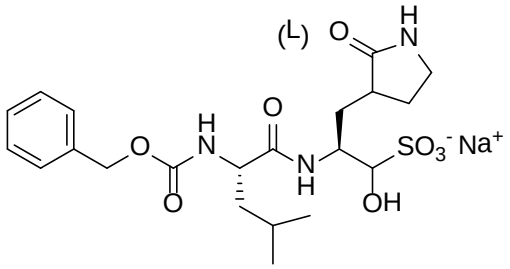
Supplemental Material

Synthesis of GC376

GC376 (compound 28). Compound **23** (GC373) (0.50 g, 1.24 mmol), sodium bisulfite (0.119 g, 1.12 mmol), ethyl acetate (2 mL), ethanol (1 mL) and water (0.40 mL) were combined and heated to 40 ^{o}C using a water bath. The reaction mixture was stirred for 2 hrs and then allowed to cool to ambient temperature. The solution was filtered and washed with ethanol (5 mL). The filtrate was dried over sodium sulfate, filtered, and concentrated leaving a yellow oil, which was treated with ethyl ether (2 x 3 mL) to give a yellow solid (0.47 g, 74.8 % yield), mp 135-137 ^{o}C. ^{1}H NMR ($CDCl_3$): δ 0.80-0.95 (d, 6H), 1.38-2.00 (m, 6H), 2.00-2.22 (d, 2H), 2.20-2.50 (m, 2H), 3.20-3.40 (m, 2H), 4.20-4.39 (m, 1H), 4.40-4.50 (m, 1H), 5.10 (s, 2H), 5.25 (d, 1H), 5.40 (d, 1H), 7.20-7.41 (m, 5H).



28

MATERIALS AND METHODS

X-ray Crystallography: Crystallization and Data Collection

Purified NV 3CLpro, PV 3Cpro and TGEV 3CLpro concentrated to 10 mg/mL in 100 mM NaCl, 50 mM PBS pH 7.2, 1 mM DTT were used to prepare complexes with GC376. All

crystallization screening was conducted in Compact Jr. (Emerald Biosystems) sitting drop vapor diffusion plates at 20 ^{o}C using equal volumes of protein and crystallization solution equilibrated against 75 μL of the latter. A 100 mM stock solution of GC376 was prepared in DMSO and complexes with the proteases were prepared as follows. NV 3CLpro-GC376: 290 μL of NV 3CLpro (0.48 mM) was mixed with 10 μL of GC376 (3.3 mM) and PV 3Cpro-GC376: 392 μL of PV 3Cpro (0.48 mM) was mixed with 8 μL of GC376 (2.0 mM). TGEV 3CLpro-GC376: 490 μL of TGEV 3CLpro (0.48 mM) was mixed with 10 μL of GC376 (2.0 mM). The complexes were incubated on ice for 1 hr and loaded onto a Superdex 75 10/300 GL column equilibrated with 100 mM NaCl, 20 mM Tris pH 8.0. Then the elution fractions were pooled and concentrated in a Vivaspin-20 concentrator (MWCO=10 kDa) to 14, 9.7, and 10.7 mg/mL for NV 3CLpro, PV 3Cpro, and TGEV 3CLpro, respectively, for crystallization.

Crystals were obtained from the following conditions. NV 3CLpro: Apo crystals, displaying a prismatic morphology, were obtained in 24 hrs from Wizard 3 screen (Emerald Biosystems) condition #10 (20% [w/v] PEG 3350, 100 mM sodium thiocyanate). Needle shaped crystals of the NV3CLpro-GC376 complex were obtained in 24 hrs from the Wizard 4 screen (Emerald Biosystems) condition #25 (30% [w/v] PEG 2000MME, 150 mM sodium bromide). PV 3Cpro-GC376: Plate shaped crystals were obtained in 24 hrs from Wizard 3 screen (Emerald Biosystems) condition #47 (30% [w/v] PEG 5000MME, 100 mM MES pH 6.5, 200 mM ammonium sulfate). TGEV 3CLpro-GC376: A cluster of plate shaped crystals were obtained in 48 hrs from Wizard 3 screen (Emerald Biosystems) condition #1 (20% [w/v] PEG 3350, 200 mM sodium acetate). All crystals, except those of the NV 3CLpro-GC376 complex, were transferred to a solution containing 80% crystallization and 20% PEG 400 and frozen in liquid nitrogen for data collection. For the NV 3CLpro-GC376 complex, 20% PEG 200 was used as the cryoprotectant. X-ray diffraction data were collected at the Advanced Photon Source beamline 17-ID using a Dectris Pilatus 6M pixel array detector.

Structure Solution and Refinement

For all structures, the following software was used unless specified otherwise. Intensities were integrated using XDS (7) and the Laue class check and data scaling were performed with Pointless and Scala (6). Structure solution was conducted by molecular replacement with Molrep (12) for the NV 3CLpro structures and Phaser (9) via the Phenix (1) interface for all other structures. Refinement and manual model building were conducted with Phenix and Coot (4) respectively. TLS refinement (10) was incorporated in the latter stages of refinement for the NV 3CLpro and PV 3Cpro structures, to model the anisotropic atomic displacement parameters of the protein atoms. Structure validation was conducted with Molprobity (2) and figures were prepared using the CCP4MG package (11). Disordered side chain atoms were truncated to the point where electron density could be observed. Relevant crystallographic data are provided in supporting table 1.

For apo NV 3CLpro, the highest probability Laue class was 2/*m* and space group $P2_1$ and the Matthew's coefficient (Vm) (8) and solvent content were estimated to be Vm=3.7 / 67.0% solvent and Vm=1.9 / 34.0% solvent for 1 and 2 molecules in the asymmetric unit, respectively. Molecular replacement searches for two molecules in the asymmetric unit were conducted in the space groups *P*2 and $P2_1$. A previously determined structure of NV 3CLpro (PDB: 2FYQ) was used as the search model. The top solution was found in the space group $P2_1$ that consisted of a non-crystallographic dimer. For the NV 3CLpro-GC376 complex, the highest probability Laue class was *mmm* and space group $P2_12_12_1$. The Matthew's coefficient (Vm) (8) and solvent content were estimated to be Vm=3.9 / 68.8% solvent and Vm=2.0 / 37.6% solvent for 1 and 2 molecules in the asymmetric unit, respectively. Molecular replacement search for two molecules in the asymmetric unit was conducted using the apo NVPro structure as a search model. The highest correlation coefficient (0.672) was obtained in the space group $P2_12_12_1$. Examination of the active site revealed prominent difference in electron density (Fo-Fc) in each

subunit greater than 3σ that was consistent with GC376. However, the bisulfite group appeared to have been removed and the inhibitor was covalently bound to Cys 139. The 6-membered aromatic ring of the inhibitor was disordered and could not be fit to the electron density maps due to disorder. Residues between Leu 122-Gly 133 of apo NV 3CLpro chain A and Leu 122-Asn 126 of chain B were disordered and could not be modeled as were the C-terminal residues from Gly 174 to Glu 181.

For the PV 3Cpro-GC376 complex, the highest probability Laue class was *mmm* and possible space groups $I2_12_12_1$ or *I*222. The Matthew's coefficient (Vm) (8) and solvent content were estimated to be Vm=2.2 / 43.0% solvent for 1 molecule in the asymmetric unit. Molecular replacement was conducted using an apo PV 3Cpro structure as the search model (PDB: 1L1N) and the top solution was found in the space group *I*222. Examination of the active site revealed prominent difference in electron density (Fo-Fc) greater than 3σ that was consistent with GC376 which was covalently bound to Cys 147. Two sulfate ions were included in the model along with a DTT molecule. The latter resides on a crystallographic 2-fold axis. Prominent electron density (Fo-Fc) was observed near the DTT molecule that appeared to be covalently connected. However, the identity of this electron density could not be conclusively confirmed and was not assigned.

For the TGEV 3CLpro-GC376 complex, the highest probability Laue class was 2/*m* and the most probable space group was $P2_1$. The Matthew's coefficient (Vm) (8) and solvent content were estimated to be Vm=3.1 / 60.2% solvent and Vm=2.3 / 46.9% solvent for 3 and 4 molecules in the asymmetric unit, respectively. Molecular replacement was conducted using a previously determined structure of 3CLpro of TGEV as the search model (PDB: 2AMP). The top solution was found in the space group $P2_1$ with 4 molecules in the asymmetric unit. Examination of the active site revealed prominent difference in electron density (Fo-Fc) in each subunit greater than 3σ that was consistent with GC376 covalently bound to Cys 144.

Table S1. Crystallographic data.

	Apo NV 3CLpro	NV 3CLpro: GC376	PV 3Cpro: GC376	TGEV 3CLpro: GC376
Data Collection				
Unit-cell parameters (Å, ^{o})	*a* = 37.24 *b* = 35.93 *c* = 112.97 β = 97.96	*a* = 37.65 *b* = 66.86 *c* = 125.11	*a* = 56.08 *b* = 58.61 *c* = 107.99	*a* = 60.66 *b* = 170.14 *c* = 66.13 β =113.32
Space group	$P2_1$	$P2_12_12_1$	*I*222	$P2_1$
Resolution (Å) *	111.89-1.50 (1.58-1.50)	125.11-1.65 (1.74-1.65)	53.99-1.60 (1.69-1.65)	170.14-2.25 (2.37-2.25)
Wavelength (Å)	1.0000	1.0000	1.0000	1.0000
Temperature (K)	100	100	100	100
Observed reflections	153,587	255,606	84,780	204,016
Unique reflections	47,402	38,969	23,731	57,334
$<I/\sigma(I)>$*	12.6 (2.1)	16.2 (2.9)	14.1 (2.3)	7.1 (2.0)
Completeness (%)*	99.1 (99.7)	100 (100)	99.5 (99.8)	98.6 (97.7)
Multiplicity*	3.2 (3.1)	6.6 (6.8)	3.6 (3.6)	3.6 (3.5)
R_{merge} (%)*, #	4.6 (56.5)	6.6 (66.5)	4.9 (53.6)	12.6 (60.4)
R_{meas} (%)*, ^	5.5 (68.0)	7.2 (72.0)	5.7 (63.0)	14.8 (71.6)
R_{pim} (%)*, ^	3.0 (37.3)	2.8 (27.4)	2.9 (32.6)	7.8 (38.1)
Refinement				
Resolution (Å)	36.57 – 1.50	45.68 – 1.65	40.52 – 1.60	55.71 – 2.25
Reflections (working/test)	42,936 / 4,442	35,204 / 3,694	21,498 / 2,228	54,368 / 2,907
R_{factor} / R_{free} (%)&	17.9 / 20.4	17.8 / 21.2	16.6 / 18.9	18.1 / 24.1
No. of atoms (protein/GC376/water)	2,510 / - / 186	2,466 / 46 / 158	1,382 / 29 / 130	9,087 / 116 / 297
Model Quality				
R.m.s deviations				
Bond lengths (Å)	0.014	0.014	0.009	0.010
Bond angles (^{o})	1.459	1.529	1.308	1.305
Average *B* factor (Å^2)				
All Atoms	24.9	26.4	22.7	31.9
Protein	24.5	26.1	21.8	31.8
GC376	-	24.4	21.2	32.0
Water	30.2	33.1	30.8	32.8
Coordinate error, maximum likelihood (Å)	0.41	0.44	0.22	0.34
Ramachandran Plot				
Most favored (%)	97.3	98.1	97.8	97.3
Additionally allowed (%)	2.7	1.9	2.2	2.4
PDB accession code	3UR6	3UR9	4DCD	4F49

***** Values in parenthesis are for the highest resolution shell.

$R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the intensity measured for the *i*th reflection and $\langle I(hkl)\rangle$ is the average intensity of all reflections with indices hkl.

& $R_{factor} = \Sigma_{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$; Rfree is calculated in an identical manner using 5% of randomly selected reflections that were not included in the refinement.

^ R_{meas} = redundancy-independent (multiplicity-weighted) R_{merge}(5, 6). R_{pim} = precision-indicating (multiplicity-weighted) R_{merge}(3, 13).

Appendix Figure Legends

Figure S1. Crystals of apo NV 3CLpro obtained from Wizard 3 #10 (A), NV 3CLpro-GC376 complex obtained from Wizard 4 #25 (B), PV 3Cpro-GC376 complex obtained from Wizard 3 #10 (C), and TGEV 3CLpro-GC376 complex obtained from Wizard 3 #1 (D).

Figure S1. **A.** Phylogenetic relationship of 3Cpro or 3CLpro of caliciviruses, picornaviruses and coronaviruses. Amino acid sequence alignment and the phylogenetic tree was made by Clustal omega and Phylip programs, respectively. **B**. Conserved residues of 3Cpro and 3CLpro of caliciviruses, picornaviruses and coronaviruses. Sequence alignment was made by Clustal omega. Amino acids for the triad (or dyad) were marked with asterisk (*) and other conserved amino acids in the active sites were marked with filled circle (•). For coronavirus 3CLpro, only domains 1 and 2 are shown in the figure.

Supplemental Movie 1.

Conformational changes of NV 3CLpro including the loop containing Gln 110 that occur during inhibitor binding. The hydrogen bonds that form between NV 3CLpro (tan) including Gln 110 and GC376 (grey) are indicated by dashed lines.

Supplemental Movie 2.

Conformational changes of PV 3Cpro including the loops containing loops containing Leu 127, Gly 128 and Gly 164 that occur during inhibitor binding. The hydrogen bonds that form between PV 3Cpro (pink) including Leu 127, Gly 128 and Gly 164, and GC376 (grey) are indicated by dashed lines.

References

1. **Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart.** 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **66:**213-21.
2. **Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson.** 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr **66:**12-21.
3. **Diederichs, K., and P. A. Karplus.** 1997. Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat Struct Biol **4:**269-75.
4. **Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan.** 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr **66:**486-501.
5. **Evans, P.** 2006. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr **62:**72-82.
6. **Evans, P. R.** 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr D Biol Crystallogr **67:**282-92.
7. **Kabsch, W.** 1988. Automatic indexing of rotation diffraction patterns. Journal of Applied Crystallography **21:**67-72.
8. **Matthews, B. W.** 1968. Solvent content of protein crystals. J Mol Biol **33:**491-7.
9. **McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read.** 2007. *Phaser* crystallographic software. J. Appl. Cryst. **40:**658-674.
10. **Painter, J., and E. A. Merritt.** 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr **62:**439-50.
11. **Potterton, L., S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis, and M. Noble.** 2004. Developments in the CCP4 molecular-graphics project. Acta Crystallogr D Biol Crystallogr **60:**2288-94.
12. **Vagin, A., and A. Teplyakov.** 2010. Molecular replacement with MOLREP. Acta Crystallographica Section D-Biological Crystallography **66:**22-25.
13. **Weiss, M. S.** 2001. Global indicators of X-ray data quality. Journal of Applied Crystallography **34:**130-135.

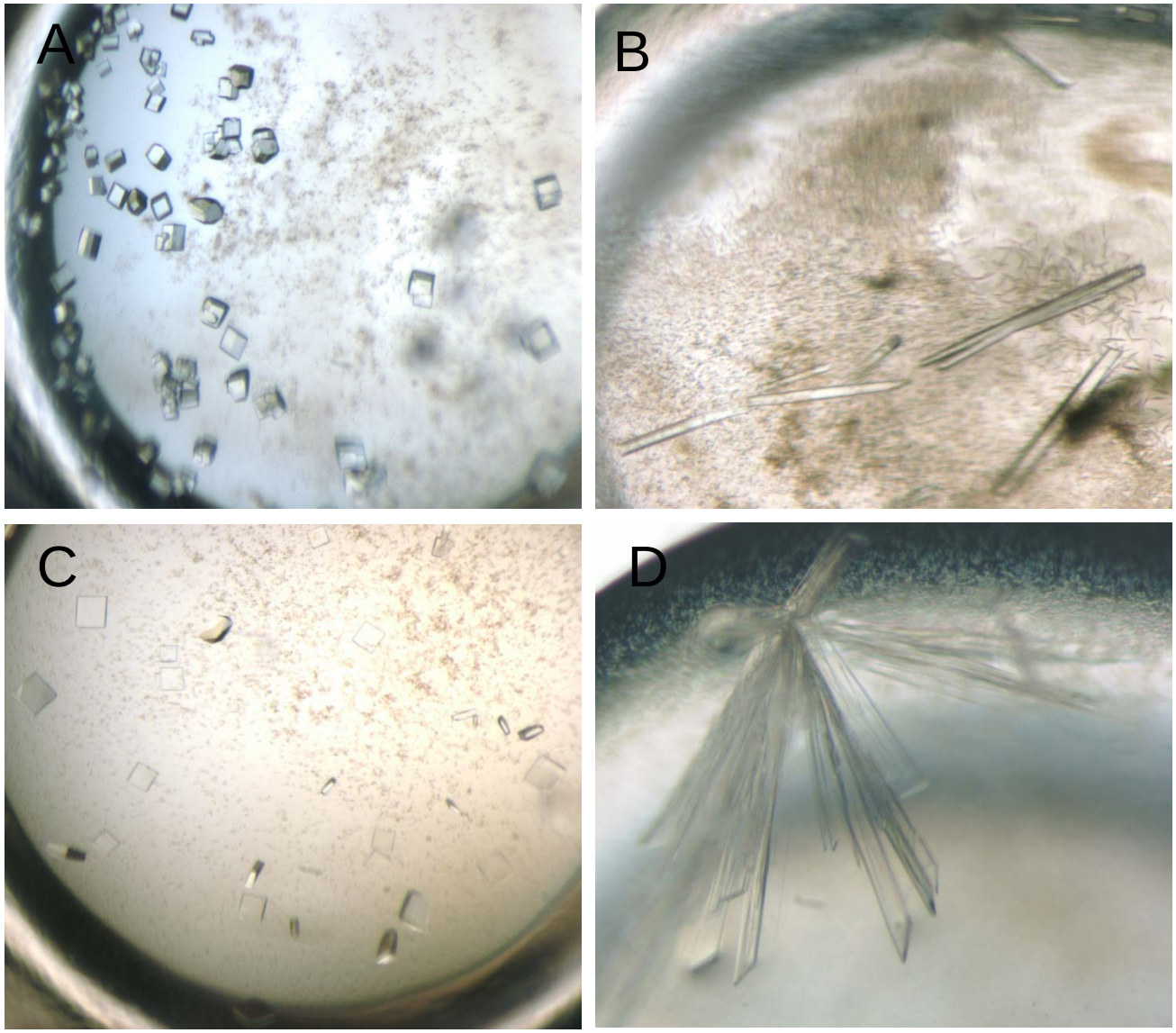


Figure S1. Crystals of apo NV 3CLpro obtained from Wizard 3 #10 (A), NV 3CLpro-GC376 complex obtained from Wizard 4 #25 (B), PV 3Cpro-GC376 complex obtained from Wizard 3 #10 (C), and TGEV 3CLpro-GC376 complex obtained from Wizard 3 #1 (D).

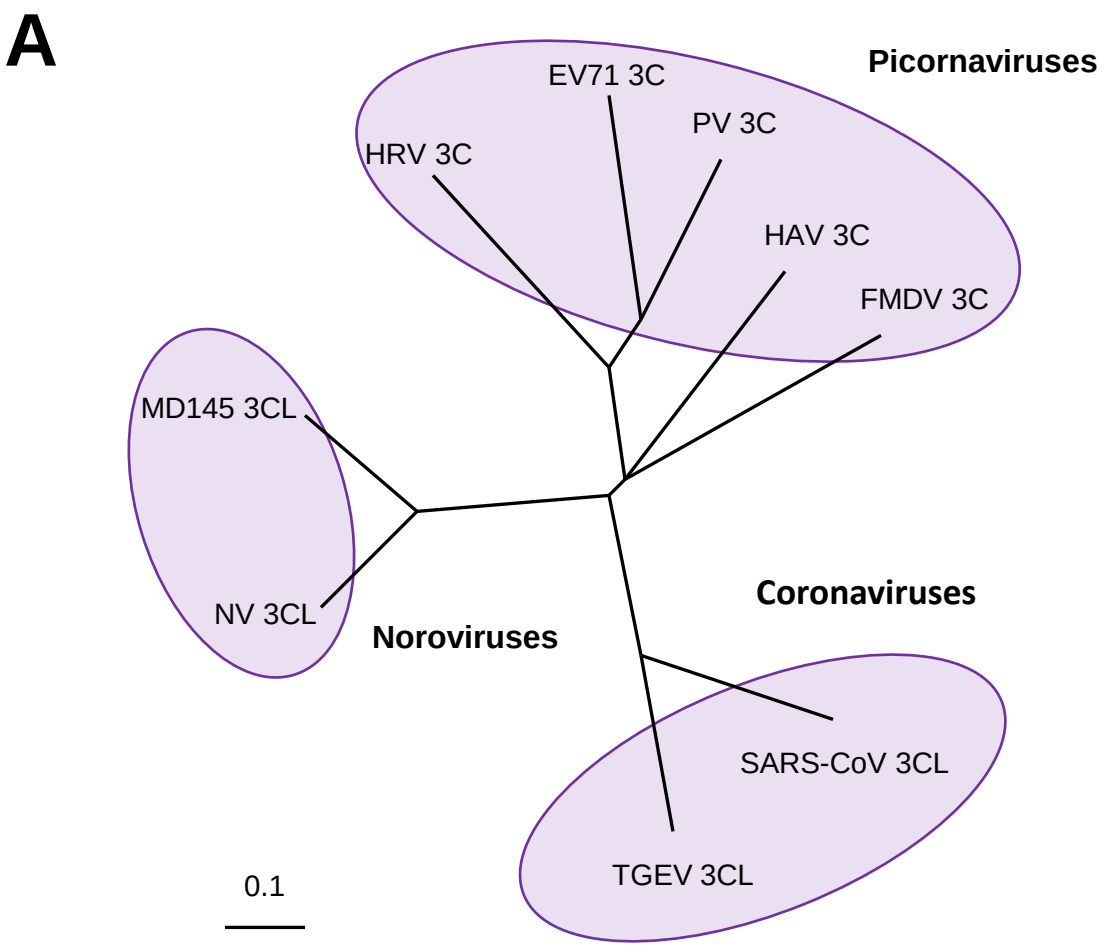


Figure S2. **A.** Phylogenetic relationship of 3Cpro or 3CLpro of caliciviruses, picornaviruses and coronaviruses. Amino acid sequence alignment and the phylogenetic tree was made by Clustal omega and Phylip programs, respectively

B

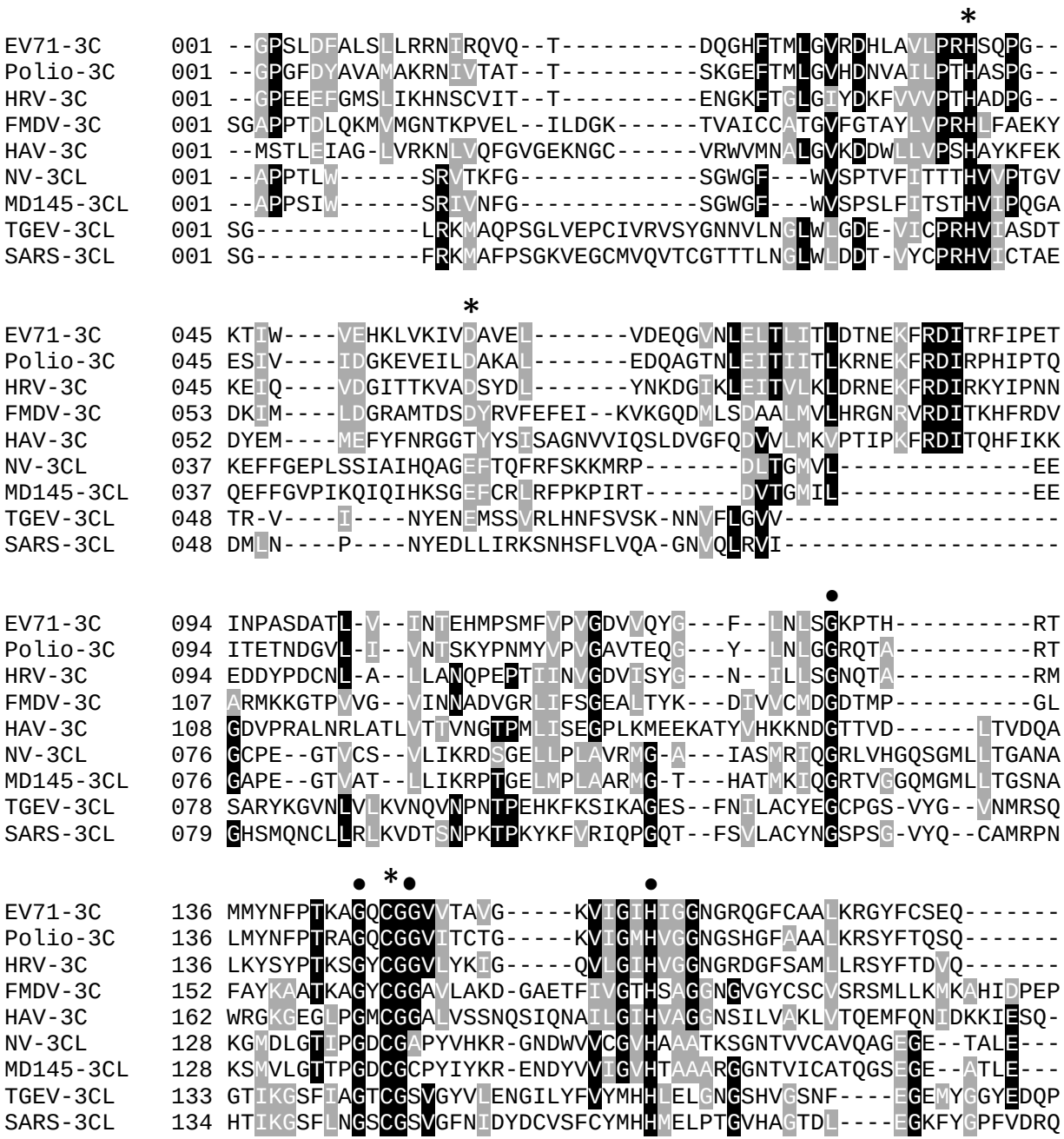


Figure S2. **B**. Conserved residues of 3Cpro and 3CLpro of caliciviruses, picornaviruses and coronaviruses. Sequence alignment was made by Clustal omega. Amino acids for the triad (or dyad) were marked with asterisk (*) and other conserved amino acids in the active sites were marked with filled circle (•). For coronavirus 3CLpro, only domains 1 and 2 are shown in the figure.