Supplementary Data

Design of the fluorogenic substrate

Since it has been well defined that the specificities of coronavirus 3C-like proteinases against the substrate involves bulky hydrophobic residues (mainly leucine/isoleucine) at the P2 position, conserved glutamine at the P1 position, and small aliphatic residues at the P1' position, a 12-amino acid peptide representing the NH_2-terminal autoprocessing site of TGEV $3CL^{pro}$ with the sequence of substrate-analog chloromethyl ketone inhibitor Cbz-Val-Asn-Ser-Thr-Leu-Gln-CMK, was devised to separate the quencher from the fluorescent donor chromophore for designing the fluorogenic substrate in this work (Figure S1). The commercially available donor/quencher pair was chosen for the study: 5-[(2'-aminoethyl)-amino] naphthelenesulfonic acid (EDANS) and 4-[[4-(dimethylamino) phenyl] azo] benzoic acid (Dabcyl).

Enzymatic activity at different pH values

The relative enzymatic activities of the full-length SARS $3CL^{pro}$ at various pH values were measured to evaluate the viability of the fluorogenic peptide as the potential substrate of $3CL^{pro}$. As indicated in Figure S2, the proteinase exhibits a stable proteolytic activity at pH 7.0-9.0, and displays only 50% activity at pH 6.0 and 10.0. However, when pH decreases to 5.0, it has almost lost its activity completely. The result is in agreement with the published data (9,26), indicating the method for determining enzymatic activity is reliable.

Proteinase-substrate docking and binding free energy prediction

The advanced docking program Autodock 3.0.3 was used to dock the substrate into the binding pocket of each conformation of the two proteinases (models 1 and 2). For each model, 100 conformations were isolated from the MD trajectory, i.e. one conformation was selected from the MD trajectory every 50 ps. The potentials of the structures of proteinases and substrate peptide were assigned according to the Amber 4.0 force field with Kollman-united-atom charges.

The Lamarckian genetic algorithm (LGA) was applied to deal with the substrate-proteinase interaction. Briefly, the LGA described the relationship between the substrate and the proteinases by the translation, orientation, and conformation of the substrate. These so-called "state variables" were the substrate's genotype, and the resulting atomic coordinates together with the interaction and the intramolecular energies were the substrate's phenotype. The environmental adaptation of the phenotype was reverse-transcribed into its genotype and became heritable traits. Each docking cycle, or generation, consisted of a regimen of fitness evaluation, crossover, mutation, and selection. A Solis and Wets local search performed the energy minimization on a user-specified proportion of the population. The docked structure of the substrate was generated after a reasonable number of evaluations.

On the basis of the traditional molecular force field model of interaction energy, a new score function at the level of binding free energy was derived and adopted in the version of Autodock 3.0.3. Not only the restriction of internal rotors, the global rotation, and the translation were modeled depending on the number of torsion angles

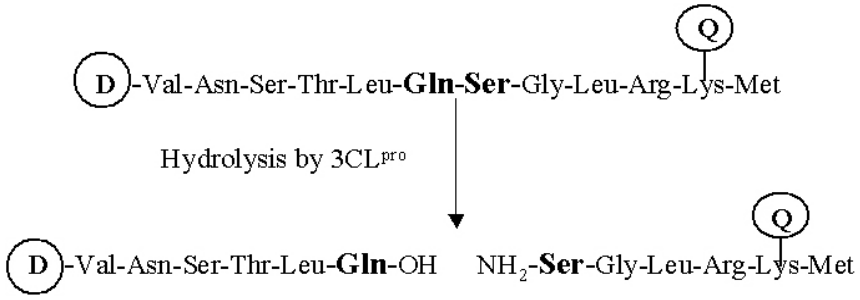
of the substrate but also the desolvation upon binding and the hydrophobic effect (solvent entropy changes at solute-solvent interfaces) were calculated. The total binding free energy was empirically calibrated based on the above-stated terms and a set of coefficient factors.

Table S1. Important residue pairs involved in the hydrogen bonds between the two protomers of the full-length and N-terminal deleted SARS 3CL^{pro}. Hydrogen bonds appeared in both proteinases are aligned and shown in bold.

H-bonds in the Full-length proteinase		H-bonds in N-terminal deleted proteinase	
Protomer A	Protomer B	Protomer A	Protomer B
Phe3(O)	Phe140(N_H)		
Phe3(N_H)	Ser139(O^{η})		
Ala7(N_H)	Val125(O)		
		Phe8(N_H3)	Ser123(O)
Ser10(N_H)	**Ser10(O^{η})**	**Ser10(N_H)**	**Ser10(O^{η})**
Ser10(O^{η})	**Ser10(N_H)**	**Ser10(O^{η})**	**Ser10(N_H)**
Ser10(O^{η})	**Gly11(N_H)**	**Ser10(O^{η})**	**Gly11(N_H)**
		Gly11(N_H)	Ser10(O^{η})
Gly11(N_H)	**Glu14($O^{\varepsilon 1}$)**	**Gly11(N_H)**	**Glu14($O^{\varepsilon 1}$)**
		Glu14($O^{\varepsilon 1}$)	Ser10(N_H)
Glu14($O^{\varepsilon 1}$)	**Gly11(N_H)**	**Glu14($O^{\varepsilon 1}$)**	**Gly11(N_H)**
		Glu14($O^{\varepsilon 2}$)	Gly11(N_H)
Val125(O)	Ala7(N_H)		
Ser139(O^{η}_H^{η})	Gly2(O)		
Ser139(O^{η}_H^{η})	Gln299($O^{\varepsilon 1}$)		
Phe140(N_H)	Ser1(O)		
Phe140(O)	Ser1(N_H)		
His172($N^{\varepsilon 2}$_$H^{\varepsilon 2}$)	Ser1(O^{η})		
Asn214($O^{\delta 1}$)	Ser139(O^{η}_H^{η})		
Met276(O)	Thr285($O^{\eta 1}$_$H^{\eta 1}$)		
Glu290($O^{\varepsilon 1}$)	Arg4($N^{\theta 2}$_$H^{\theta 22}$)		
		Thr169(N_H)	Asn214($O^{\delta 1}$)
		Thr169(N_H)	Asn214(O)
		Thr169($O^{\eta 1}$_$H^{\eta 1}$)	Asp216($O^{\delta 1}$)
		Thr169($O^{\eta 1}$_$H^{\eta 1}$)	Asp216($O^{\delta 2}$)
		Gly170(N_H)	Leu282(O)
		Asn214(O)	Gly170(N_H)
		Asn214($N^{\delta 2}$_$H^{\delta 22}$)	Asn142($N^{\delta 2}$)

Figure S1

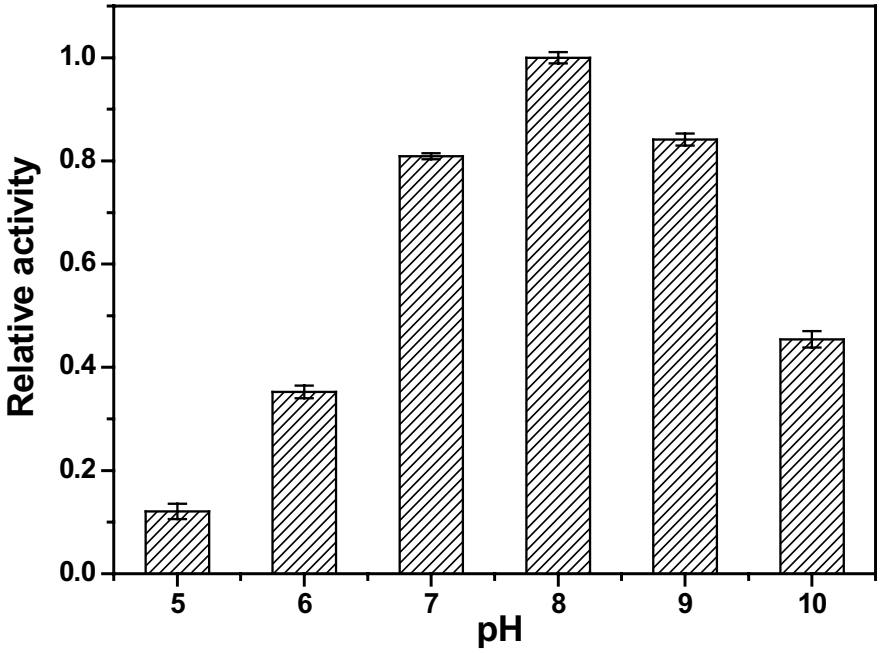
Design of internally quenched substrate of SARS 3CL^{pro}



Proposed design of fluorogenic substrate

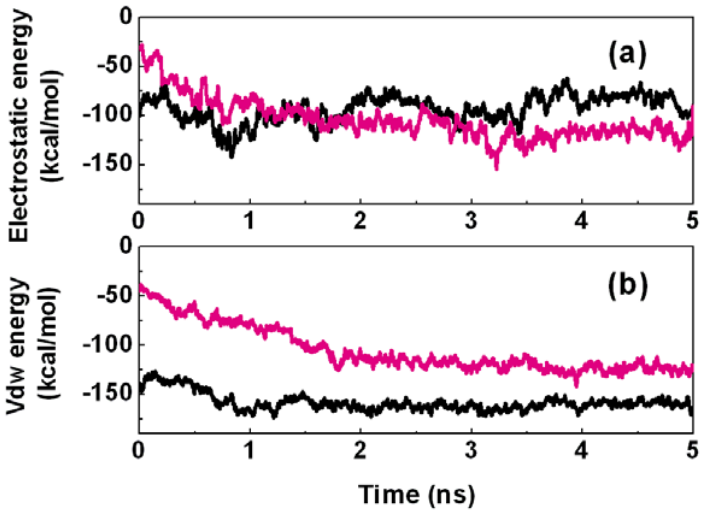
Figure S2

Enzymatic activity of the full-length SARS 3CL^{pro} at different pH



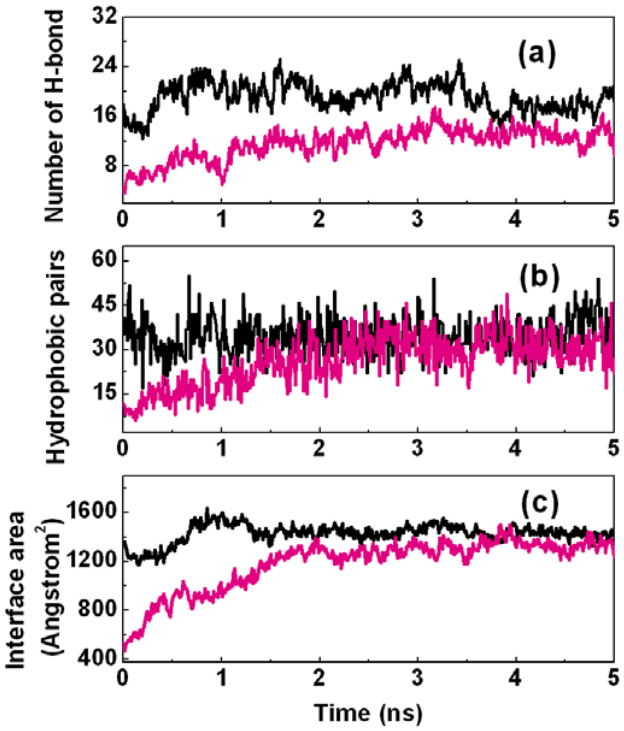
The proteolytic activity was determined at 25°C in sodium phosphate buffer (pH=5, 6, 7, 8) or glycine/NaOH buffer (pH=9, 10) containing 5mM DDT, 1mM EDTA, 1μM SARS 3CL^{pro} and 10μM substrate. The proteolytic activity at each pH value was measured in triplicate and averaged, and the enzymatic activity at pH 8.0 was taken as 1.0.

Figure S3



Electrostatic (a) and van der Waals (b) interaction energies for the full-length (black) and N-terminal deleted (pink) SARS $3CL^{pro}$s versus MD simulation time (shown as 10 ps average).

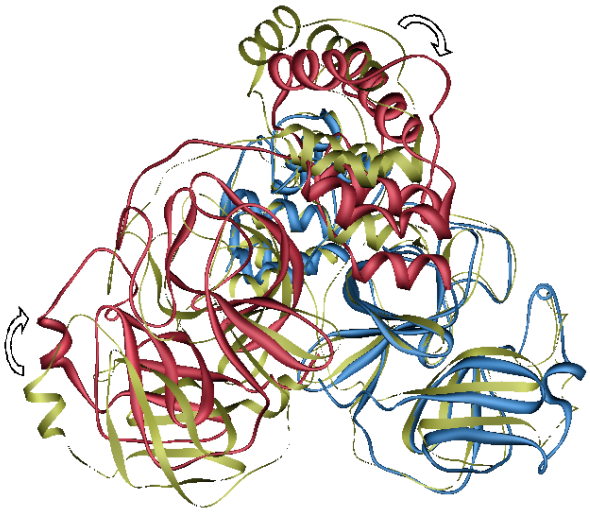
Figure S4



(a) Number of hydrogen bonds and (b) hydrophobic interaction pairs (HIPs) at the interface of the dimer for the full-length (black) and N-terminal deleted (pink) SARS $3CL^{pro}$s versus MD simulation time (shown as 10 ps average).

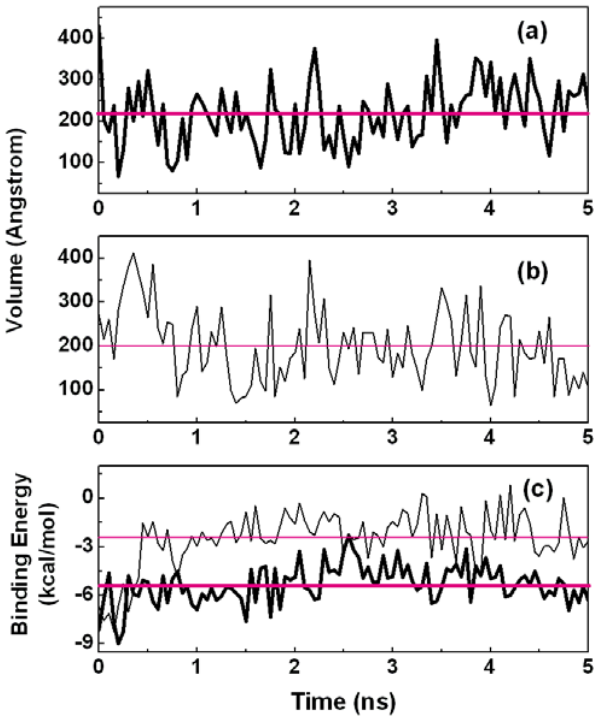
(c) The interface areas between the two protomers of the full-length (black) and N-terminal deleted (pink) SARS $3CL^{pro}$s versus MD simulation time (shown as 10 ps average).

Figure S5



Superposition of the N-terminal deleted proteinase dimer (protomers A and B are represented as red and blue, respectively) with the full-length proteinase dimer (yellow). To clearly demonstrate the dimerization state change, the two protomer Bs of the full-length and N-terminal deleted proteinases were superposed each other with the smallest root mean-square deviation (RMSD). The arc arrows represent the rotation direction of protomer A of the N-terminal deleted proteinase dimer.

Figure S6



The fluctuations of the binding pocket volumes of the full-length (a) and N-terminal deleted (b) SARS $3CL^{pro}$s during the MD simulation. The pink lines are the average values of the volumes.

(c) The fluctuations of the binding free energies of the full-length (thick curve) and N-terminal deleted (thin curve) proteinases to the model peptide substrate (TSAVLQ) versus MD simulation time. The pink lines represent the average free energies for the two proteinases to the substrate.