

CHEMBIOCHEM

Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2007

CHEM**BIO**CHEM

Supporting Information

for

Structure-Based Design and Synthesis of Highly Potent SARS-CoV 3CL Protease Inhibitors

Yi-Ming Shao, Wen-Bin Yang, Hung-Pin Peng, Min-Feng Hsu, Keng-Chang Tsai,
Tun-Hsun Kuo, Andrew H.-J. Wang, Po-Huang Liang, Chun-Hung Lin,
An-Suei Yang*, and Chi-Huey Wong*

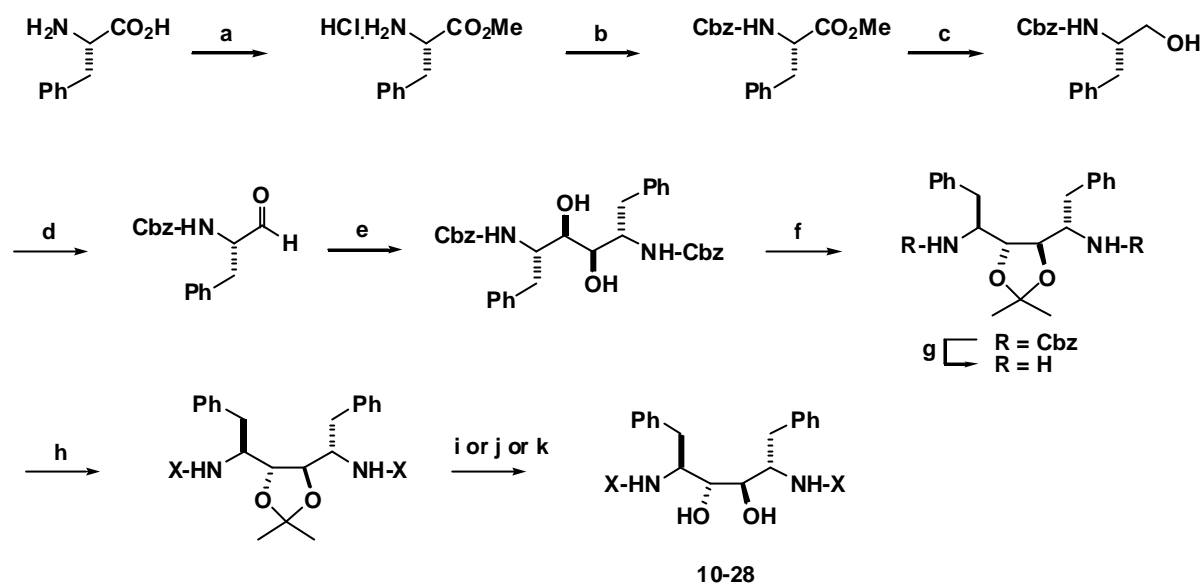
- I General synthetic procedures, materials and instrumentation
- II Synthesis and activity of **10-28**
- III Synthesis and activity of **4** and **9**
- IV Computer modeling and crystallization

I. General Synthetic Procedures, Materials and Instrumentation

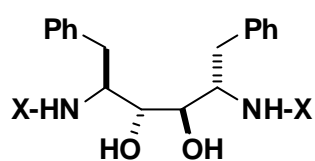
All reactions with air- and moisture-sensitive materials were performed in oven-dried glassware fitted with rubber septa or three-way T taps under a positive pressure of argon or nitrogen. Air- and moisture-sensitive liquids and solutions were transferred via syringe. Organic solutions were concentrated by rotary evaporation at 23-80 °C (water-bath temperature). Column chromatography was performed employing Merck silica gel (60 Å pore size, 70–230 mesh ASTM). Analytical thin-layer chromatography (TLC) was performed using glass plates pre-coated with Merck silica gel (60 Å pore size) impregnated with a fluorescent indicator (254 nm). TLC plates were visualized by UV lamp, phosphomolybdic acid solution in ethanol, or 0.5% ninhydrin in ethanol followed by brief heating on a hot plate. Commercial solvents and reagents were used as received without further purification. They were purchased from Aldrich, ACROS, BACHEM, or other commercial sources.

Compounds are characterized by nuclear magnetic resonance spectroscopy, infrared spectroscopy, optical rotations, and high resolution mass spectroscopy. Proton nuclear magnetic resonance (^1H NMR) spectra and carbon nuclear magnetic resonance (^{13}C NMR) spectra were recorded with Bruker Avance 600 (600 MHz/ 150 MHz) and Bruker Avance 400 (400 MHz/100 MHz) NMR spectrometers. Chemical shifts for protons are reported in parts per million (ppm; δ scale) and are referenced to residual protium in the NMR solvents (CHCl_3 : δ 7.26, D_2HCO : δ 3.31, $\text{C}_2\text{D}_5\text{HSO}$: δ 2.50). Chemical shifts for carbon are reported in parts per million (ppm; δ scale) and are referenced to the carbon resonances of the solvent (CDCl_3 : δ 77.23, D_3CO : δ 49.15, $(\text{CD}_3)_2\text{SO}$: δ 39.50). Data are represented as follows: chemical shift, multiplicity, coupling constant in Hz, integration, and assignment. Infrared (IR) spectra were obtained using a Nicolet 380 FTIR-ATR spectrophotometer. Data are represented as follows: frequency of the absorption (cm^{-1}), intensity of absorption, and assignment (where appropriate). Optical rotations were determined on a PerkinElmer 341 digital polarimeter equipped with a sodium lamp source using a 1-mL solution cell. Reported readings are the average of five measurements for each sample. High resolution mass spectra were obtained using Bruker Daltonics BioTOF III.

II. Synthesis and Activity of 10-28



Scheme S1. Synthesis of TL-3 and its derivatives. Reagents and conditions: (a) SOCl_2 , MeOH, reflux, 16 h, 100%; (b) $\text{Na}_2\text{CO}_{3(\text{aq})}$, CbzCl, CHCl_3 , 22 °C, 2 h, 100%; (c) LiCl/NaBH₄ (2 equiv.), THF/EtOH (1:2), 23 °C, 8 h, 94%; (d) TEMPO, NaBr, $\text{NaOCl}_{(\text{aq})}$, NaHCO_3 , EtOAc/H₂O (12:1), 0 ? 23 °C, 1 h; (e) $\text{VCl}_3(\text{thf})_3$ (1.2 equiv), Zn (1.4 equiv), CH_2Cl_2 , 23 °C, 3 h, 60%; (f) 2,2-dimethoxypropane, *p*-TsOH·H₂O, 70-80 °C, 1.5 h, then 21 °C, 12.5 h, 80%; (g) Pd/C, H₂, EtOH/ CHCl_3 (2:1), 99%; (h) amino acid coupling; (i) DDQ (2 equiv), MeCN/H₂O (5:1), 75 °C, 31-92%; (j) $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ (7 equiv), CH_2Cl_2 , 23 °C, 40-83%; (k) *p*-TsOH·H₂O, MeOH, reflux, 27 h, 93%.

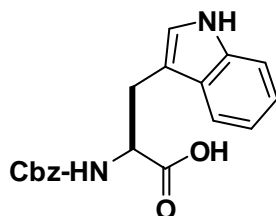
Table S1. Activity of TL-3 derivatives against SARS-CoV 3CL protease.**10-28**

No.	X	IC ₅₀ [μ M]
10	Cbz-Ala-Val	8
11	Cbz-Ala-Ser	40
12	Cbz-Ala-Thr	25
13	Cbz-Ile	30
14	Cbz-Ala-Thr(O t Bu)	40
15	Cbz-Thr(O t Bu)	>50
16	Cbz-Val	50
17	Cbz-Leu	>50
18	Cba-Ala-Leu	>50
19	Ac-Gln-Ala-Val	>50
20	Ac-Gln(Trt)-Ala-Val	>50
21	Cbz-Gly-Val	40
22	Cbz-Val-Thr	25
23	Cbz-Phe-Thr	30
24	Cbz-Ala-Phe	25
25	Cbz-Ala-Ile	20
26	Cbz-Gly-Ile	>50
27	Cbz-Ser-Ile	>50
28	Cbz-Thr-Ile	>50

III. Synthesis and Activity of 4 and 9

General procedure A: Deprotection of Cbz group. To a stirred solution of Cbz-protected compound in THF/MeOH (1:1) was added catalytic amounts of 10% Pd/C. The mixture was stirred under H₂ (1atm) at 23 °C until no starting material could be detected by TLC analysis. The amine compound was confirmed by developing on TLC and stained by ninhydrin. The catalyst was removed by filtration through a pad of Celite and washed with MeOH. The filtrate was concentrated under reduced pressure to give the corresponding amine, which was used in the amide-formation reaction without further purification.

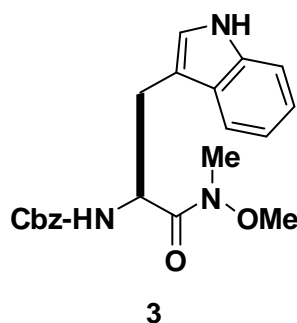
General procedure B: Peptide bond formation. To a stirred solution of diamine in DMF was added *N*-Cbz-protected amino acid, HBTU, and DIEA. The reaction mixture was stirred for 3~24 h at 23 °C under Ar. When no starting material (diamine) was present by TLC analysis, DMF was evaporated under reduced pressure. The residue was diluted with EtOAc and H₂O. Water layer was separated and extracted with EtOAc (3x). The organic layers were collected and washed with brine, dried (Na₂SO₄ or MgSO₄), filtered, and concentrated under reduced pressure. Purification by SiO₂ chromatography gave the desired product as a white solid.



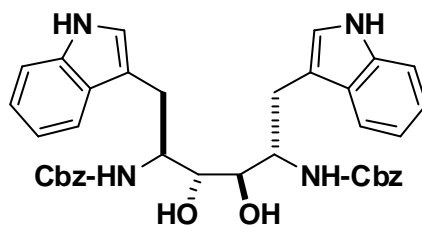
2

***N*-Cbz-L-Trp-OH (2).** To a clear solution of tryptophan (**1**) (3000 mg, 14.7 mmol) in 125 mL of CH₃CN-H₂O (2:3) were added NaHCO₃ (1852 mg, 22 mmol) and Na₂CO₃ (2337 mg, 22 mmol). The turbid solution was tested by pH sticks (pH 10~11). The mixture was cooled to 0 °C (H₂O/ice bath) and stirred. To this stirred mixture was added Cbz-Cl (2153 μL, 14.7 mmol) slowly. The resulting solution was stirred for 2 h at 23 °C, at which time no starting material remained (TLC analysis). After acidified by dropwise addition of a 1 N HCl aqueous solution and removal of CH₃CN by rotary evaporation, the reaction mixture was transferred to a separatory funnel and washed three times with EtOAc. The combined organic phase was washed with brine, dried over Na₂SO₄,

and filtered. Concentration under reduced pressure gave the desired product (**2**) as a white solid. (4495 mg, 90%). R_f 0.67 (4:1 $\text{CH}_2\text{Cl}_2/\text{CH}_3\text{OH}$); ^1H NMR (400 MHz, CD_3OD) δ 7.56 (d, $J = 7.9$ Hz, 1H, ArH), 7.35-6.97 (m, 9H, ArH), 5.03 (AB, $J = 12.7$ Hz, $\tau_{ab} = 17.5$ Hz, 2H, PhCH_2O), 4.50 (m, 1H, CH_a), 3.35 (m, 1H, $\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$), 3.15 (dd, $J = 14.7, 8.1$ Hz, 1H, $\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$).



***N*-Cbz-L-tryptophan *N,O*-dimethylhydroxylamide (**3**).** To an ice-cooled and stirred solution of **2** (58.3 mg, 0.17 mmol), *N, O*-dimethylhydroxylamine hydrochloride (18 mg, 0.18 mmol), and triethylamine (26 μL , 0.18 mmol) in DMF (6 mL) were added HOBt (26 mg, 0.19 mmol) and WSC (36 mg, 0.19 mmol), and the whole was stirred at room temperature for 15 h. The reaction mixture was concentrated in vacuo. The residue was extracted with ethyl acetate, and the organic layers were combined and washed with 1 N HCl, brine, dried over MgSO_4 , and concentrated under reduced pressure. Purification by column chromatography (1:2 to 1:1 ethyl acetate-hexanes) afforded the title compound (**3**) as a white solid (56.8 mg, 86%). R_f 0.33 (1:1 hexanes/EtOAc); ^1H NMR (600 MHz, CDCl_3) δ 8.27 (br s, 1H, ArH), 7.59 (d, $J = 6$ Hz, 1H, ArH), 7.34-7.31 (m, 6H, ArH), 7.18 (t, $J = 6$ Hz, 1H, ArH), 7.11 (t, $J = 6$ Hz, 1H, ArH), 6.95 (s, 1H, ArH), 5.57 (d, $J = 6$ Hz, 1H, NH amide), 5.07 (AB, $J = 12$ Hz, $\tau_{ab} = 30$ Hz, 2H, PhCH_2O ; 1H, CH_a), 3.66 (s, 3H, NOCH_3), 3.27 (dd, $J = 18, 6$ Hz, 1H, $\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$), 3.17 (s, 3H, NCH_3), 3.14 (dd, $J = 18, 6$ Hz, 1H, $\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$); ^{13}C NMR (150 MHz, CDCl_3) δ 172.4, 156.1, 136.5, 136.3, 128.6, 128.2, 128.3, 127.8, 123.1, 122.2, 119.6, 118.7, 111.4, 110.4, 66.9, 61.7, 51.7, 32.3, 28.3; HRMS (ESI) calcd for $\text{C}_{21}\text{H}_{23}\text{N}_3\text{O}_4\text{Na}$ [$M+\text{Na}$] $^+$: 404.1586, found: 404.1664.

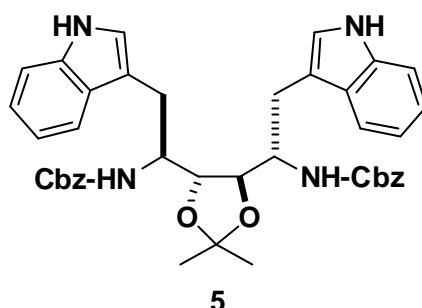


4

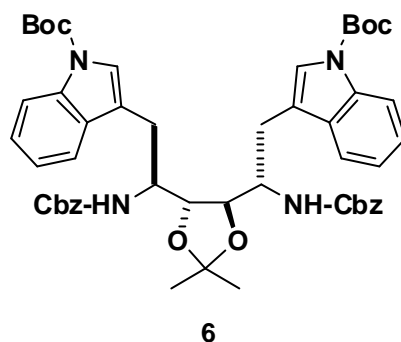
(2S,3R,4R,5S)-2,5-Bis[N-(benzyloxycarbonyl)amino]-1,6-di(3-indolyl)-3,4-hexane diol (4). To a stirred solution of **3** (106 mg, 0.28 mmol) in dry THF (15 mL) was added dropwise 20 wt% diisobutylaluminum hydride toluene solution (1840 μ L, 2.23 mmol) at -78 $^{\circ}$ C. After stirring at -78 $^{\circ}$ C for 15 min, the reaction mixture was poured into 1 N HCl_(aq), extracted with EtOAc, washed with brine, dried over Na₂SO₄, and concentrated under reduced pressure to give a yellow solid. The crude ¹H NMR spectrum is consistent with the formulated structure, so the residue was used in the pinacol homo-coupling immediately without further purification.

A round bottom flask containing a magnetic stirring bar was charged with Zn powder (279 mg, 4.3 mmol) and VC₃(THF)₃ (420 mg, 1.1 mmol). The flask was evacuated and purged with Ar (3x) balloon. The dry CH₂Cl₂ was added and the solution was stirred, causing after stirring for 20 min a color change from red to green. A solution of crude aldehyde in dry CH₂Cl₂ was added, causing a color change from green to brown. The resulting mixture was stirred for 17 h at 23 $^{\circ}$ C, at which time no starting material could be detected by TLC analysis. The reaction solution was opened to air and poured into 1 N HCl. The two phases were stirred together for 2 h giving a nearly colorless CH₂Cl₂ layer and a blue aqueous layer. The two layers were separated and the aqueous layer was extracted with CH₂Cl₂ for three times. The combined organic layers were washed with brine, and then dried with MgSO₄, filtered, and evaporated to give a brown oil. The residue was purified by silica gel chromatography (1:1 to 2:1 ethyl acetate-hexanes) to give the desired diol (**4**) as a white solid (30.9 mg, 34% overall). *R*_f 0.52 (1:2 hexanes/EtOAc); ¹H NMR (600 MHz, [D₆]DMSO) δ 10.8 (br s, 1H, NH), 7.59 (d, *J* = 12 Hz, 1H, ArH), 7.31 (d, *J* = 6 Hz, 2H, ArH), 7.27 (d, *J* = 6 Hz, 1H, ArH), 7.22 (d, *J* = 12 Hz, 2H, ArH), 7.08 (s, 1H, ArH), 7.03 (t, *J* = 6 Hz, 1H, ArH), 6.92 (t, *J* = 6 Hz, 1H, ArH), 6.76 (d, *J* = 6 Hz, 1H, ArH), 4.87 (AB, *J* = 12 Hz, *J*_{ab} = 36 Hz, 2H, PhCH₂O), 4.53 (d, *J* = 6 Hz, 1H, NH amide), 4.19 (m, 1H, CH_a), 3.44 (br s, 1H, OH), 3.33 (s, 1H, CHOH), 2.86 (dd, *J* = 18, 6 Hz, 1H, CH_βH_β(Trp)), 2.78 (dd, *J* = 18, 6 Hz, 1H,

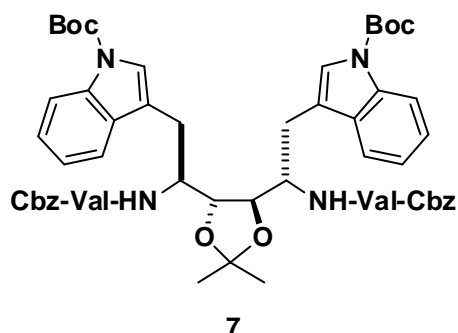
$\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$; ^{13}C NMR (150 MHz, DMSO- d_6) δ 155.9, 137.5, 136.1, 128.3, 128.1, 127.7, 127.6, 127.5, 127.4, 123.2, 120.6, 118.7, 118.1, 111.6, 111.1, 72.6, 64.9, 52.5, 27.9; HRMS (ESI) calcd for $\text{C}_{38}\text{H}_{39}\text{N}_4\text{O}_6$ $[M+\text{H}]^+$: 647.2870, found: 647.2812.



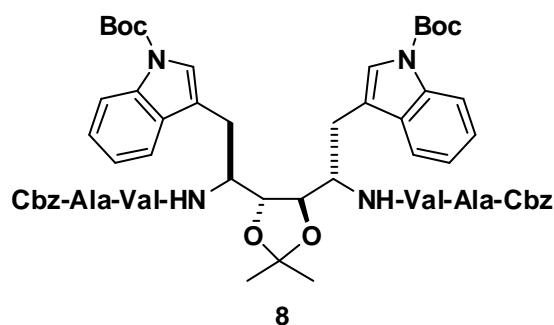
(2S,3R,4R,5S)-2,5-Bis(benzyloxycarbonylamino)-1,6-di(3-indolyl)-3,4-O-isopropylidenehexane (5). 2,2-Dimethoxypropane (2744 μL , 22 mmol) and *p*-TsOH (41 mg, 0.22 mmol) were added to a suspended solution of **4** (1403 mg, 2.2 mmol) in CHCl_3 (50 mL). After stirring for 30 min, the mixture became light brown clear solution, and it was stirred for another 1.5 h. At this time, TLC showed completion of the reaction, and the reaction mixture was washed with sat. aq. NaHCO_3 and extracted with CHCl_3 (3x). The organic layers were collected and washed with brine, dried over MgSO_4 , filtered, and concentrated under reduced pressure. The residue was loaded on a column of silica gel and the column was eluted with 25% ? 30% ? 40% ? 66% EtOAc/hexanes to afford acetonide **5** as a white solid (1171.2 mg, 79%). R_f 0.50 (1:1 hexanes/EtOAc); ^1H NMR (600 MHz, CDCl_3) δ 7.89 (br s, 1H, ArH), 7.59 (d, $J = 6$ Hz, 1H, ArH), 7.33 (d, $J = 6$ Hz, 2H, ArH), 7.29 (d, $J = 6$ Hz, 2H, ArH), 7.22 (d, $J = 6$ Hz, 2H, ArH), 7.15 (t, $J = 6$ Hz, 1H, ArH), 7.08 (t, $J = 6$ Hz, 1H, ArH), 6.98 (s, 1H, ArH), 5.00 (d, $J = 6$ Hz, 1H, NH amide), 4.90 (d, $J = 12$ Hz, 1H, OCHH'Ph), 4.76 (d, $J = 12$ Hz, 1H, OCHH'Ph), 4.07 (m, 1H, CH_a), 3.85 (s, 1H, CHOCCH₃), 2.99 (dd, $J = 18, 6$ Hz, 1H, CH_βH'_{β(Trp)}), 2.92 (dd, $J = 18, 6$ Hz, 1H, CH_βH'_{β(Trp)}), 1.39 (s, 3H, OCCH₃); ^{13}C NMR (150 MHz, CDCl_3) δ 156.2, 136.8, 136.2, 128.7, 128.2, 122.8, 122.1, 119.5, 119.0, 111.9, 111.1, 109.1, 77.9, 66.7, 50.5, 29.4, 27.3; HRMS (ESI) calcd for $\text{C}_{41}\text{H}_{42}\text{N}_4\text{O}_6\text{Na}$ $[M+\text{Na}]^+$: 709.3002, found: 709.2783.



(2S,3R,4R,5S)-2,5-Bis(benzyloxycarbonylamino)-1,6-di(3-N-Boc-indolyl)-3,4-O-isopropylidenehexane (6). To a stirred solution of **5** (1144.7 mg, 1.67 mmol) in CH₃CN (50 mL) at 23 °C were added di-*tert*-butyl dicarbonate (735 μL, 3.33 mmol) and 4-(dimethylamino)pyridine (41 mg, 0.33 mmol). After 24 h of stirring, TLC analysis indicated that a minor amount of starting material was still present. Additional di-*tert*-butyl dicarbonate (2.0 equiv) and 4-(dimethylamino)pyridine (0.1 equiv) were added, and the mixture was stirred for another 2 h at 23 °C. The mixture was diluted with ethyl acetate (100 mL) and washed with saturated aqueous sodium hydrogen carbonate. After extraction with EtOAc, the organic layer was washed with 10% KHSO₄, brine, dried over MgSO₄, filtered, and the solvent was removed. Silica gel chromatography with 4:1 hexanes/ethyl acetate as eluant delivered the desired product (**6**) as a white solid (986.9 mg, 67%). *R*_f 0.56 (2:1 hexanes/EtOAc); ¹H NMR (600 MHz, CDCl₃) *δ* 8.15 (br s, 1H, ArH), 7.57 (d, *J* = 7.7 Hz, 1H, ArH), 7.48 (s, 1H, ArH), 7.30-7.16 (m, 7H, ArH), 4.96 (AB, *J* = 12.3 Hz, *ν*_{ab} = 22.7 Hz, 2H, PhCH₂O), 4.72 (d, *J* = 12.2 Hz, 1H, NH), 4.15 (m, 1H, CH_a), 3.77 (s, 1H, CHOCCH₃), 2.93 (dd, *J* = 14.6, 8 Hz, 1H, CH_βH_β(Trp)), 2.85 (dd, *J* = 14.6, 8 Hz, 1H, CH_βH_β(Trp)), 1.63 (s, 9H, C(CH₃)₃), 1.39 (s, 3H, OCCH₃); ¹³C NMR (150 MHz, CDCl₃) *δ* 155.9, 149.7, 136.5, 135.5, 130.9, 128.5, 128.1, 128.0, 124.3, 124.2, 122.5, 119.2, 116.4, 115.2, 108.9, 83.3, 77.9, 66.7, 49.6, 29.1, 28.3, 27.2; HRMS (ESI) calcd for C₅₁H₅₈N₄NaO₁₀ [*M*+Na]⁺: 909.4051, found: 909.3993.

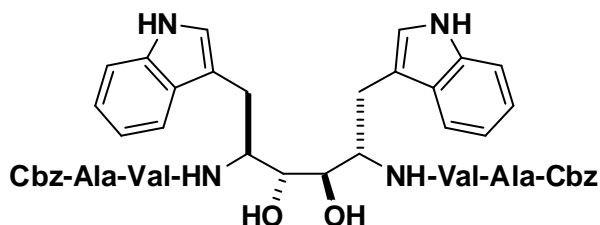


(2S,3R,4R,5S)-2,5-Bis(N-benzyloxycarbonyl-L-Val)-1,6-di(3-N-Boc-indolyl)-3,4-O-isopropylidenehexane (7). Cbz group of **6** (271.9 mg, 0.3 mmol) was deprotected by the general procedure A. The general procedure B was followed using the obtained diamine (186.1 mg, 0.3 mmol), Cbz-Val-OH (166 mg, 0.66 mmol), HBTU (285 mg, 0.75 mmol), and DIEA (397 μ L, 2.4 mmol). The reaction mixture was stirred for 13.5 h. Silica gel column chromatography (3:1 hexanes-ethyl acetate) gave the desired product (**7**) as a white solid (260.3 mg, 78% overall). R_f 0.34 (2:1 hexanes/EtOAc); ^1H NMR (600 MHz, CDCl_3) δ 8.09 (br s, 1H, ArH), 7.58 (d, $J = 6$ Hz, 1H, ArH), 7.44 (s, 1H, ArH), 7.37-7.31 (m, 5H, ArH), 7.28 (m, 1H, ArH), 7.22 (m, 1H, ArH), 6.20 (d, $J = 12$ Hz, 1H, NH), 5.05 (AB, $J = 12$ Hz, $\tau_{ab} = 36$ Hz, 2H, PhCH_2O), 5.05 (hidden within PhCH_2O , 1H, NH), 4.34 (dd, $J = 18, 6$ Hz, 1H, CH_a), 3.76 (m, 1H, CH_a), 3.70 (s, 1H, CHOCCCH_3), 2.95 (dd, $J = 12, 6$ Hz, 1H, $\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$), 2.87 (dd, $J = 12, 6$ Hz, 1H, $\text{CH}_\beta\text{H}'_{\beta(\text{Trp})}$), 1.88 (m, 1H, $\text{CH}_\beta(\text{Val})$), 1.64 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.38 (s, 3H, OCCH_3), 0.66 (d, $J = 6$ Hz, 3H, $\text{CH}_{3?}(\text{Val})$), 0.50 (d, $J = 6$ Hz, 3H, $\text{CH}_{3?}(\text{Val})$); ^{13}C NMR (150 MHz, CDCl_3) δ 170.7, 156.4, 149.7, 136.3, 135.5, 130.9, 128.7, 128.4, 128.2, 124.5, 124.3, 122.7, 119.1, 116.2, 115.3, 108.9, 83.5, 77.4, 67.2, 60.6, 47.6, 30.2, 28.8, 28.3, 27.2, 19.0, 17.0; HRMS (ESI) calcd for $\text{C}_{61}\text{H}_{77}\text{N}_6\text{O}_{12}$ $[M+\text{H}]^+$: 1085.5599, found: 1085.5614.



(2S,3R,4R,5S)-2,5-Bis(N-benzyloxycarbonyl-L-Ala-L-Val)-1,6-di(3-N-Boc-indolyl)-3,4-O-isopropylidenehexane (8). Cbz group of **7** (517.6 mg, 0.477 mmol) was de-

protected by the general procedure A. The general procedure B was followed using the obtained diamine (381.6 mg, 0.47 mmol), Cbz-Ala-OH (229 mg, 1.03 mmol), HBTU (443 mg, 1.17 mmol), and DIEA (618 μ L, 3.74 mmol). The reaction mixture was stirred for 14.5 h. Silica gel column chromatography (40% ? 50% ? 60% EtOAc/hexanes) gave the desired product (**8**) as a white solid (571.1 mg, 98% overall). R_f 0.36 (1:1 hexanes/EtOAc); ^1H NMR (600 MHz, CDCl_3) δ 8.06 (br s, 1H, ArH), 7.58 (d, J = 12 Hz, 1H, ArH), 7.42 (s, 1H, ArH), 7.34-7.26 (m, 6H, ArH), 7.21 (t, J = 6 Hz, 1H, ArH), 6.40 (br s, 1H, NH), 6.26 (d, J = 12 Hz, 1H, NH), 5.36 (d, J = 6 Hz, 1H, NH), 5.08 (s, 2H, PhCH_2O), 4.33 (d, J = 6 Hz, 1H, CH_a), 4.15 (t, J = 6 Hz, 1H, CH_a), 3.94 (t, J = 6 Hz, 1H, CH_a), 3.67 (s, 1H, CHOCCH_3), 2.93-2.84 (m, 2H, $\text{CH}_{2\beta(\text{Trp})}$), 1.86 (br s, 1H, $\text{CH}_{\beta(\text{Val})}$), 1.64 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.38 (s, 3H, OCCH_3), 1.29 (d, J = 12 Hz, 3H, $\text{CH}_{3\beta(\text{Ala})}$), 0.59 (d, J = 6 Hz, 3H, $\text{CH}_{3\gamma(\text{Val})}$), 0.55 (d, J = 6 Hz, 3H, $\text{CH}_{3\gamma'(\text{Val})}$); ^{13}C NMR (150 MHz, CDCl_3) δ 172.4, 170.3, 156.2, 149.8, 136.2, 135.5, 130.9, 128.7, 128.4, 128.3, 124.5, 124.3, 122.7, 119.4, 116.3, 115.2, 108.9, 83.6, 77.2, 67.3, 58.9, 50.8, 47.7, 30.2, 28.7, 28.4, 27.2, 19.0, 18.5, 17.5; HRMS (ESI) calcd for $\text{C}_{67}\text{H}_{87}\text{N}_8\text{O}_{14}$ $[M+\text{H}]^+$: 1227.6342, found: 1227.6254.



9

(2S,3R,4R,5S)-2,5-Bis(N-benzyloxycarbonyl-L-Ala-L-Val)-3,4-dihydroxy-1,6-di(3-indolyl)hexane (9). Compound **8** (46.4 mg, 0.038 mmol) was dissolved in 4 N HCl/dioxane/ H_2O (10 mL), and the mixture was stirred at 23 $^\circ\text{C}$ for 14 h, at which time no starting material could be detected by TLC analysis. The solvent was evaporated, and the residue was partitioned between sat. aq. NaHCO_3 and EtOAc. The water layer was extracted with EtOAc (3x), and the organic layers were collected. After washing with brine, dried (MgSO_4), filtered, and concentration, the crude was purified by SiO_2 column (3% ? 10% $\text{MeOH-CH}_2\text{Cl}_2$) to give the desired product **9** as a white solid (15.9 mg, 43%). R_f 0.33 (9:1 $\text{CH}_2\text{Cl}_2\text{-MeOH}$); $[\alpha]_{\text{D}}^{20}$ -64.9 $^\circ$ (c 0.00325, MeOH); FTIR (neat, ATR, cm^{-1}) 3306 (w, OH), 2927 (w), 2361 (w), 1699 (m, C=O), 1651 (s, C=O),

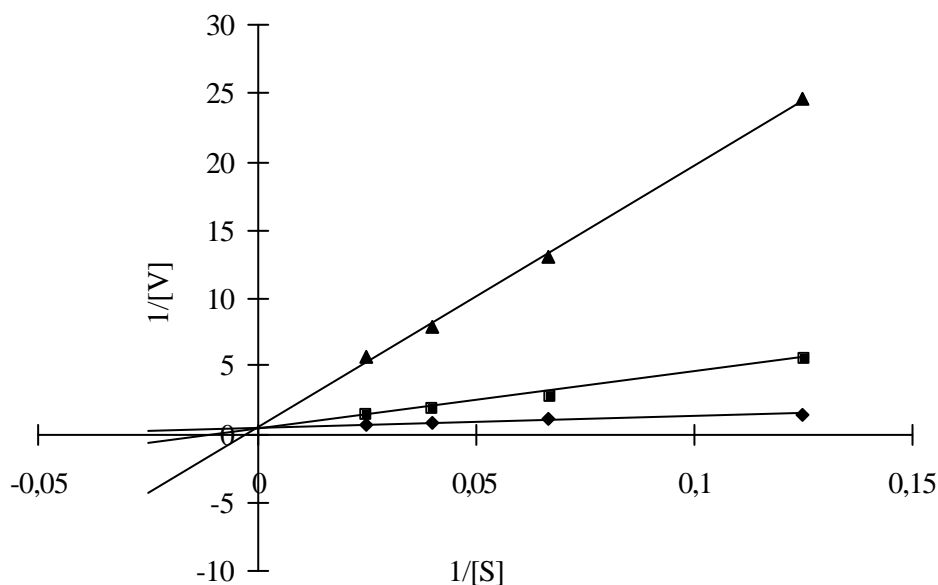
1645 (s, C=O), 1520 (s), 1506 (s), 1456 (m), 1339 (w), 1235 (m), 1071 (m); ^1H NMR (600 MHz, CD_3OD) δ 7.63 (d, $J = 12$ Hz, 1H, ArH), 7.35-7.26 (m, 6H, ArH), 7.03 (t, $J = 6$ Hz, 2H, ArH), 6.96 (t, $J = 6$ Hz, 1H, ArH), 5.07 (AB, $J = 12$ Hz, $J_{\text{ab}} = 18$ Hz, 2H, PhCH_2O), 4.52 (br s, 1H, CH_a), 4.14 (q, $J = 6$ Hz, 1H, CH_a), 4.00 (d, $J = 6$ Hz, 1H, CH_a), 3.58 (s, 1H, CHOH), 2.99 (m, 2H, $\text{CH}_{2\beta}(\text{Trp})$), 1.82 (m, 1H, $\text{CH}_\beta(\text{Val})$), 1.28 (d, $J = 6$ Hz, 3H, $\text{CH}_{3\beta}(\text{Ala})$), 0.71 (d, $J = 6$ Hz, 3H, $\text{CH}_{3\gamma}(\text{Val})$), 0.66 (d, $J = 6$ Hz, 3H, $\text{CH}_{3\gamma}(\text{Val})$); ^{13}C NMR (150 MHz, CD_3OD) δ 175.7, 173.2, 158.5, 138.2, 138.1, 129.6, 129.4, 129.2, 129.1, 124.3, 122.2, 119.9, 119.7, 112.5, 112.2, 74.1, 67.9, 60.7, 52.4, 52.2, 31.8, 29.2, 19.9, 18.5, 18.3; HRMS (ESI) calcd for $\text{C}_{54}\text{H}_{67}\text{N}_8\text{O}_{10}$ $[M+H]^+$: 987.4980, found: 987.5106.

SARS-CoV 3CL Protease Inhibition Assay

The cloning, expression, purification, and kinetics determination of this enzyme, the preparation of buffers and the utilization of the fluorogenic peptide substrate were performed by methods described in full detail elsewhere.^[S1] The illustrations on the double reciprocal plots of **4** and **9** for inhibition of SARS-CoV 3CL protease are provided in the Figure S1.

A)

Compound	IC_{50} [μM]	K_i [μM]
4	4	0.340 ± 0.042



B)

Compound	IC ₅₀ [μM]	K _i [μM]
9	0.8	0.073 ± 0.008

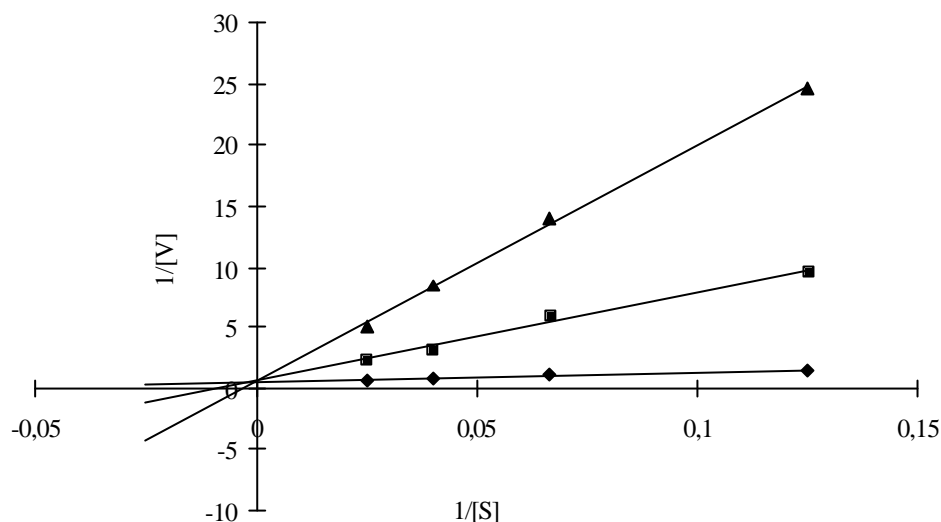


Figure S1. The Lineweaver-Burk plot for competitive inhibition of SARS-CoV 3CL protease exhibited by **4** (A) and **9** (B). The enzyme activities were measured using 8-40 μM fluorogenic substrate in the absence (?) or presence of 1 × IC₅₀ (!) and 2 × IC₅₀ (?) inhibitor. The data were fitted with Equation (1) by using the KinetAsyst II program to obtain the K_i values.

$$1/V = K_m/V_m (1 + [I]/K_i)1/[S] + 1/V_m \quad (1)$$

K_m is Michaelis constant of the substrate, V_m is maximal velocity, K_i is inhibition constant, and [I] and [S] represent the inhibitor and substrate concentrations in the reaction mixture, respectively.

IV. Computer Modeling and Crystallization

1. Preliminary Modeling Prediction

The crystal structure of SARS-CoV 3CL protease (PDB ID code 1UK4:A) was used as docking target. Di-peptide structures with CH₃ capped at both N- and C-terminus were generated by perl script and Accelrys Insight II biopolymer. 20x20 di-peptides (all 20 nature amino acid sidechains were used at each of the residue positions of the dipeptide) were docked onto the target protein with LIGANDFIT (Accelrys DISCOVERY STUDIO 1.2 SBD) program using default parameters. The resultant ligand-protein complex structures were ranked with the LIGANDFIT utility (LIGSCORE2).

2. Modeling Refinement

We docked compound **4** to SARS-CoV 3CL protease structure solved with the ligand in the co-solvent using three docking programs (DOCK 4.0.2,^[S2] GOLD 3.0.1^[S3, S4] and AutoDOCK 3.0.5^[S5]). One hundred 3D structures of compound **4** were generated with the Genetic algorithm (GA) and random search procedure in the SYBYL 7.0.1 program^[S6] – 50 conformations with the lowest conformal energies were derived from each of the algorithms with default settings. As described in previous work on SARS-CoV 3CL protease,^[S7, S8] Kollmann-all atom charges^[S9] were assigned to the protein atoms, and Gasteiger-Hückel^[S10-S12] charges were assigned to inhibitor atoms using the SYBYL 7.0.1 program.^[S2] The atom charge assignment was applied in the docking task with all three software package except Kollmann-United atom charges^[S9] were assigned to the protein atoms using Autodock.

DOCK 4.0.2 program^[S2] was used to dock the 100 conformations of compound **4** as described above. The ligand conformations were optimized in the active site first by the search of the anchor fragment orientations and then by optimizing the torsion angles connecting the fragments of the compound with the computational procedures implemented in the DOCK program.^[S2] Fifty configurations and a maximum of 3000 anchor orientations for each initial conformation were generated, and all of the docked configurations were energy minimized by 150 iterations. All other parameters were default settings. The final model complex structures were energy-ranked based on electrostatic and Van der Waals terms. Only the top 20 model complex structures were chosen for manual graphic analysis.

The same set of 100 conformation of compound **4** were docked to the target protein with GOLD (Genetic Optimization for Ligand Docking) (Cambridge Crystallographic Data Center (CCDC), version 3.0.1), which utilizes a genetic algorithm described by Jones et al.^[S3, S4] The most extensive search settings were applied so as to explore the largest conformational space for the ligand-protein complex structure. The automatic searching efficiency was set at 200% along with other default parameters. Each of the initial conformation was docked to the target protein with 100 independent searches, and the top energy-rank structure from each of the searches was saved for manual graphic analysis.

AutoDock version 3.0.5^[S5] was used for the docking simulation. We selected the Lamarckian genetic algorithm (LGA) for the ligand-protein complex structure search

because it was expected to have enhanced performance in comparison with simple genetic algorithm ^[S5]. The docking parameters were as follows: the number of individuals in population was increased to 500 (ga_pop_size); the maximum number of energy evaluations was increased to 2500000 per run (ga_num_evals); the maximum number of generations in the Lamarckian genetic algorithm was increased to 90000 (ga_num_generations); elitism was set at 1; mutation rate was set at 0.02; crossover rate was set at 0.80; local search rate was set at 0.06. All other parameters were maintained at their default settings. As in the docking task with GOLD, 100 independent searches were carried out for each of the initial conformation, and the top energy-ranked structure from each of the searches was saved for manual graphic analysis. The docking tasks with DOCK 4.02 ^[S2], GOLD 3.01 ^[S3, S4] and Autodock 3.0.5^[S5] programs were distributed on a 40-processor Linux cluster with Intel(R) Xeon(TM) CPU (3.00 GHz) for 10 days.

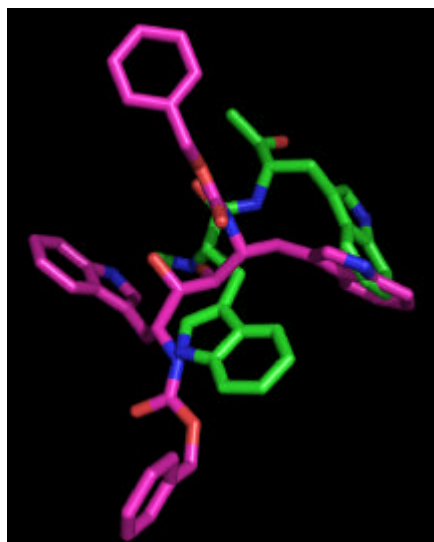


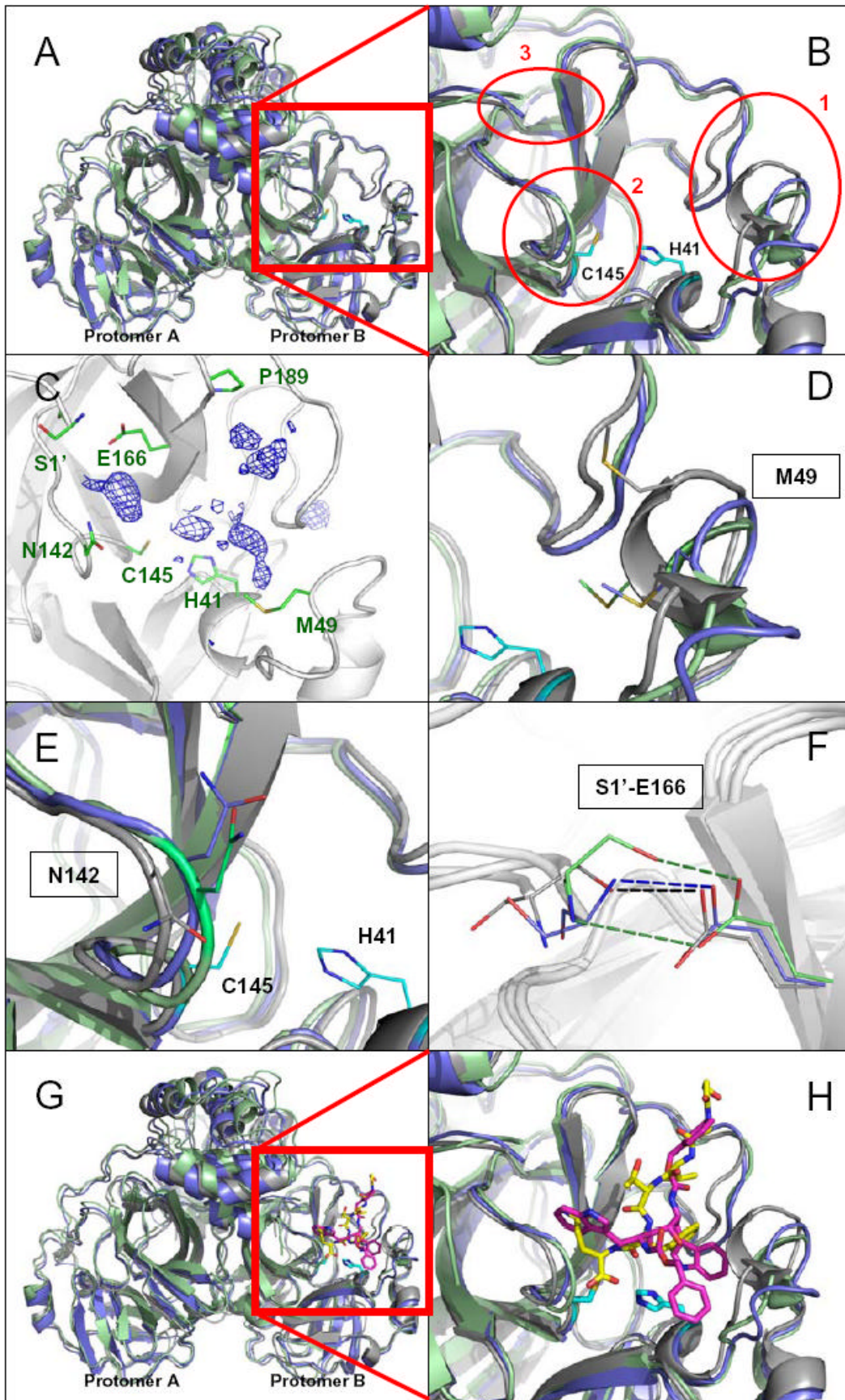
Figure S2. The preliminary modeling Trp-Trp (green) and the refined modeling of compound 4 (magenta). The P1 moieties of the two ligands occupied the S1 pocket in the protein with similar conformation, while the P2 moieties of the two ligands showed significant discrepancies in conformation.

3. Crystallization Condition

SARS 3CL protease was prepared as previous report ^[S13]. The protein was mixed with the freshly prepared inhibitor dissolved in dimethylsulfoxide (DMSO) in the molar ratio 1:10 and preincubated for at least 15 min. prior to crystallization. The complex was co-crystallized by the sitting drop diffusion method at 18°C. The best crystals were ob

-tained by mixing 2 μL of 3CL^{pro} protein-inhibitor solution with 2 μL of the reservoir solution (0.1 M MES, pH 6.5 and 16 % (v/v) DMSO, 2 mM DTT and 10 % (v/v) PEG6000) onto a sitting drop post, equilibrated with 0.5 mL of the reservoir solution for three days. Crystals used for data collection were rinsed with reservoir and cryo-cooled in liquid nitrogen. The X-ray data was collected at Taiwan beamline BL12B2 in SPring-8 (Ja-an). Data integration and scaling were performed by using the programs HKL-2000.^[S14] Data were reduced in the $P2_1$ space group. The unit cell dimensions are $a = 52.2 \text{ \AA}$, $b = 96.4 \text{ \AA}$, $c = 96.4 \text{ \AA}$, $\beta = 103^\circ$ with 1.8 \AA resolution. There is one dimer in an asymmetric unit.

Figure S3. (over) A) Superposition of *apo* form, product bound form and compound **4** complex 3CL proteases's dimer. B) an enlarged view of the active site in protomer B. Red circle 1 showed the variety of the loop region, aa 45~52; red circle 2 showed the variety of the residue, Asn142; red circle 3 showed the variety of the interactions between N-finger (Ser1) in protomer A and Glu166 in protomer B. C) important residues in the active site shown in green, and differential density maps in the active site of 3CL^{pro}-compound **4** complex. D) an enlarged view of red circle 1, and the major difference is Met49. E) an enlarged view of red circle 2, and the major difference is Asn142. F) an enlarged view of red circle 3. In product bound form structure (green), two hydrogen bonds existed between Ser1' in protomer A and Glu166 in protomer B. In *apo* form structure (inactive form; gray), one hydrogen bond existed between Ser1' in protomer A and Glu 166 in protomer B. In compound **4** complexed structure (blue), one hydrogen bond existed between Ser1' in protomer A and Glu166 in protomer B. G) superposition of compound **4** and product (SGVTFQ) in the active site of protomer B. H) an enlarged view of the active site in protomer B.



Protease of SARS 3CL^{pro}-compound 4 complex Compound 4
 Protease of SARS 3CL^{pro} (product bound form) Product (SGVTFQ)
 Protease of SARS 3CL^{pro} (apo form)

- [S1] C.-J. Kuo, Y.-H. Chi, J. T.-A. Hsu, P.-H. Liang, *Biochem. Biophys. Res. Commun.* **2004**, *318*, 862-867.
- [S2] T. J. A. Ewing, I. D. Kuntz, *J. Comput. Chem.* **1997**, *18*, 1175-1189.
- [S3] G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43-53.
- [S4] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727-748.
- [S5] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639-1662.
- [S6] SYBYL 7.0.1; The Tripos Associates: St. Louis, MO.
- [S7] B. Xiong, C. S. Gui, X. Y. Xu, C. Luo, J. Chen, H. B. Luo, L. L. Chen, G. W. Li, T. Sun, C. Y. Yu, L. D. Yue, W. H. Duan, J. K. Shen, L. Qin, T. L. Shi, Y. X. Li, K. X. Chen, X. M. Luo, X. Shen, J. H. Shen, H. L. Jiang, *Acta Pharm. Sin.* **2003**, *24*, 497-504.
- [S8] K.-C. Tsai, S.-Y. Chen, P.-H. Liang, I.-L. Lu, N. Mahindroo, H.-P. Hsieh, Y.-S. Chao, L. Liu, D. Liu, W. Lien, T.-H. Lin, S.-Y. Wu, *J. Med. Chem.* **2006**, *49*, 3485-3495.
- [S9] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- [S10] J. Gasteiger, M. Marsili, *Tetrahedron* **1980**, *36*, 3219-3228.
- [S11] M. Marsili, J. Gasteiger, *Croat. Chem. Acta* **1980**, *53*, 601-614.
- [S12] W. P. Purcell, J. A. Singer, *J. Chem. Eng. Data* **1967**, *12*, 235-246.
- [S13] M.-F. Hsu, C.-J. Kuo, K.-T. Chang, H.-C. Chang, C.-C. Chou, T.-P. Ko, H.-L. Shr, G.-G. Chang, A. H.-J. Wang, P.-H. Liang, *J. Biol. Chem.* **2005**, *280*, 31257-31266.
- [S14] Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307-326.