Tianjin 300350, People's Republic of China; Center of Basic Molecular Science, Department of Chemistry, Tsinghua University, Beijing 100084, People's Republic of China*; orcid.org/0000-0002-3974-4053

Complete contact information is available at: https://pubs.acs.org/10.1021/acscatal.0c00110

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We sincerely thank Prof. Zhao, Yilei., School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, for the PRS analysis and Prof. Wang, Binghe., Department of Chemistry, Georgia State University, for the suggestion of this work. This work was supported by the National Key Research and Development Program of China (grant no. 2018YFA0507204), the National Natural Science Foundation of China (grant nos. 21672115), the Natural Science Foundation of Tianjin (grant nos. 19JCZDJC33300), and the "111" Project of the Ministry of Education of China. In addition, we thank the staff of Tsinghua University for their support in collecting the diffraction data.

■ REFERENCES

(1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727−733.

(2) Wang, C.; Horby, P. W.; Hayden, F. G.; Gao, G. F. A Novel Coronavirus Outbreak of Global Health Concern. *Lancet* **2020**, *395*, 470−473.

(3) Nassar, M. S.; Bakhrebah, M. A.; Meo, S. A.; Alsuabeyl, M. S.; Zaher, W. A. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Infection: Epidemiology, Pathogenesis and Clinical Characteristics. *Eur. Rev. Med. Pharmacol. Sci.* **2018**, *22*, 4956−4961.

(4) Wise, J. Patient with New Strain of Coronavirus is Treated in Intensive Care at London Hospital. *Br. Med. J.* **2012**, *345*, e6455.
(5) Yeung, M. L.; Yao, Y.; Jia, L.; Chan, J. F. W.; Chan, K. H.; Cheung, K. F.; Chen, H.; Poon, V. K. M.; Tsang, A. K. L.; To, K. K. W.; et al. MERS Coronavirus Induces Apoptosis in Kidney and Lung by Upregulating Smad7 and FGF2. *Nat. Microbiol.* **2016**, *1*, 16004−16011.
(6) Ko, J. H.; Park, G. E.; Lee, J. Y.; Lee, J. Y.; Cho, S. Y.; Ha, Y. E.; Kang, C. I.; Kang, J. M.; Kim, Y. J.; Huh, H. J.; Ki, C. S.; Jeong, B. H.; Park, J.; Chung, C. R.; Chung, D. R.; Song, J. H.; Peck, K. R. Predictive Factors for Pneumonia Development and Progression to Respiratory Failure in MERS-CoV Infected Patients. *J. Infect.* **2016**, *73*, 468−475.
(7) Abuhammad, A.; Al-Aqtash, R. A.; Anson, B. J.; Mesecar, A. D.; Taha, M. O. Computational Modeling of the Bat HKU4 Coronavirus 3CL(pro) Inhibitors as a Tool for the Development of Antivirals against the Emerging Middle East Respiratory Syndrome (MERS) Coronavirus. *J. Mol. Recognit.* **2017**, *30*, e2644.
(8) Galasiti Kankanamalage, A. C.; Kim, Y.; Damalanka, V. C.; Rathnayake, A. D.; Fehr, A. R.; Mehzabeen, N.; Battaile, K. P.; Lovell, S.; Lushington, G. H.; Perlman, S.; Chang, K.-O.; Groutas, W. C. Structure-Guided Design of Potent and Permeable Inhibitors of MERS Coronavirus 3CL Protease that Utilize a Piperidine Moiety as a Novel Design Element. *Eur. J. Med. Chem.* **2018**, *150*, 334−346.
(9) Kumar, V.; Tan, K. P.; Wang, Y. M.; Lin, S. W.; Liang, P. H. Identification, Synthesis and Evaluation of SARS-CoV and MERS-CoV 3C-Like Protease Inhibitors. *Bioorg. Med. Chem.* **2016**, *24*, 3035−3042.
(10) van den Brand, J. M.; Smits, S. L; Haagmans, B. L Pathogenesis of Middle East Respiratory Syndrome Coronavirus. *J. Pathol.* **2015**, *235*, 175−184.
(11) Lee, H.-J.; Shieh, C.-K.; Gorbalenya, A. E.; Koonin, E. V.; La Monica, N.; Tuler, J.; Bagdzhadzhyan, A.; Lai, M. M.C. The Complete Sequence (22 kilobases) of Murine Coronavirus Gene 1 Encoding the Putative Proteases and RNA Polymerase. *Virology* **1991**, *180*, 567−582.
(12) Marra, M. A.; Jones, S. J. M.; Astell, C. R.; Holt, R. A.; Angela, B. W.; Butterfield, Y. S. N.; Jaswinder, K.; Asano, J. K.; Barber, S. A.; Chan, S. Y. The Genome Sequence of the SARS-Associated Coronavirus. *Science* **2003**, *300*, 1399−1404.
(13) Woo, P. C. Y.; Huang, Y.; Lau, S. K. P.; Yuen, K.-Y. Coronavirus Genomics and Bioinformatics Analysis. *Viruses* **2010**, *2*, 1804−1820.
(14) St John, S. E.; Tomar, S.; Stauffer, S. R.; Mesecar, A. D. Targeting Zoonotic Viruses: Structure-Based Inhibition of the 3C-Like Protease from Bat Coronavirus HKU4-The Likely Reservoir Host to the Human Coronavirus that Causes Middle East Respiratory Syndrome (MERS). *Bioorg. Med. Chem.* **2015**, *23*, 6036−6048.
(15) Ratia, K.; Saikatendu, K. S.; Santarsiero, B. D.; Barretto, N.; Baker, S. C.; Stevens, R. C.; Mesecar, A. D. Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease: Structure of a Viral Deubiquitinating Enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5717−5722.
(16) Chen, S.; Chen, L.; Tan, J.; Chen, J.; Du, L.; Sun, T.; Shen, J.; Chen, K.; Jiang, H.; Shen, X. Severe Acute Respiratory Syndrome Coronavirus 3C-Like Proteinase N Terminus is Indispensable for Proteolytic Activity but not for Enzyme Dimerization. Biochemical and Thermodynamic Investigation in Conjunction with Molecular Dynamics Simulations. *J. Biol. Chem.* **2005**, *280*, 164−173.
(17) Wojdyla, J. A.; Manolaridis, I.; van Kasteren, P. B.; Kikkert, M.; Snijder, E. J.; Gorbalenya, A. E.; Tucker, P. A. Papain-Like Protease 1 from Transmissible Gastroenteritis Virus: Crystal Structure and Enzymatic Activity toward Viral and Cellular Substrates. *J. Virol.* **2010**, *84*, 10063−10073.
(18) Needle, D.; Lountos, G. T.; Waugh, D. S. Structures of the Middle East Respiratory Syndrome Coronavirus 3C-Like Protease Reveal Insights into Substrate Specificity. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 1102−1111.
(19) Tomar, S.; Johnston, M. L.; St John, S. E.; Osswald, H. L.; Nyalapatla, P. R.; Paul, L. N.; Ghosh, A. K.; Denison, M. R.; Mesecar, A. D. Ligand-Induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus Nsp5 Protease (3CL Pro). *J. Biol. Chem.* **2015**, *290*, 19403−19422.
(20) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The Crystal Structures of Severe Acute Respiratory Syndrome Virus Main Protease and its Complex with an Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13190−13195.
(21) Anand, K. Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs. *Science* **2003**, *300*, 1763−1767.
(22) Hegyi, A.; Ziebuhr, J. Conservation of Substrate Specificities among Coronavirus Main Proteases. *J. Gen. Virol.* **2002**, *83*, 595−599.
(23) Chou, K. C.; Wei, D. Q.; Zhong, W. Z. Binding Mechanism of Coronavirus Main Proteinase with Ligands and its Implication to Drug Design against SARS. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 148−151.
(24) Hsu, M. F.; Kuo, C. J.; Chang, K. T.; Chang, H. C.; Chou, C. C.; Ko, T. P.; Shr, H. L.; Chang, G. G.; Wang, A. H. J.; Liang, P. H. Mechanism of the Maturation Process of SARS-CoV 3CL Protease. *J. Biol. Chem.* **2005**, *280*, 31257−31266.
(25) Huang, C.; Wei, P.; Fan, K.; Liu, Y.; Lai, L. 3C-like Proteinase from SARS Coronavirus Catalyzes Substrate Hydrolysis by a General Base Mechanism. *Biochemistry* **2004**, *43*, 4568−4574.
(26) Shang, L.; Zhang, S.; Yang, X.; Sun, J.; Li, L.; Cui, Z.; He, Q.; Guo, Y.; Sun, Y.; Yin, Z. Biochemical Characterization of Recombinant Enterovirus 71 3C Protease with Fluorogenic Model Peptide Substrates and Development of a Biochemical Assay. *Antimicrob. Agents Chemother.* **2015**, *59*, 1827−1836.
(27) Zhai, Y.; Ma, Y.; Ma, F.; Nie, Q.; Ren, X.; Wang, Y.; Shang, L.; Yin, Z. Structure−Activity Relationship Study of Peptidomimetic Aldehydes as Enterovirus 71 3C Protease Inhibitors. *Eur. J. Med. Chem.* **2016**, *124*, 559−573.
(28) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(29) Fletcher, R.; Powell, M. J. D. A Rapidly Convergent Descent Method for Minimization. *Comput. J.* **1963**, *6*, 163−168.
(30) Fletcher, R.; Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput. J.* **1964**, *7*, 149−154.
(31) Pastor, R. W.; Brooks, B. R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409−1419.
(32) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(33) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(34) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084−3095.
(35) Durrant, J. D.; de Oliveira, C. A. F.; McCammon, J. A. POVME: an Algorithm for Measuring Binding-Pocket Volumes. *J. Mol. Graphics Modell.* **2011**, *29*, 773−776.
(36) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.
(37) Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. Molecular Dynamics and Free-Energy Calculations Applied to Affinity Maturation in Antibody 48G7. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 14330−14335.
(38) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(39) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275−282.

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890

(40) Paasche, A. *Mechanistic Insights into SARS Coronavirus Main Protease by Computational Chemistry Methods*; Julius-Maximilians-Universität: Würzburg, Germany, 2013.
(41) Taranto, A. G.; Carvalho, P.; Avery, M. A. QM/QM Studies for Michael Reaction in Coronavirus Main Protease (3CLPro). *J. Mol. Graphics Modell.* **2008**, *27*, 275−285.
(42) Zhou, J.; Wang, Y.; Xu, G.; Wu, L.; Han, R.; Schwaneberg, U.; Rao, Y.; Zhao, Y.; Zhou, J.; Ni, Y. Structural Insight into Enantioselective Inversion of an Alcohol Dehydrogenase Reveals a "Polar Gate" in Stereorecognition of Diaryl Ketones. *J. Am. Chem. Soc.* **2018**, *140*, 12645−12654.
(43) Shi, T.; Liu, L.; Tao, W.; Luo, S.; Fan, S.; Wang, X. L.; Bai, L.; Zhao, Y. Theoretical Studies on the Catalytic Mechanism and Substrate Diversity for Macrocyclization of Pikromycin Thioesterase. *ACS Catal.* **2018**, *8*, 4323−4332.
(44) Chen, X.; Shi, T.; Wang, X.; Wang, J.; Chen, Q.; Bai, L.; Zhao, Y. Theoretical Studies on the Mechanism of Thioesterase-Catalyzed Macrocyclization in Erythromycin Biosynthesis. *ACS Catal.* **2016**, *6*, 4369−4378.
(45) Raghavachari, K. Perspective on "Density functional thermochemistry. III. The role of exact exchange. *Theor. Chem. Acc.* **2000**, *103*, 361−363.
(46) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy into a function of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(47) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.
(48) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.01, Gaussian, Inc.: Wallingford CT, 2009.
(49) Shi, J.; Wei, Z.; Song, J. 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.* **2004**, *279*, 24765−24773.
(50) Zhai, Y.; Zhao, X.; Cui, Z.; Wang, M.; Wang, Y.; Li, L.; Sun, Q.; Yang, X.; Zeng, D.; Liu, Y.; et al. Cyanohydrin as an Anchoring Group for Potent and Selective Inhibitors of Enterovirus 71 3C Protease. *J. Med. Chem.* **2015**, *58*, 9414−9420.
(51) Wang, Y.; Cao, L.; Zhai, Y.; Yin, Z.; Sun, Y.; Shang, L. Structure of the Enterovirus 71 3C Protease in Complex with NK-1.8k and Indications for the Development of Antienterovirus Protease Inhibitor. *Antimicrob. Agents Chemother.* **2017**, *61*, 00298−17.
(52) Ma, Y.; Shang, C.; Yang, P.; Li, L.; Zhai, Y.; Yin, Z.; Wang, B.; Shang, L. 4-Iminooxazolidin-2-one as a Bioisostere of the Cyanohydrin Moiety: Inhibitors of Enterovirus 71 3C Protease. *J. Med. Chem.* **2018**, *61*, 10333−10339.
(53) Ma, Y.; Li, L.; He, S.; Shang, C.; Sun, Y.; Liu, N.; Meek, T. D.; Wang, Y.; Shang, L. Application of Dually Activated Michael Acceptor to the Rational Design of Reversible Covalent Inhibitor for Enterovirus 71 3C Protease. *J. Med. Chem.* **2019**, *62*, 6146−6162.
(54) Wang, X.; Jing, X.; Deng, Y.; Nie, Y.; Xu, F.; Xu, Y.; Zhao, Y.-L.; Hunt, J. F.; Montelione, G. T.; Szyperski, T. Evolutionary Coupling Saturation Mutagenesis: Coevolution-Guided Identification of Distant Sites Influencing Bacillus Naganoensis Pullulanase Activity. *FEBS Lett.* **2020**, *594*, 799−812.
(55) Yi, L.; Gebhard, M. C.; Li, Q.; Taft, J. M.; Georgiou, G.; Iverson, B. L. Engineering of TEV Protease Variants by Yeast ER Sequestration Screening (YESS) of Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7229−7234.
(56) Yu, X.; Sun, J.; Wang, W.; Jiang, L.; Cheng, B.; Fan, J. Assessment of the Fusion Tags on Increasing Soluble Production of the Active TEV Protease Variant and Other Target Proteins in *E. coli*. *Appl. Biochem. Biotechnol.* **2017**, *182*, 769−781.

https://dx.doi.org/10.1021/acscatal.0c00110
ACS Catal. 2020, 10, 5871−5890