

Supporting information

3-Chymotrypsin-like Protease (3CLpro): Validation as a Molecular Target, Proposal of a Novel Catalytic Mechanism, and Inhibitors in Preclinical and Clinical Trials

Vitor Martins de Freitas Amorim¹, Eduardo Pereira Soares¹, Anielle Salviano de Almeida Ferrari¹, Davi Gabriel Salustiano Merighi¹, Robson Francisco de Souza¹, Cristiane Rodrigues Guzzo^{1,*}, Anacleto Silva de Souza^{1,*}

¹Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil.

* Corresponding author

E-mail: crisguzzo@usp.br and crisguzzo@gmail.com

E-mail: anacetosilvadesouza@usp.br and anacetosilvadesouza@gmail.com

Running title: main protease and drugs in clinical trials

To whom correspondence should be addressed:

Cristiane R. Guzzo, Ph.D, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes, 1374, Cidade Universitária, 5508-000, São Paulo/SP, Brazil, +55 11 3091-7298; E-mail: crisguzzo@usp.br

Anacleto Silva de Souza, Ph.D, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes, 1374, Cidade Universitária, 5508-000, São Paulo/SP, Brazil, +55 11 3091-7298; E-mail: anacetosilvadesouza@usp.br

Keywords: SARS-CoV-2, 3CLpro, Novel Mechanism of Catalysis, FDA-approved drugs, Non-Covalent and Covalent Inhibitors.

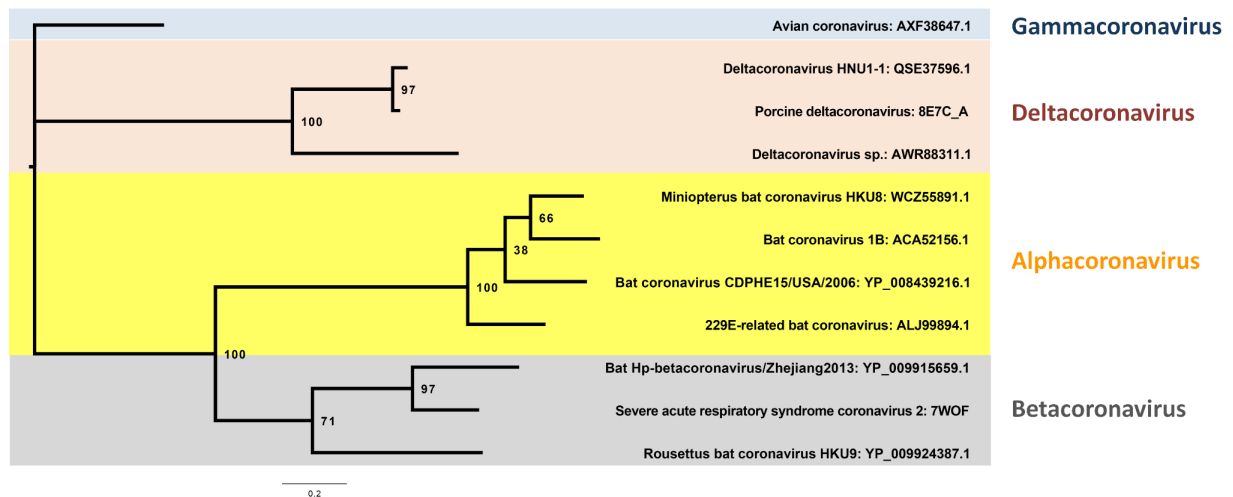


Figure S1. Phylogenetic tree illustrating the evolution of 3CLpro in different *Othocoronavirinae* genus. Through bioinformatic analysis of representative members (ACA52156.1, ALJ99894.1, AWR88311.1, AXF38647.1, QSE37596.1, WCZ55891.1, YP_008439216.1, YP_009915659.1, YP_009924387.1, 7WOF_A, 8E7C_A), we observed their classification into four genus *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*.

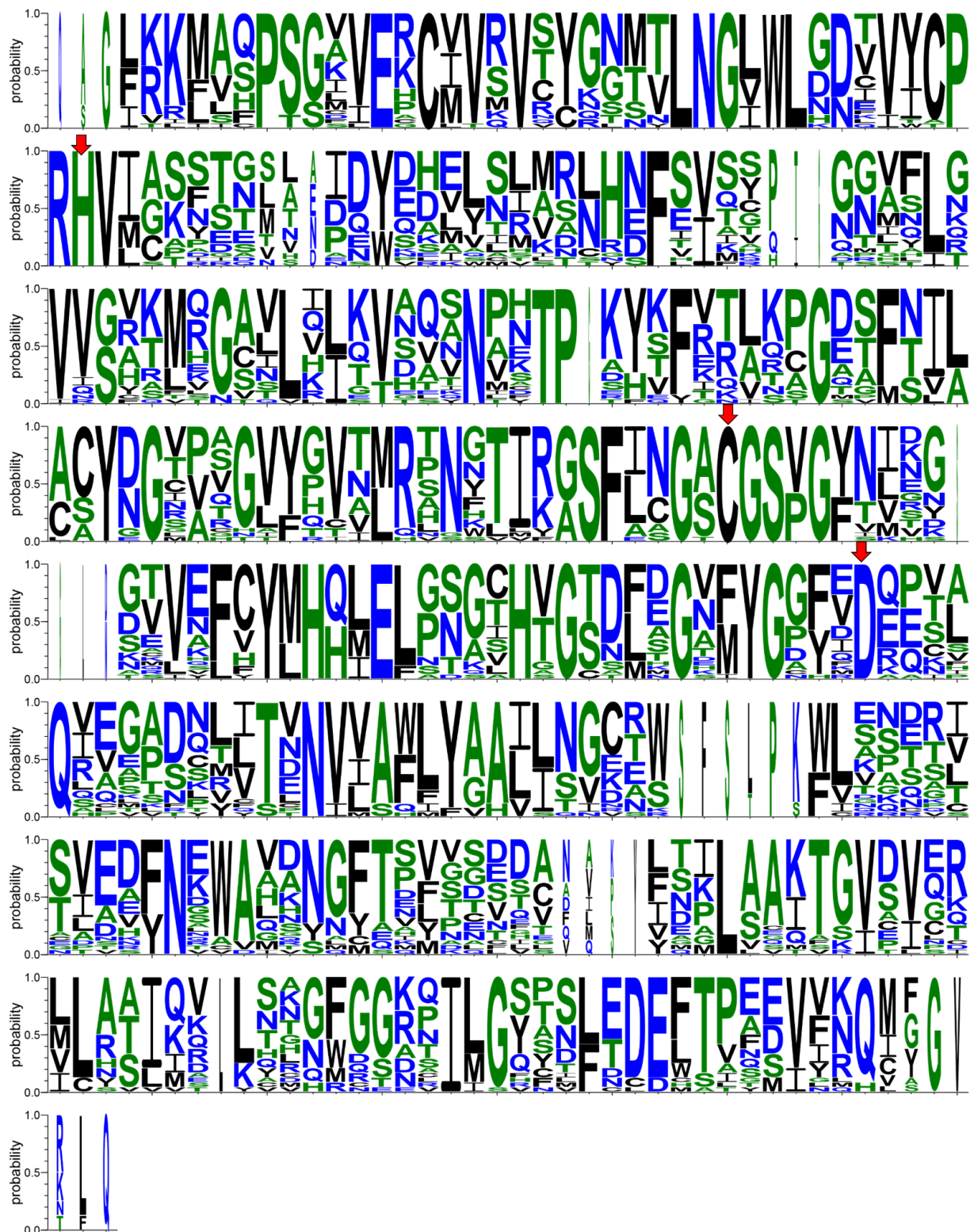


Figure S2. Representation of the multiple sequence alignment. Through alignment with the sequences ACA52156.1, ALJ99894.1, AWR88311.1, AXF38647.1, QSE37596.1, WCZ55891.1, YP_008439216.1, YP_009915659.1, YP_009924387.1, 7WOF_A, 8E7C_A, we observed a high conservation of amino acid residues, including H41, C145, and D187. The probability of these residues being important for the enzyme in the catalytic process based on phylogeny is 100%, corroborating with the experimental data obtained by Adem and co-authors (Al Adem et al., 2023).

Table S1. Representatives of the groups for multiple sequence alignment. ID, sequence, and Organism represent the accession code, sequence in fasta format, and coronavirus strains, respectively.

id	sequence	Organism
ACA52156.1	>ACA52156.1:3206-3501 ORF1ab polyprotein [Bat coronavirus 1B] LRKFAQPSGVIEHCIVRVSYGNMVLNGIWLGDDEVICPRHVIASSTNT TIDYDHEYTMMLRHNFSSVSSGNL FIGVVSAKMRGASLVIKVNQNNPHTPKHVFKTLKAGDAFNILACYDG VPSGVYGTILRHNKTIIRGSFING ACGSPGFNNGDTVEFVYLHQLELGSGCHVGSNMEGAMYGGFEDQPS LQIEGADCLVTNVNIAFLYGAIL NGCTWFLSNERITAEVFNGWAHANNFTEVGSGLDCFNILAAKTGVDIQ RVLASIQKLSKGFGRNIIGYAS LTDEFTVTEVVKQMYG	Bat coronavirus 1B
ALJ99894.1	>ALJ99894.1:2973-3268 orf1a [229E-related bat coronavirus] LRKMAQPSGIVERCIVRVCYGNTVLNGLWLGDVVYCPRHVIASNTTM AIDYEHEYSVMRLHNFSITSGSV FLGVVGATMHGATLRIKVSQTNMCTPKHSFRTLKSGEGFNILACYDG CAHGVFGVMNRTNWTIRGSFING ACGSPGYNLRGSEVEFCYMHQIELGSGSHVGSSFDGVMYGGFDDQPN LQVESASQMLTTNVVAFLYAAIL NGCTWWLKGDRLSVEHYNEWAQSNNGFTAMNGEDAFSILAAKTGVCVE RLLHAIQVLSNGFGGRQILGYSS LNDEFSINEVVRQMFG	229E-related bat coronavirus
AWR88311.1	>AWR88311.1:2447-2748 replicase polyprotein, partial [Deltacoronavirus sp.] QAGIKKLLSPSGMVEHCVVSVTYQGSTLNGVWLHDTVYTPRHVIGKY TANQWKDMVNLVDNREFSVFCPT QGVNLVVSVRMQGAVLQLKVHAKNLKTPDYVFERATPGTPMTIACA YDGVVQTIYHVVLQNNGLLFAF LNGACGSGVGYTVRGTKLVLYYMHHIEFSNKTHGGTTMDGHFYGDYVD EEKIQHLKSSSTLTDNVMISHIV HLLTVNRKPSWLASNGIEIADFNEWAATSNTFKYPLNVENQKYLEAL SGIVKVPINIILATIVKLHANWG EARILQSQSDFELDWTPQMVYNQ	Deltacoronavirus sp.
AXF38647.1	>AXF38647.1:600-905 MAG: non-structural polyprotein, partial [Avian coronavirus] GFKKFVSPSSAVEKCVVMVRYKGTVLNGLWLNDVCYCPRHVMGKFSG DEWQDVLSLASNHEFEVTGPNGA SLQVVGRKLQGAVLILQTAAVNANTPKYKFVQAQCGDTFTIACSUNG NVVGLYPVTMRANGTIRASFACG SCGSGVGFNMDDGVVAFCYMHHLPLNAIHTGTDLAGNFYGGFIDEEC AQRVEPDLRITNNILAWLYAAII SVRESSFSTPKWLESTTVSIEDYNKWAVDNGFTSFVSCTAITKLSAI TGVDVCKLLRTIMVKSQWQSEP ILGQYNFEDEMTPEVFNHVGGVRLQ	Avian coronavirus

QSE37596.1	>QSE37596.1:2588-2889 ORF1ab polyprotein [Deltacoronavirus HNU1-1] QAGIKLLHPSGVVEKCMVSVVYNGSTLNGIWLKNVVYCPRHVIGKY RGEQWSHMVSIADCRDFIVKCPT QGIQLNVQSVKMVGALLQLTVHTSNTATPDYKFERLTPGSSMTIACA YDGVVRHVYHVVLQLNNLIYASF LNGACGSVGYTLKGKTLYLHYMHIEFNKTHSGTDLEGNFYGPYVD EEVAQQQTAFQYYTDNVVAQLYA HLLTVDARPKWLAQSQISIEDFNAWAANNLFANFPCEQANMSYIMGL SQIARIPIERVLNTIIQLTLNRD GAVIMGSPDFECDWTPEMVYNQ	Deltacoronavirus HNU1-1
WCZ55891.1	>WCZ55891.1:3106-3401 MAG: ORF1a protein [Miniopterus bat coronavirus HKU8] LKKMAQPSGIIERCVRVCYGNMTLNGIWLGDVVICPRHVIASSTNS TINYEHEYTMRLHNFSSVSGNV FLGINGVTMKGVLNHIKVNQSNPHTPAHSFRTLKPGDSFNILACYDG TPAGVYGITLRPNYTIRGSFING ACGSVGYNVVNGQVEFCYLHQLELGSGCHVGSNFDGVMYGAQDQPS LQIEGADQLVTPNVVAFLYGALL NGCNWFLVNERLTVEAYNEWAHNGFTDMVGVECFMMLAAKTGVDVQ RLLAAIQKLTKGFGGKNILGFTS LTDEFTAAEVIKQMYG	Miniopterus bat coronavirus HKU8
YP_008439216.1	>YP_008439216.1 3C-like proteinase nsp5 [Bat coronavirus CDPHE15/USA/2006] AGLKKMAQPSGVVERCIVRVSYGNMVLNGLWLGDTVYCPRHVLAAS TTHIDYEHALAIMRLHNFSSISYG NTFLGVVGCSTKGALLLIKVGQNNVHTPGSYTFRTLKPGDSFNILAC YDGTATGVYGVNLRTNHTIRGSF INGACGSPGFNIHGSTVEFCYLHQLELGSGCHVGSMDNGAMYGGYED QPSFQIEGVSNLVSENVVAFLYG ALLNGCNWWLDQCGVTVEAYNDWAHSNGCTSLSSDCTILAAKTGV EVSRLAAIQRLNVNFVGGKAILG YTSLTDEFTVGEIIKQMFVGNLQ	Bat coronavirus CDPHE15/USA/2006
YP_009915659.1	>YP_009915659.1:3408-3704 ORF1a polyprotein [Bat Hp-betacoronavirus/Zhejiang2013] RKMSCPTGKVERCMVRVTCGTMTLNLGLWLDNTVYCPRHVMCTPEELL APDYDSILLRKATHSFTVQYGT YLKVVSYKMTGSVLQLGVDQINPETPKYKFVRKPGATFSVLACYNG MPAGVYQVAMRPNHTIKGSFLNG SCGSVGYTLGYDRVEFCYMHMELPTGVHTGTDLGTFYGFVDRQT SQSAGSDNTLTNLVLAWLAAVI NGERWFIVPQTCALTDNFNTAVLKYGYQSLTEDGVAALDPLVAQTGIS VQTMCASLKDLLVHGMRGRCILS SPTLEDEFTPFDIVRQC	Bat Hp-betacoronavirus/Zhejiang g2013
YP_009924387.1	>YP_009924387.1:4-300 nsp5 [Rousettus bat coronavirus HKU9] TRMAHPSGLVEPCLVKVNYGSMTLNGIWLDFVICPRHVMCSRDELA NPDPRLSMRAANYDFHVSQNGH NIRVIGHTMEGSLKLTVDVNNPKTPAYSFIRVSTGQAMSLACDYDG LPTGVYTCTLRNGTMRASFLCG SCGSPGFVMNGKEVQFCYLHQLELPNGTHTGTDFSGVFYGPFDKQV PQLAAPDCTITVNLVLAWLAAVL SGENWFLTKSSISPAEFNNCAVKYMCQSVTSESLLQVLQPLAAKTGIS	Rousettus bat coronavirus HKU9

	VERMLSALKVLLSAGFCGRTIMG SCSLEDEHTPYDIGRQM	
7WOF_A	>pdb 7WOF A Chain A, 3C-like proteinase SGFRKMAFSPGKVEGCMVQVTCGTTTLNGLWLDVVCPRHVICTSE DMLNPNYEDLLIRKSNHNFLVQA GNVQLRVIGHSMQNCVLKLVDTANPKTPKYKFVRIQPGQTFSVLAC YNGSPSGVYQCAMRPNFTIKGSF LNGSCGSVGFNIDYDCVSFCYMHMELPTGVHAGTDLEGNFYGPFFVD RQTAQAAGTDTTITVNVLAWLYA AVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQT GIAVLDMCASLKELLQNGMNGRT ILGSALLEDEFTPFDDVVRQCSGVTFQ	Severe acute respiratory syndrome coronavirus 2
8E7C_A	>pdb 8E7C A Chain A, Main Protease AGIKILLHPSGVVERCMVSVVYNGSALNGIWLKNVVYCPRHVIGKFR GDQWTHMVSIAADCRDFIVKCPIQ GIQLNVQSVKMGALLQLTVHTNNTATPDYKFERLQPGSSMTIACAY DGIVRHVYHVVLQNNLIYASFL NGACGSVGYTLKGKTLYLHYMHMIEFNKTHSGTDLEGNFYGPYVDE EVIQQQTAFQYYTDNVVAQLYAH LLTVDARPKWLAQSQISIEDFNWAANNSFANFPCEQTNMSYIMGLS QTARVPVERILNTIIQLTTNRDG ACIMGSYDFECDWTPEMVYNQAPISLQ	Porcine deltacoronavirus

Table S2. The results of the search for structural homologues of 3CLpro in human proteins, obtained from the Dali server. PDB ID, which identifies protein structures, the Z-score, indicating structural similarity, the RMSD to measure discrepancies between structures, the ID (%) to evaluate the similarity of amino acid sequences and the name of the protein under analysis.

PDB ID	Z-score	RMSD	ID (%)	Protein
8s9k	11.5	3.3	16	SERINE PROTEASE FAM111A
5to3	10.5	2.9	10	PROTHROMBIN
1z8g	10.4	2.9	16	SERINE PROTEASE HEPsin
4kkd	10.1	3.4	11	MANNAN-BINDING LECTIN SERINE PROTEASE
2xwb	10.1	3.2	8	COMPLEMENT C3B BETA CHAIN
7wqx	9.7	3.5	9	ENTEROPEPTIDASE
3h5c	9.6	2.7	12	PROTEIN Z-DEPENDENT PROTEASE INHIBITOR
2xrc	9.6	3.1	14	HUMAN COMPLEMENT FACTOR I
5i25	9.4	2.8	9	COAGULATION FACTOR XI
4a5t	9.1	3.1	8	PLASMINOGEN