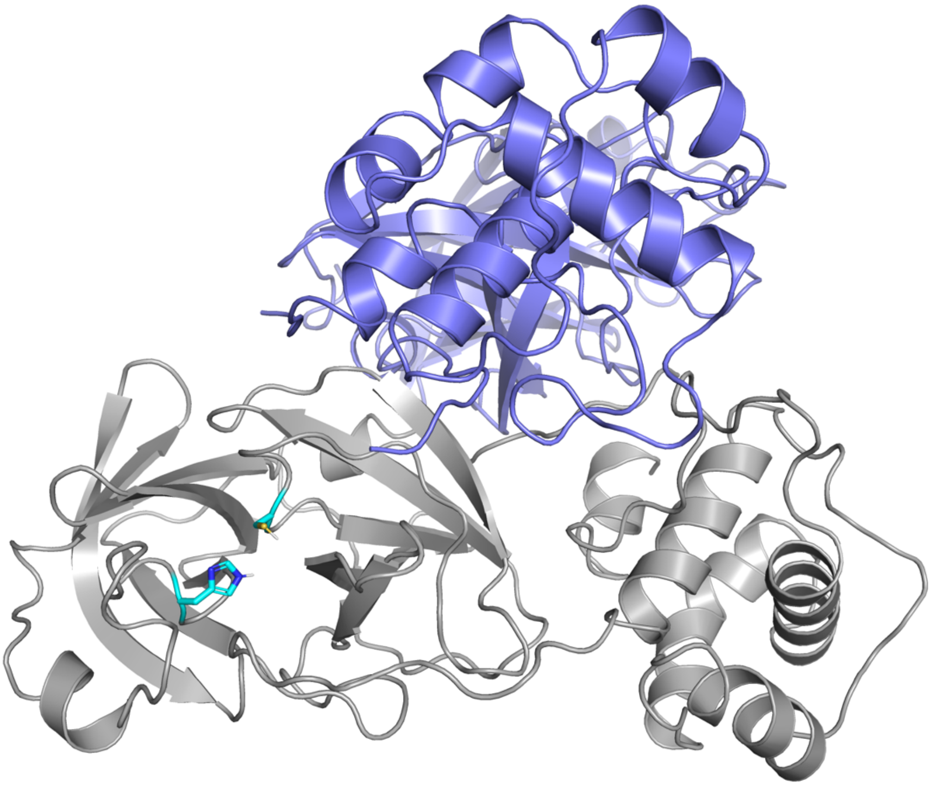
Homology models of Coronavirus 2019-nCoV 3CL^{pro} Protease

Martin J. Stoermer*
Division of Chemistry and Structural Biology, Institute for Molecular Bioscience, The University of Queensland, St. Lucia, 4072, Queensland, Australia.

*Corresponding author
Dr. Martin J. Stoermer
Division of Chemistry and Structural Biology, Institute for Molecular Bioscience
The University of Queensland
St. Lucia, 4072, Queensland, Australia
E-mail: martin.stoermer@uq.edu.au
Orcid: https://orcid.org/0000-0003-3445-2104
Twitter: @MartinStoermer

Abstract.



A regional outbreak of pneumonia in Wuhan, Hubei Province of China in late 2019 was associated with a novel coronavirus. Rapid release of genomic data for the isolated virus enabled the construction of first-generation homology models of the new CoV 3CL^{pro} cysteine protease. Whilst the overall viral genome was most closely associated with bat coronaviruses, the main protease is most closely related (96% identity) to SARS CoV protease.

Keywords: Wuhan coronavirus, 3CL^{pro}, protease, homology modelling

Foreword and Acknowledgements

The author is indebted to Professor Yong-Zhen Zhang, Fudan University, Shanghai, the Shanghai Public Health Clinical Center & School of Public Health, the Central Hospital of Wuhan, Huazhong University of Science and Technology, the Wuhan Center for Disease Control and Prevention, the National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control, and Professor Edward Holmes, The University of Sydney, Sydney, Australia for the rapid provision of the Wuhan coronavirus genome sequence WH-Human_1 and subsequent GenBank submissions to the international research community and the public.

Introduction.

In December 2019, the World Health Organisation (WHO) was officially informed of a cluster of cases of pneumonia of unknown cause detected in Wuhan City, Hubei Province of China.[1] Subsequently researchers from the Shanghai Public Health Clinical Center & School of Public Health reported that the evidence suggested an origin in a seafood market in Wuhan, and the isolation of a new type of coronavirus (novel coronavirus, nCoV) on 7 January 2020. Laboratory testing ruled out other respiratory pathogens including Severe Acute Respiratory Syndrome coronavirus (SARS-CoV).[2]

On 10th January 2020 a preliminary sequence (WH-Human_1.fasta.gz) of the novel coronavirus was made available on virological.org,[3] and subsequently on 12th January 2020, was released on Genbank (ID MN908947).[4] Preliminary analysis of the Wuhan virus suggests it is most closely related to coronaviruses found in bats.

Proteases are an important class of drug target,[6] most successfully in the cases of HIV protease[7] and Hepatitis C[8]. Recent human outbreaks of related coronaviruses SARS[9,10] in 2002 and MERS[11] in 2012 have focused the world's attention on the global risks of such emerging viruses. Coronaviruses are positive strand, enveloped RNA viruses with exceptionally large genomes which usually encode two or three viral proteases. In the case of Severe Acute Respiratory Syndrome coronavirus, the two proteases are a papain-like cysteine protease ($\mathrm{PL}^{\mathrm{pro}}$)[12] and a chymotrypsin-like cysteine protease $3\mathrm{CL}^{\mathrm{pro}}$,[13] also known as the Main protease. SARS-CoV $3\mathrm{CL}^{\mathrm{pro}}$ is essential for viral replication and is thus recognized as a potential drug target for the treatment of SARS infections. To date there are no clinically used inhibitors of the SARS protease yet they remain in development.[14]

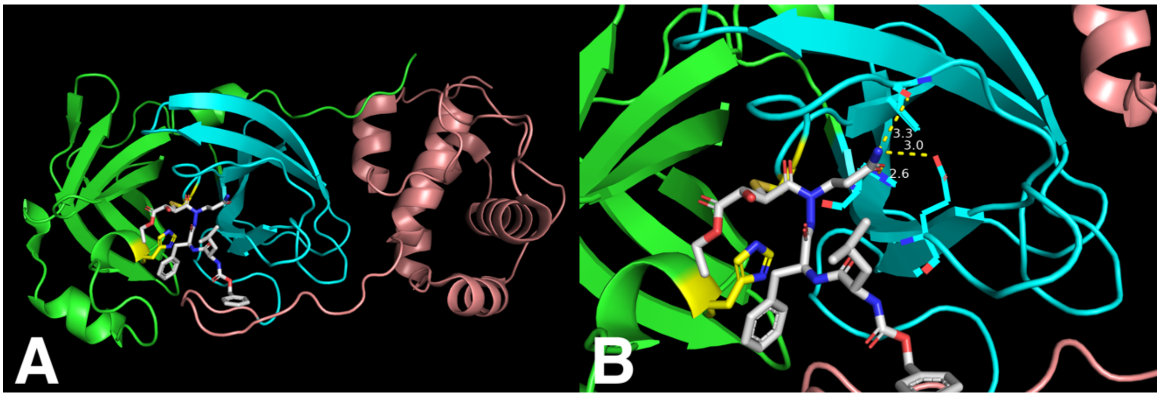


Figure 1. X-ray crystal structure of SARS-CoV 3CL^{pro} bound to a covalent inhibitor (PDB code 2a5i). Panel A: Domain I shown as yellow ribbons, domain II cyan ribbons, domain III pale pink ribbons. Catalytic dyad His41, Cys145 shown as yellow sticks, covalent azapeptide inhibitor shown as grey sticks. Panel B: Active site at the interface of domains I, II with P1 Gln residue shown hydrogen bonding to S1 subsite residues Phe140, His163, and Glu166 (blue lines).

The active site of SARS-CoV 3CL^{pro} is located in the cleft between domains I and II of the protein, and the two domains each contribute one residue to the catalytic dyad composed of His41 and Cys145 (Figure 1). Substrates and inhibitors typically contain Gln residues at the P1 position and large hydrophobic residues Phe/Leu/Met at P2.[15] The wealth of medicinal chemistry and biochemical information gained over the last 18 years, coupled with good-high sequence similarity between the members of the coronavirus family makes the homology modelling of new and emerging viruses an inviting prospect.

Results and Discussion.

The release of the viral genome of the novel Wuhan coronavirus prompted us to construct first-generation homology models of the new CoV 3CL^{pro}. Initial studies and BLAST analysis indicated that the virus was most closely related to a clade of bat coronaviruses (Figure 2), of which the closest neighbour for which there was protease structural information available in the Protein Data Bank (www.rcsb.org) was bat coronavirus HKU4.[16,17] EMBOSS Needle[18] alignment of the Wuhan genome with the sequence extracted from PDB:2YNB identified a putative 306ss viral 3CL^{pro} protein with 49% sequence identity (65% sequence similarity) (Figure 3).

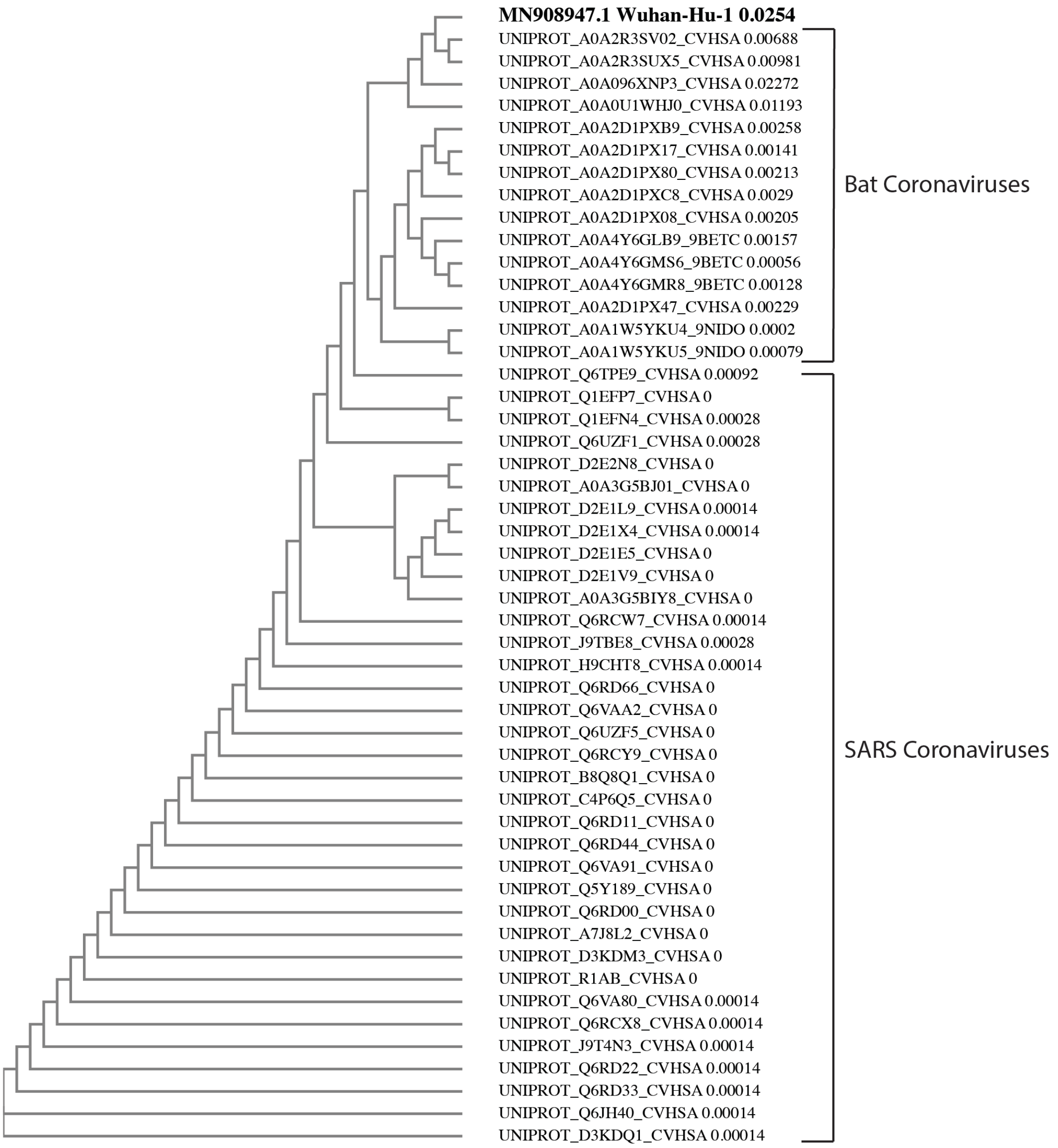

MN908947.1 Wuhan-Hu-1 0.0254
UNIPROT_A0A2R3SV02_CVHSA 0.00688
UNIPROT_A0A2R3SUX5_CVHSA 0.00981
UNIPROT_A0A096XNP3_CVHSA 0.02272
UNIPROT_A0A0U1WHJ0_CVHSA 0.01193
UNIPROT_A0A2D1PXB9_CVHSA 0.00258
UNIPROT_A0A2D1PX17_CVHSA 0.00141
UNIPROT_A0A2D1PX80_CVHSA 0.00213
UNIPROT_A0A2D1PXC8_CVHSA 0.0029
UNIPROT_A0A2D1PX08_CVHSA 0.00205
UNIPROT_A0A4Y6GLB9_9BETC 0.00157
UNIPROT_A0A4Y6GMS6_9BETC 0.00056
UNIPROT_A0A4Y6GMR8_9BETC 0.00128
UNIPROT_A0A2D1PX47_CVHSA 0.00229
UNIPROT_A0A1W5YKU4_9NIDO 0.0002
UNIPROT_A0A1W5YKU5_9NIDO 0.00079
UNIPROT_Q6TPE9_CVHSA 0.00092
UNIPROT_Q1EFP7_CVHSA 0
UNIPROT_Q1EFN4_CVHSA 0.00028
UNIPROT_Q6UZF1_CVHSA 0.00028
UNIPROT_D2E2N8_CVHSA 0
UNIPROT_A0A3G5BJ01_CVHSA 0
UNIPROT_D2E1L9_CVHSA 0.00014
UNIPROT_D2E1X4_CVHSA 0.00014
UNIPROT_D2E1E5_CVHSA 0
UNIPROT_D2E1V9_CVHSA 0
UNIPROT_A0A3G5BIY8_CVHSA 0
UNIPROT_Q6RCW7_CVHSA 0.00014
UNIPROT_J9TBE8_CVHSA 0.00028
UNIPROT_H9CHT8_CVHSA 0.00014
UNIPROT_Q6RD66_CVHSA 0
UNIPROT_Q6VAA2_CVHSA 0
UNIPROT_Q6UZF5_CVHSA 0
UNIPROT_Q6RCY9_CVHSA 0
UNIPROT_B8Q8Q1_CVHSA 0
UNIPROT_C4P6Q5_CVHSA 0
UNIPROT_Q6RD11_CVHSA 0
UNIPROT_Q6RD44_CVHSA 0
UNIPROT_Q6VA91_CVHSA 0
UNIPROT_Q5Y189_CVHSA 0
UNIPROT_Q6RD00_CVHSA 0
UNIPROT_A7J8L2_CVHSA 0
UNIPROT_D3KDM3_CVHSA 0
UNIPROT_R1AB_CVHSA 0
UNIPROT_Q6VA80_CVHSA 0.00014
UNIPROT_Q6RCX8_CVHSA 0.00014
UNIPROT_J9T4N3_CVHSA 0.00014
UNIPROT_Q6RD22_CVHSA 0.00014
UNIPROT_Q6RD33_CVHSA 0.00014
UNIPROT_Q6JH40_CVHSA 0.00014
UNIPROT_D3KDQ1_CVHSA 0.00014
Bat Coronaviruses
SARS Coronaviruses

Figure 2. Phylogenetic Neighbour-joining tree without distance corrections for Wuhan coronavirus and top BLAST hits generated with Clustal Omega.

```
BatHKU4_pr        SGLVKMSAPSGAVENCIVQVTCGSMTLNGLWLDNTVWCPRHIMCPADQLTDPNYDALLIS  60
WH-Human_1pr      SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIR  60
SARS_pr           SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIR  60
                  **: **: *** **.*:******: ********:.*:****::* :::: :***: ***

BatHKU4_pr        KTNHSFIVQKHIGAQANLRVVAHSMVGVLLKLTVDVANPSTPAYTFSTVKPGASFSVLAC  120
WH-Human_1pr      KSNHNFLVQA---GNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLAC  117
SARS_pr           KSNHSFLVQA---GNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTFSVLAC  117
                  *:**.*:**    .:.:***:.*** . :*:*.**.:**.** *.*  ::** :******

BatHKU4_pr        YNGKPTGVFTVNLRHNSTIKGSFLCGSCGSVGYTENGGVINFVYMHQMELSNGTHTGSSF  180
WH-Human_1pr      YNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDL  177
SARS_pr           YNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDL  177
                  ***.*:**:   :* * ******* *******:. : . :.* ***:*** .*.*:*:.:

BatHKU4_pr        DGVMYGAFEDKQTHQLQLTDKYCTINVVAWLYAAVLNGCKWFVKPTRVGIVTYNEWALSN  240
WH-Human_1pr      EGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKY  237
SARS_pr           EGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKY  237
                  :* :** * *:** *   **.  *:**:*******:** :**::   . :  :*  *:.

BatHKU4_pr        QFTEFVGT--QSIDMLAHRTGVSVEQMLAAI-QSLHAGFQGKTILGQSTLEDEFTPDDVN  297
WH-Human_1pr      NYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVV  297
SARS_pr           NYEPLTQDHVDILGPLSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTPFDVV  297
                  ::  :.    : :. *: :**::* :* *:: : *: *::*:****.: ******* **

BatHKU4_pr        MQVMGVVMQ      306
WH-Human_1pr      RQCSGVTFQ      306
SARS_pr           RQCSGVTFQ      306
                   *  **.:*
```

Figure 3. EMBOSS Needle alignment of the extracted protease sequence of Wuhan coronavirus against Bat coronavirus (PDB:2ynb) and SARS (PDB:2z9j). Catalytic dyads shown in red bold.

An initial Swissmodel[19] search for possible homology model templates against this search sequence however provided only SARS protease templates with much higher sequence identity and similarity (96, 99% respectively). This was confirmed via a second EMBOSS Needle alignment (Figure 3), and visually by a mapping of differing residues onto the respective crystal structures (Figure 4).

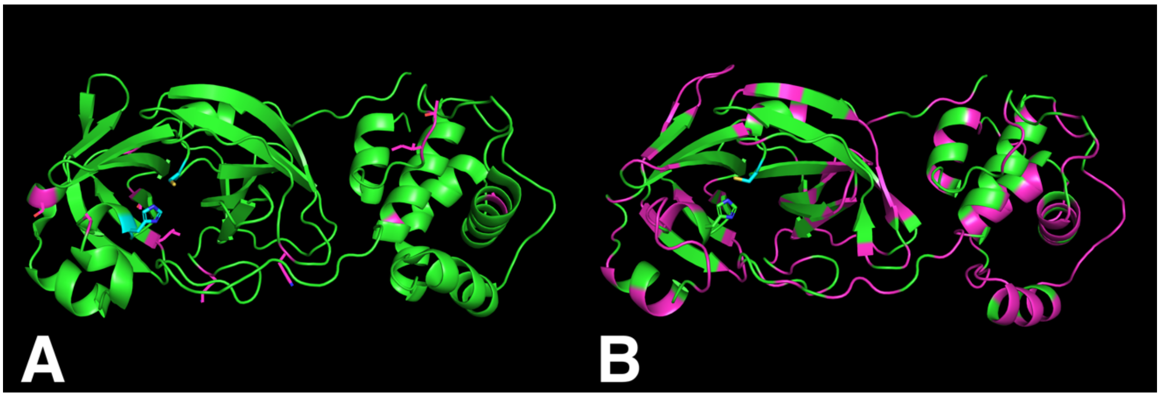


Figure 4. Visualisation of sequence differences between bat HKU4, SARS and Wuhan coronaviruses. Panel A: X-ray crystal structure (PDB:2z9j) of SARS CoV protease shown as green ribbons, with catalytic His41 and Cys145 residues shown as cyan sticks. Residues which differ in the Wuhan sequence are marked in magenta. Panel B: X-ray crystal structure (PDB:2ynb) of bat coronavirus HKU4 protease shown as green ribbons, with residues which differ in the Wuhan sequence are marked in magenta.

Alignment of the SARS 2z9j and HKU4 2ynb protease crystal structures however showed that the two had a very similar overall fold with an RMSD of 0.66Å across 229 well conserved Cα atoms. Accordingly, due to their higher overall sequence identities, SARS crystal structures were chosen for the future development of Wuhan CoV protease homology models. Coronavirus crystal structures are predominantly homodimers, and as a consequence Swissmodel template searches returned some duplicates arising from both chains A&B being selected. Models were built as homodimers where appropriate however for visualisation and initial refinement purposes the monomers were used (Table 1, Figure 5A). Pairs 1&3, 4&5 were functionally identical so only 6 models were taken through for further refinement. The 5 Swissmodel dimer models 1,2,4,6,7 were refined as both dimers and monomers, using Schrodinger Suite 2019 and monomer model 8 was refined without reconstruction to a dimer. The models were overlaid as their monomers and compared by manual inspection. The models were very structurally similar, especially in the active site region near His41 and Cys145 (Figure 5B).

Model	Template name	Title	% Identity	QSQE score	Method (Resolution)	Oligo State	Ligands
1	2z9j.1.B	3C-like proteinase	96.1	0.91	X-ray, 2.0Å	homodimer	2x DTZ
2	3vb3.1.A	3C-like proteinase	96.1	0.91	X-ray, 2.2Å	homodimer	None
3	2z9j.1.A	3C-like proteinase	96.1	0.91	X-ray, 2.0Å	homodimer	2x DTZ
4	1uk3.1.B	3C-like proteinase	96.1	0.88	X-ray, 2.4Å	homodimer	None
5	1uk3.1.A	3C-like proteinase	96.1	0.88	X-ray, 2.4Å	homodimer	None
6	2a5i.1.A	3C-like peptidase	96.1	0.86	X-ray, 1.9Å	homodimer	2x AZP
7	1uj1.1.B	3C-like proteinase	96.1	0.86	X-ray, 1.9Å	homodimer	None
8	1z1i.1.A	3C-like proteinase	96.1	-	X-ray, 2.8Å	monomer	None

Table 1. PDB templates used for Swissmodel homology model building. Templates ranked by internal Swissmodel QSQE score

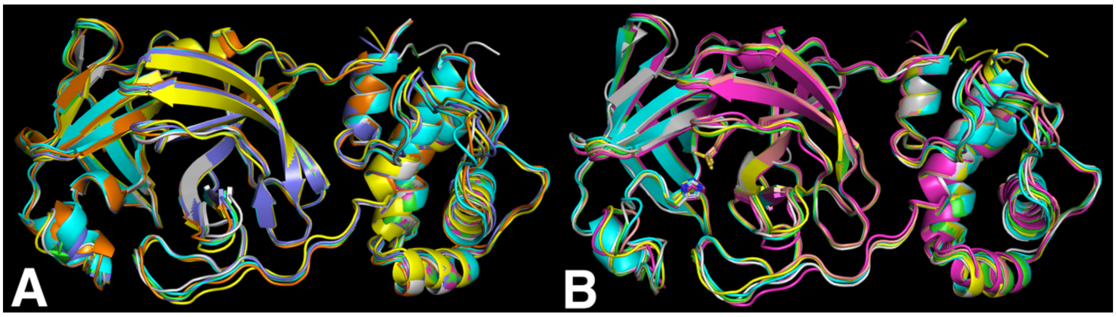


Figure 5. Panel A: Overlay of the 8 Wuhan coronavirus protease monomers produced via Swissmodel. Panel B: Overlay of the 6 energy minimised Wuhan coronavirus protease monomers produced in Schrödinger Suite. Model1 from 2z9j, green ribbons, model2 from 3vb3 cyan, model4 from 1uk3 magenta, model6 from 2a5i yellow, model7 from 1uj1 pale brown, and model8 from 1z1i grey ribbons.

By a combination of Swissmodel scoring of the initial models and MolProbity analysis of the refined overall model quality, the final two models selected were Model2 from PDB:3vb3 representing the unbound or apo form of the protease, and Model6 from PDB:2a5i which represents a ligand-bound form. The original SARS CoV 3CL^{pro} inhibitor was carried through the modelling protocol and is shown in the active site of the Wuhan 3CL^{pro} model (Figure 5). The apo-like and bound-like models 2,4 are very similar overall except for a significant difference in the loop ^{45}TSEDM^{49} which forms the boundary of the S2 subsite. This results in the bulky Met49 side chain moving out of the S2 pocket allowing the ligand Phe residue to be accommodated

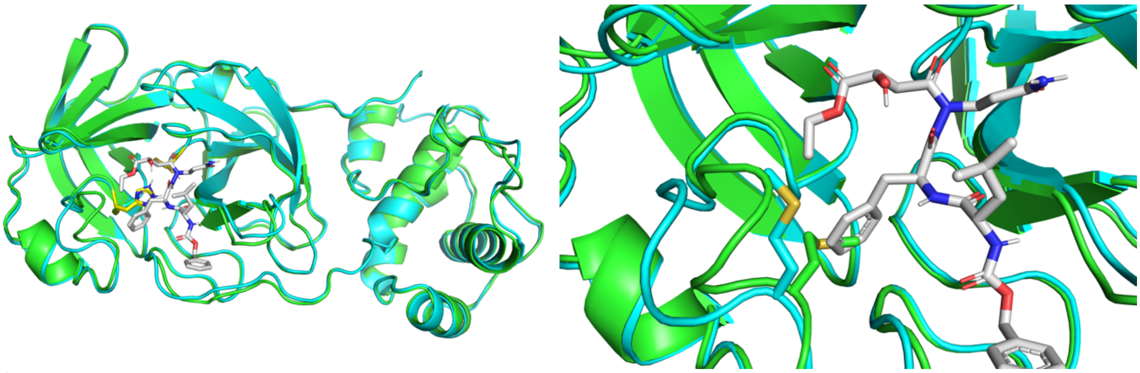


Figure 5. Overlay of ligand-bound and unbound forms of Wuhan coronavirus 3CL^{pro}. Panel A: Apo Model2 from PDB:3vb3 shown as green ribbons, bound Model6 from PDB:2a5i cyan ribbons. Catalytic triad yellow sticks, SARS CoV 3CL^{pro} inhibitor shown as grey sticks. Panel B: Active site of the overlaid bound and unbound forms showing Met49 and mobile loop in S2 pocket.

These models were also analysed using 100ns molecular dynamics simulations (see Supporting Information) to check the systems for overall structural stability (Figure 6). In general the dimeric forms were slightly more stable overall, as the *C*- and *N*-termini of the chains form part of the dimer interface and are held more closely than in the monomeric state where they are more mobile.

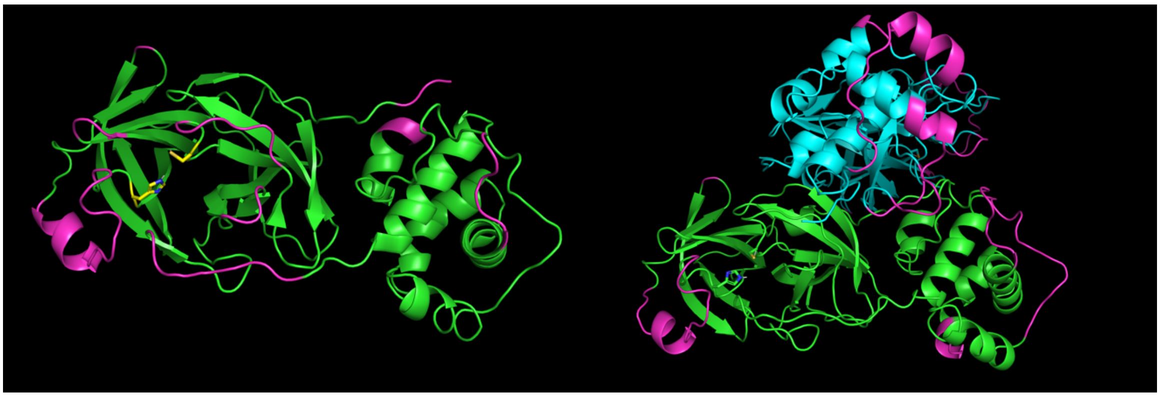


Figure 6. Comparison of mobile regions in molecular dynamics simulations. Panel A: Monomeric model2 as green ribbons except magenta ribbons for moderately mobile regions. Catalytic dyad yellow sticks. Panel B: Dimeric model12 shown as green ribbons Chain A, cyan ribbons Chain B, except magenta ribbons for moderately mobile regions. Catalytic dyad green and cyan sticks.

Finally an analysis of the new Wuhan coronavirus genome revealed multiple sites, analogous to those found or proposed for SARS[14] that may be cleaved by the 3CLpro after (L/F/M)-Gln residues. These are mostly identical to the SARS sequences (Table 2).

SARS 3CL^{pro} cleavage site	Site	Potential Wuhan coronavirus sites
AVLQ ↓ SGFR	TM2/3CLpro	AVLQ ↓ SGFR
VTFQ ↓ **GKF**K	3CLpro/TM3	VTFQ ↓ **SAV**K
ATVQ ↓ SKMS	TM3/?	ATVQ ↓ SKMS
ATLQ ↓ AIAS	?	ATLQ ↓ AIAS
VKLQ ↓ NNEL	?	VKLQ ↓ NNEL
VRLQ ↓ AGNA	?/GFL	VRLQ ↓ AGNA
P**LM**Q ↓ SADA	GFL/?	P**ML**Q ↓ SADA
TVLG ↓ AVGA	?/RdRp	TVLQ ↓ AVGA
ATLQ ↓ AENV	RdRp/NTPase	ATLQ ↓ AENV
TRLQ ↓ SLEN	NTPase, etc./exonuclease	TRLQ ↓ SLEN
PKLQ ↓ **A**SQA	exonuclease/2′-O-MT	PKLQ ↓ **S**SQA

Table 2. Proposed polyprotein cleavage sites of Wuhan coronovirus compared to SARS CoV 3CL^{pro} (Table adapted from Pillaiyar et al).[14] Bold residues represent differences; TM = Transmembrane; GFL = growth factor-like domain; RdRp = RNA- dependent RNA polymerase; 2′-O-MT, = 2′-O-methyltransferase

Methods.

The Wuhan coronavirus genome was obtained from Genbank/NCBI (release MN908947.1).[4] Sequences from crystal structure sequences were obtained from the Protein Data Bank (rcsb.org). Sequence analysis was carried out using the web interface at the European Bioinformatics Institute (EMBL-EBI).[18] Pairwise sequence alignments were performed using the EMBOSS-Needle method with default EBLOSUM62 matrices. Multiple sequence alignments were carried out using Clustal Omega using default HMM profile-profile parameters. Blast searches were carried out against the full UniProt Knowledgebase using blastp 2.9.0+ and default parameters.

Homology models were prepared using the publicly accessible online Swissmodel[19-23] programs, searching the SWISS-MODEL template library (SMTL version 2020-01-02, and PDB release 2019-12-27), or using a directed user-defined template approach. Coronavirus crystal structures are predominantly homodimers and were modelled as such, but the further refinement was initially performed on the monomeric forms. Newer releases of the Swissmodel software enable the inclusion of bound inhibitors when appropriate. In this case, models derived from PDB:2a5i included a bound covalent inhibitor. The initial heavy atom-only models were further refined using the Protein Preparation module in Schrödinger Suite 2019-2[24] to add hydrogen atoms and optimise the internal hydrogen bonding network. Finally, the models were energy minimised using the OPLS3e force field with charges from the force field and implicit water solvent, and the Polak-Ribier Conjugate Gradient (PRCG) method to gradient <0.05Å. The final minimised models were then analysed using MolProbity[25] and visualised in Pymol v2.1.[26] Molecular dynamics simulations were performed with the Desmond Molecular Dynamics System (D. E. Shaw Research, New York, NY) by using the tools incorporated in the Schrödinger Suite 2019-2.[24]

Conclusion.

The main protease encoded by the Wuhan fish market coronavirus is very highly homologous to SARS CoV 3CL^{pro}, whereas the virus overall is more highly related to bat coronaviruses. Homology models of the Wuhan coronavirus protease were built from known SARS 3CL^{pro} crystal structures in both ligand-bound and apo forms. These represent potential starting points for structure-based drug design and for suggesting point mutations to inform future biochemical experiments.

Supporting Information

EMBOSS and Clustal Omega sequence alignments, MolProbity statistics and Ramachandran plots for refined monomer and dimer models. PDB files for raw Swissmodel and individual refined monomer and dimer models. Molecular Dynamics simulations of monomer and dimeric models.

References.

1. https://www.who.int/westernpacific/emergencies/pneumonia-in-wuhan-china. Accessed 12th January 2020.

2. https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/. Accessed 12th January 2020.

3. Holmes EC on behalf of Zhang Y-Z. http://virological.org/t/initial-genome-release-of-novel-coronavirus/319. Accessed 12th January 2020.

4. Zhang Y-Z, Wu F, Chen Y-M, Pei Y-Y, Xu L, Wang W, Zhao S, Yu B, Hu Y, Tao Z-W, Song Z-G, Tian J-H, Zhang Y-L, Liu Y, Zheng J-J, Dai F-H, Wang Q-M, She J-L, and Zhu T-Y Nucleotide Accession No. MN908947.1, Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]. Accessed 13th January 2020. Available from: https://www.ncbi.nlm.nih.gov/nuccore/MN908947

5. https://www.sciencemag.org/news/2020/01/mystery-virus-found-wuhan-resembles-bat-viruses-not-sars-chinese-scientist-says. Accessed 12th January 2020.

6. Agbowuro AA, Huston WM, Gamble AB, Tyndall JDA. Proteases and protease inhibitors in infectious diseases. *Med Res Rev*. 2018;38(4): 1295-1331. (DOI:10.1002/med.21475)

7. Piot P, Abdool Karim SS, Hecht R, et al. Defeating AIDS--advancing global health. *Lancet*. 2015;386(9989): 171-218. (DOI:10.1016/s0140-6736(15)60658-4)

8. Howe AY, Venkatraman S. The Discovery and Development of Boceprevir: A Novel, First-generation Inhibitor of the Hepatitis C Virus NS3/4A Serine Protease. *J Clin Transl Hepatol*. 2013;1(1): 22-32. (DOI:10.14218/jcth.2013.002xx)

9. Drosten C, Gunther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med*. 2003;348(20): 1967-1976. (DOI:10.1056/NEJMoa030747)

10. Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*. 2003;348(20): 1953-1966. (DOI:10.1056/NEJMoa030781)

11. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367(19): 1814-1820. (DOI:10.1056/NEJMoa1211721)

12. Ratia K, Saikatendu KS, Santarsiero BD, et al. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc Natl Acad Sci U S A*. 2006;103(15): 5717-5722. (DOI:10.1073/pnas.0510851103)

13. Chen S, Chen L, Tan J, et al. 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(1): 164-173. (DOI:10.1074/jbc.M408211200)

14. Pillaiyar T, Manickam M, Namasivayam V, Hayashi Y, Jung S-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *Journal of Medicinal Chemistry*. 2016;59(14): 6595-6628. (DOI:10.1021/acs.jmedchem.5b01461)

15. Fan K, Wei P, Feng Q, et al. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem*. 2004;279(3): 1637-1642. (DOI:10.1074/jbc.M310875200)

16. 6 Bat coronavirus HKU4 crystal structures are available in the PDB; Accession codes 2YNA, 2YNB, 4YO9, 4YOG, 4YOI, 4YOJ.

17. St John SE, Tomar S, Stauffer SR, Mesecar AD. 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(17): 6036-6048. (DOI:10.1016/j.bmc.2015.06.039)

18. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*. 2019;47(W1): W636-W641. (DOI:10.1093/nar/gkz268)

19. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22(2): 195-201. (DOI:10.1093/bioinformatics/bti770)

20. Kopp J, Schwede T. The SWISS-MODEL Repository: new features and functionalities. *Nucleic Acids Research*. 2006;34(suppl_1): D315-D318. (DOI:10.1093/nar/gkj056)

21. Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42(Web Server issue): W252-258. (DOI:10.1093/nar/gku340)

22. Bienert S, Waterhouse A, de Beer Tjaart AP, et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research*. 2017;45(D1): D313-D319. (DOI:10.1093/nar/gkw1132)

23. Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*. 2018;46(W1): W296-W303. (DOI:10.1093/nar/gky427)

24. Schrodinger L. Small-Molecule Drug Discovery Suite 2019-3. 2017-2 ed. New York, NY: Schrodinger, LLC; 2019.

25. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*. 2007;35(Web Server issue): W375-383. (DOI:10.1093/nar/gkm216)

26. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.