Classification and structural analysis of coronavirus papain-like proteases based on the active site residues of SARS-CoV-PLpro

Rima Chaudhuri[1, 2], Guijun Zhao[2], Hui Lu[2] and Michael E. Johnson[1]
[1]Center for Pharmaceutical Biotechnology [2]Department of Bioengineering, Bioinformatics Program, University of Illinois at Chicago, Chicago, IL -60607

Abstract:

The two Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) proteases, Papain-like protease (PLpro) and 3Chymotrypsin-like protease (3CLpro) play primary roles in processing the viral polyproteins pp1a and pp1ab, and are fundamental to the replication of the genomic RNA of SARS-CoV. The replicase gene of coronaviruses often encodes two papain-like proteases, PL1 and PL2, whereas SARS-CoV encodes only one papain-like protease domain.

In this paper, we propose 1) a new multiple sequence alignment (MSA) of the catalytic domain of 20 coronaviral papain-like proteases, based on our MSA, we propose that the position analogous to Y274 (Y1813) in SARS-CoV-PLpro decides the open and closed nature of the active sites, 2) a majority rule consensus phylogenetic tree built based on the new set of 14 residues identified from the crystal structure and the improved MSA which classifies the active sites of 20 papain-like proteases showing the minimum number of amino acid substitutions needed for the enzyme to acquire or lose functionality, 3) structural models of a few coronaviral PLpros based on the recent crystal structure of SARS-CoV-PLpro, and, 4) the identification of the closest analog to SARS-CoV-PLpro, using structural alignment of these models that provides useful insights for the physico-chemical properties of their active sites that may be helpful for subsequent drug design efforts.

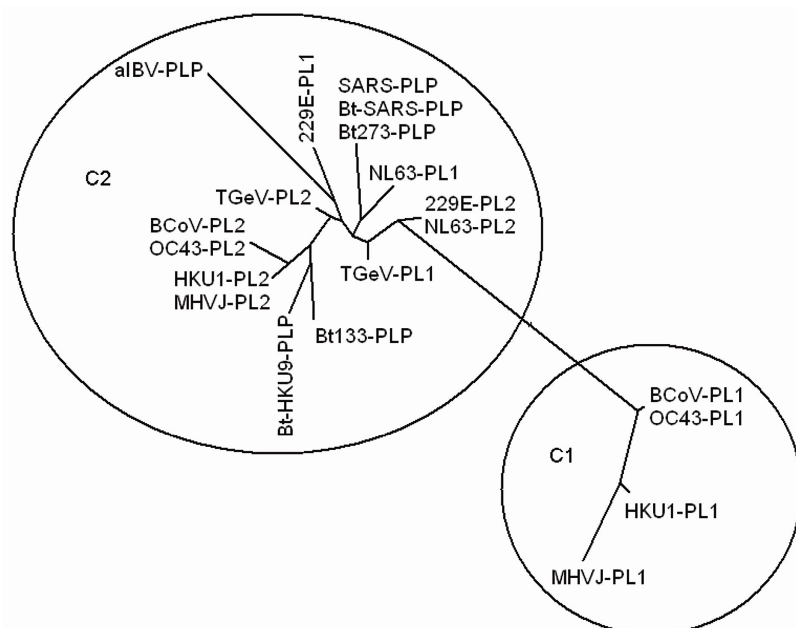
Our MSA, the crystal structure and analysis of the homology based models suggested inclusion of an important residue in our analysis that was not part of prior analyses: the residue analogous to E168 (E1707) in SARS-CoV-PLpro (Figure A) Based on our hypothesis when mutated in NL63-PL2, the enzyme lost its deubiquitinating activity (Chen and Baker, unpublished). Phylogenetic studies suggest that the active site residues in HCoV-NL63-PLP1 and TGEV-PLP1 closest to those of SARS-CoV-PLpro (Figure B). The recently solved X-ray structure of SARS-CoV-PLpro now provides a solid foundation for construction and refinement of reliable structural models for other coronaviral papain-like proteases. Our structural alignment identifies important physico-chemical characteristics of the predicted binding pockets of the other 19 proteases. Our results suggest that the active site of HCoV-NL63-PL2 is physico-chemically similar to that of SARS-CoV-PLpro. We have focused particularly on the structural analysis of the homology based models of NL63-PLP1 and NL63-PLP2 which suggests reasons for the apparent lack of DUB activity by NL63-PLP1. The P5 substrate position of HCoV-NL63-PLP1 is electropositive, in contrast to that of other enzymes, due to the presence of K163 at a position analogous to the position of E168 in SARS-CoV-PLpro. The entrance to the active site is bracketed by the two lysines which also form a strong positive potential that would likely hinder the positioning of a LRLRGG substrate in the HCoV-NL63-PLP1 active site. The structural models also highlight the differences in substrate cleavage site specificity between NL63-PLP2 and SARS-CoV-PLpro, Figure C summarizes the substrate cleavage sites for the enzymes SARS-CoV-PLpro and NL63-PLP2. Identification of the papain-like proteases of HCoV-NL63 as the closest kin to SARS-CoV-PLpro and the fact that HCoV-NL63 also shares the same cellular receptor as SARS-CoV is potentially important for therapeutic development.

Created with novaPDF Printer (www.novaPDF.com). Please register to remove this message.

Figure A: The 14 active site residues of SARS-CoV-PLpro and their corresponding aligned residues in the active sites of other coronaviral PLpros as extracted from the MSA. The catalytic residues are highlighted in yellow. The residues highlighted in green are a few of the important residues that differ significantly among SARS-PLpro, NL63-Plp2 and NL63-Plp1. Papain-like protease 1/2 (PLP, PL1, PL2), Bat-SARS-CoV-PLP (BtSARSPLP), New Haven CoV(NL63), Bovine-CoV (BCoV), Murine hepatitis virus(MHV), Human coronavirus OC43, 229E, HKU1 (OC43, 229E, HKU1), Transmissible gastroenteritis virus(Tgev), avian infectious bronchitis virus (aIBV)

Organism-PLP														
SARS-PLP	W107	N110	C110	Y113	L163	G164	D165	E168	Y265	G272	G267	H273	Y274	D287
BtSARS-PLP	W	N	C	Y	L	G	D	E	Y	G	G	H	Y	D
Bt273- PLP	W	N	C	Y	L	G	D	E	Y	G	G	H	Y	D
Bt133 -PLP	L	N	C	Y	P	D	D	E	F	G	G	H	Y	D
BtHKU9-PLP	Q	N	C	Y	S	D	D	M	F	G	G	H	Y	D
NL63-PL2	T	N	C	W	K	G	D	E	F	G	G	H	Y	D
NL63-PL1	Q	N	C	W	L	G	D	K	Y	G	G	H	Y	D
BCoV-PL2	Q	N	C	F	P	A	D	D	F	G	G	H	Y	D
MHV-PL2	Q	N	C	Y	P	S	D	D	F	G	G	H	Y	D
OC43-PL2	Q	N	C	F	P	A	D	D	F	G	G	H	Y	D
Tgev-PL2	Q	N	C	W	P	G	D	E	Y	G	G	H	Y	D
229E-PL2	T	N	C	W	K	G	D	E	F	G	G	H	Y	D
HKU1-PL2	Q	N	C	Y	P	S	D	D	F	G	G	H	Y	D
BCoV-PL1	A	N	C	W	G	G	Y	D	Y	A	S	H	S	D
Tgev-PL1	Q	N	C	W	S	G	D	E	Y	G	G	H	Y	D
229E-PL1	Q	N	C	W	M	G	D	L	F	G	G	H	Y	D
OC43-PL1	A	N	C	W	G	G	Y	D	Y	A	S	H	S	D
IBV PLP	W	N	C	W	F	S	D	W	F	G	G	H	C	D
MHV-PL1	I	T	C	W	G	G	Y	D	Y	A	C	H	S	D
HKU1-PL1	I	N	C	W	G	G	F	D	F	A	C	H	S	D

Figure B: Majority rule based consensus tree shows classification of the coronaviral PLPs based on the fourteen active site residue alignment observed in Figure A.



Created with novaPDF Printer (www.novaPDF.com). Please register to remove this message.

Figure C: Substrate specificity and substrate cleavage sites for the enzymes SARS-CoV-PLpro and NL63-Plp2 is summarized here. The SARS-CoV-PLpro substrates have a consensus sequence of LXGG which is also recognized by NL63-Plp2, but the difference lies in the position of the cleavage site. The table shows the sequence of the substrates for SARS-CoV-PLpro and NL63-PLp2. Substrate1, 2 denote the fragment of pp1a (replicase polyprotein 1a) processed by the SARS-CoV-PLpro to release nsp1 and nsp2 respectively. The subscripts indicate the amino acid ID in the sequence of pp1a. Synthetic peptide1, 2 were used by Chen et al to identify the precise cleavage sites by NL63-Plp2. The cleavage site is between P1 and P1' marked by the '/' symbol. The substrates are aligned from the P6 to P2' pockets of the enzymes. The substrate specificity between the two enzymes can be partially explained by the electrostatic potential (EP) of their binding pockets as described in table 3. The difference in potential of the P3 pocket is clear from this table, contributed mainly due to the mutation of L163 in SARS-CoV-PLpro to a Lysine in NL63-PLp2 which requires inducing a one residue shift to accommodate substrates with a cationic residue at S3 substrate position.

Substrates sites	S6	S5	S4	S3	S2	S1	S1'	S2'
SARS-PLPRO: SUBSTRATE1 sequence (pp1a)NSP1	$_{175}$R	E	L	N	G	G	/A	V_{182}
SUBSTRATE2 sequence (pp1a) NSP2	$_{813}$F	R	L	K	G	G	/A	P_{820}
NL63-PLP2: Synthetic peptide1 NSP2-NSP3	F	T	K	L	A	G	/G	K
Synthetic peptide2 NSP3-NSP4	V	A	K	Q	G	A	/G	F
Enzyme Binding Pockets	P6	P5	P4	P3	P2	P1	P1'	P2'
EP of SARS-CoV-PLpro binding pocket		(-)	(-)	(-)	weakly(+)/ neutral	weakly(+)/ neutral		
EP of NL63-PLp2 binding pocket		(-)	(-)	(+)	weakly(+) /neutral	weakly(+) /neutral		

Created with novaPDF Printer (www.novaPDF.com). Please register to remove this message.