

REVIEW ARTICLE

Potential 3-chymotrypsin-like cysteine protease cleavage sites in the coronavirus polyproteins pp1a and pp1ab and their possible relevance to COVID-19 vaccine and drug development

Shaomin Yan  | Guang Wu 

National Engineering Research Center for Non-Food Biorefinery, State Key Laboratory of Non-Food Biomass and Enzyme Technology, Guangxi Biomass Engineering Technology Research Center, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, Nanning, China

Correspondence

Guang Wu, National Engineering Research Center for Non-Food Biorefinery, State Key Laboratory of Non-Food Biomass and Enzyme Technology, Guangxi Biomass Engineering Technology Research Center, Guangxi Key Laboratory of Biorefinery, Guangxi Academy of Sciences, 98 Daling Road, Nanning, Guangxi 530007, China. Email: hongguanglishibahao@gxas.cn

Abstract

Coronavirus (CoV) 3-chymotrypsin (C)-like cysteine protease (3CL^{pro}) is a target for anti-CoV drug development and drug repurposing because along with papain-like protease, it cleaves CoV-encoded polyproteins (pp1a and pp1ab) into nonstructural proteins (nsps) for viral replication. However, the cleavage sites of 3CL^{pro} and their relevant nsps remain unclear, which is the subject of this perspective. Here, we address the subject from three standpoints. First, we explore the inconsistency in the cleavage sites and relevant nsps across CoVs, and investigate the function of nsp11. Second, we consider the nsp16 mRNA overlapping of the spike protein mRNA, and analyze the effect of this overlapping on mRNA vaccines. Finally, we study nsp12, whose existence depends on ribosomal frameshifting, and investigate whether 3CL^{pro} requires a large number of inhibitors to achieve full inhibition. This perspective helps us to clarify viral replication and is useful for developing anti-CoV drugs with 3CL^{pro} as a target in the current coronavirus disease 2019 (COVID-19) pandemic.

KEYWORDS

3CL^{pro}, COVID-19, MERS-CoV, nsp, SARS-CoV, SARS-CoV-2

1 | INTRODUCTION

The 3-chymotrypsin (C)-like cysteine protease (3CL^{pro}) of coronaviruses (CoVs) is a target for developing antiviral drugs against SARS-CoV,¹⁻³ MERS-CoV,⁴⁻⁶ and SARS-CoV-2.⁷⁻¹⁰ CoVs encode four structural and accessory proteins: spike protein (S-protein), envelope protein (E-protein),

membrane protein (M-protein), and nucleocapsid protein (N-protein); and two replicase polyproteins (pp1a and pp1ab). The open reading frames (ORFs) 1a and 1b encode pp1a and pp1ab,¹¹⁻¹³ which are cleaved by papain-like protease (PL^{pro}) and 3CL^{pro}.¹⁴

pp1ab can be cleaved into 16 nonstructural proteins (nsps).^{15,16} PL^{pro} is released from nsp3 through autocleavage,

Abbreviations: 2'-O-MTase, 2'-O-methyltransferase; 3CL^{pro}, 3-chymotrypsin-like cysteine protease; BCoV, bovine coronavirus; BEV, Berne virus; CoV, coronavirus; COVID-19, coronavirus disease 2019; DMV, double-membraned vesicles; E-protein, envelope protein; ExoN, 3' to 5' exonuclease; FCoV, feline coronavirus; HCoV, human coronavirus; HCV, hepatitis C virus; IBV, infectious bronchitis virus; MERS, Middle East Respiratory Syndrome; MHV, murine hepatitis virus; M^{pro}, main protease; M-protein, membrane protein; NendoU, nidoviral endoribonuclease specific for U; N-protein, nucleocapsid protein; nsp, nonstructural protein; nt, nucleotide; ORF, open reading frame; PEDV, porcine epidemic diarrhea virus; PL^{pro}, papain-like protease; RdRp, RNA-dependent RNA polymerase; S-protein, spike protein; SADS-CoV, swine acute diarrhea syndrome coronavirus; SARS, severe acute respiratory syndrome; SARS-CoV, severe acute respiratory syndrome coronavirus; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; TRS, transcription-regulating sequence.

and then cleaves three sites: between nsp1/2, nsp2/3, and nsp3/4 with the LXGG recognition motif.¹⁷ 3CL^{pro}—also called the main protease (M^{pro})—is released from nsp5 through autocleavage.¹⁸ Using recognition motifs such as XXXLQAXXX and XXXLQSXXX,¹⁹ 3CL^{pro} cleaves the remaining nsps.²⁰

3CL^{pro} is a desirable target for developing wide-spectrum anti-CoV drugs and drug repurposing^{21–24} because (a) the nsps cleaved by 3CL^{pro} influence the formation of the double-membraned vesicles (DMV),^{18,25–31} (b) 3CL^{pro} shares significant sequence identity and 3D structure among CoVs,^{32,33} (c) human proteases do not have identical cleavage specificity,⁸ and (d) 3CL^{pro} inhibitors are unlikely to be toxic to humans.⁸

However, 3CL^{pro} has the potential for industrial application because its autocleavage occurs near the membrane³⁴ or in the cytoplasm;^{35–37} therefore, its efficacy is higher than that of other proteases.³⁵

Much information on the 3D structure,³⁸ mechanism,³⁹ and substrate specificity⁴⁰ of 3CL^{pro} has been elucidated. For example, 3CL^{pro} has at least three crystal structures⁴¹: the wild-type active dimer (wt-dimer) with one active site per subunit^{1,35}; the monomeric form or the G11A, R298A, and S139A mutants that cannot dimerize^{42–45}; and the super-active octamer form.³ Therefore, two strategies have been employed to develop 3CL^{pro} inhibitors based on their active sites and dimerization.^{46–48}

So far, however the cleavage sites by 3CL^{pro} and cleaved nsps remain controversial. For instance, a study based on 24 completely sequenced CoV genomes predicted 11 cleavage sites between nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp9, nsp10, nsp11, nsp12, and nsp13,⁴⁹ but found that the cleavage site between nsp7/8 in pp1a and the cleavage site between nsp7/9 in pp1ab are the same.⁴⁹ Here, there is no nsp8.

3CL^{pro} has 11 cleavage sites, that is, between nsp4/5, nsp5/6, nsp6/7, nsp7/8, nsp8/9, nsp9/10, nsp10/11, nsp12/13, nsp13/14, nsp14/15, and nsp15/16.^{26,49–51} Here, there is no cleavage site between nsp11/12.

Sometimes, a cleavage site by 3CL^{pro} is defined between nsp10/12.^{10,21} Here, there is no nsp11.

Although these discrepancies seem trivial, they are essential for 3CL^{pro} action and are instrumental for developing 3CL^{pro} inhibitors. In this perspective, we closely explore these inconsistent cleavage sites by 3CL^{pro} and their nsps, and discuss their possible relevance to coronavirus disease 2019 (COVID-19) vaccine and drug development.

2 | CLEAVAGE SITES BY 3CL^{pro}

Initially, 13 probable cleavage sites with dipeptides Q/S(G) were predicted for 3CL^{pro} from infectious bronchitis virus (IBV), and cleavage sites 10, 11, 12, and 13 were located in F2 polyprotein (pp1b).²⁰ Subsequently, it was revealed that

these 13 cleavage sites include 2 cleavage sites—sites 3 and 4—flanking 3CL^{pro} for autocleavage.²⁰ Thus, 3CL^{pro} cleaves 11 sites.²⁰ These findings seem to show that 3CL^{pro} is located in nsp3 in IBV, but 3CL^{pro} is actually located in nsp2 in IBV per the current annotation (accession no. NC_001451.1). In addition, nsp8 is positioned from nucleotide (nt) 12313 to nt 12381 in pp1a, but is absent in pp1ab, where it merges with nsp9 from nt 12313 to nt 12354, and from nt 13354 to nt 15131.

Shortly after, seven cleavage sites were experimentally identified in ORF1b from murine CoV mouse hepatitis virus (MHV),⁵² which generated eight nsps rather than the five nsps generally accepted in pp1b.

Later, it was accepted that PL^{pro} and 3CL^{pro} in HCoV-229E RNA-directed RNA polymerase ORF1A (accession no. 464694) have 12 cleavage sites, 1 site cleaved by PL^{pro} and 11 sites cleaved by 3CL^{pro}, generating 13 nsps.⁵⁰ Here, the number of nsps in pp1a is more than that currently accepted—pp1a has 11 nsps from nsp1 to nsp11 and pp1ab has 16 nsps because ORF1b encodes 5 nsps from nsp12 to nsp16.¹⁵ In addition, HCoV-229E has two copies of PL^{pro} and generates two nsps, whereas most CoVs have a single PL^{pro} copy and generate three nsps.

Eventually, 3CL^{pro} cleaves 9 sites rather than 11 cleavage sites, 2 of which form its own flanking N- and C-terminal autoprocessing sites.^{38,53} This is plausible because 3CL^{pro} is located in nsp5, and its own N- and C-terminal autoprocessing sites are located at the cleavage sites between nsp4/5 and nsp5/6. In other words, 3CL^{pro} can trans-cleave five sites in pp1a—between nsp6/7, nsp7/8, nsp8/9, nsp9/10, and nsp10/11 and trans-cleave four sites in pp1b—between nsp12/13, nsp13/14, nsp14/15, and nsp15/16.

Of the frequently referenced CoVs, nsps were clearly annotated in nine CoVs (Table 1). As shown in Table 1, the identified nsps were inconsistent across different CoVs. For example, nsp11 is absent but merges with nsp12 in MHV (A59 C12 mutant) and SARS-CoV (Tor2). In another example, nsp8, nsp14, nsp15, and nsp16 are absent from avian IBV (Beaudette), porcine epidemic diarrhea virus (PEDV) CV777, and two SARS-CoVs (CUHK-W1 and TW1), but their nsp8 and nsp9 merge.

This inconsistency is interesting because the sizes of nsps cleaved by 3CL^{pro} are highly conserved among different groups of CoVs, whereas the sizes of nsps cleaved by PL^{pro} are irregular.⁴⁹

Table 1 shows the existence of 16 nsps for several CoVs. This is interesting because 15 cleavage sites are required to generate 16 nsps. However, PL^{pro} cleaves 3 sites and 3CL^{pro} cleaves 11 sites—which sums up to 14 cleavage sites—generating 15 nsps only. Indeed, 15 nsps are often mentioned,⁵⁴ and therefore, the generation of 15 nsps is arithmetically correct. Intriguingly, a cleavage site is missing from such calculations.

TABLE 1 nsp5 in nine clearly annotated CoVs among the frequently referenced CoVs

Type	Avian IBV	FCoV	MHV	PEDV	SARS-CoV	SARS-CoV	SARS-CoV	MERS-CoV	SARS-CoV-2
Strain	Beaudette	FIPV 79-1146	A59 C12 mutant	CV777	Tor2	CUHK-W1	TW1	HCoV-EMC/2012	Wuhan-Hu-1
ID	NC_001451.1	DQ010921	NC_001846.1	NC_003436.1	NC_004718.3	AY278554.2	AY291451.1	NC_019843.3	NC_045512.2
Total bp	27608	29147	31357	28033	29751	29736	29729	30119	29903
nsp1	2548-8865	312-641	210-950	2982-7847	265-804	2704-9969	2719-9984	279-857	266-805
nsp2	8866-9786	642-2948	951-2705	9288-10193	805-2718	9970-10887	9985-10902	858-2837	806-2719
nsp3	9787-10665	2949-7319	2706-9632	10194-11033	2719-8484	10888-11757	10903-11772	2838-8498	2720-8554
nsp4	10666-10914	7320-8789	9633-10208	11034-11282	8485-9984	11758-12006	11773-12021	8499-10019	8555-10054
nsp5	10915-11544	8790-9686	10209-11117	11283-11867	9985-10902	12007-12600	12022-12615	10020-10937	10055-10972
nsp6	11545-11877	9687-10577	11118-11978	11868-12191	10903-11772	12601-12939	12616-12954	10938-11813	10973-11842
nsp7	11878-12312	10578-10826	11979-12245	12192-12596	11773-12021	12940-13356	12955-13371	11814-12062	11843-12091
nsp8		10827-11411	12246-12836		12022-12615			12063-12659	12092-12685
nsp9	Join 12313-12354, 12354-15131	11412-11744	12837-13166	Join 12597-12620, 12620-15376	12616-12954	Join 13357-13383, 13383-16151	Join 13372-13398, 13398-16166	12660-12989	12686-13024
nsp10	15132-16931	11745-12149	13167-13577	15377-17167	12955-13371	16152-17954	16167-17969	12990-13409	13025-13441
nsp11	16932-18494	12150-12206		17168-18718		17955-19535	17970-19550	13410-13451	13442-13480
nsp12	18495-19508	Join 12150-12176, 12176-14935	Join 13578-13619, 13619-16360	18719-19735	Join 13372-13392, 13392-13394	19536-20573	19551-20588	Join 13410-13433, 13433-16207	Join 13442-13468, 13468-16236
nsp13	19509-20414	14936-16732	16361-18160	19736-20638	16167-17969	20574-21467	20589-21482	16208-18001	16237-18039
nsp14		16733-18289	18161-19723		17970-19550			18002-19573	18040-19620
nsp15		18290-19306	19724-20845		19551-20588			19574-20602	19621-20658
nsp16		19307-20206	20846-21742		20589-21482			20603-21511	20659-21552

No predictable cleavage sites were observed between nsp10/11 and nsp11/12 in CoVs using the NetCorona 1.0 webserver.¹⁹ Additionally, the cleavage site between nsp11/12 in SARS-CoV, MERS-CoV, and SARS-CoV-2 cannot be found in literature although that between nsp10/12 is mentioned.²¹ pp1ab is a fusion between nsp11 and nsp12 when a ribosomal frameshifting occurs between ORF1a and ORF1b⁵⁵⁻⁵⁷; thus, it is likely that the missing cleavage site is between nsp11/12, which is the cleavage site between pp1a and pp1b. Hence, a convincing explanation is that nsp11 appears only when pp1a exists, in which case the ribosomal frameshifting does not occur. Consequently, there are 15 nsps for pp1ab, but the co-existence of pp1a and pp1ab provides 16 nsps. Here, the function of nsp11 remains unclear?

nsp14, nsp15, and nsp16 are absent from the four CoVs (Table 1). Meanwhile, a comparison based on the data^{58,59} shows the absence of nsp11 and nsp16, and total 14 nsps in both SARS-CoV and SARS-CoV-2.¹⁰ The absence of nsp16 in pp1ab draws our attention, leading us to explore nsp16 in the next section.

In summary, this section indicates that (a) nsp11 exists in pp1a only but not in pp1ab, and there is no cleavage site between nsp11/12; (b) the 11 3CL^{pro} cleavage sites are only applicable to pp1ab; and (c) the function of nsp11 is yet to be determined.

3 | nsp16 IN CoVs

nsp16 is an RNA cap-modifying enzyme,⁶⁰ and forms a complex with nsp10.⁶⁰⁻⁶² Therefore, it plays the role of 2'-O-methyltransferase (2'-O-MTase) in CoVs, which was first found in the nsp16 from feline CoV (FCoV) FIPV 79-1146⁶³; in the FASTA format of the FCoV genome (accession no, DQ010921), the mRNA for nsp16 is positioned at nt 19307—nt 20206, whereas the mRNA for S-protein is positioned at nt 20206—nt 24564, thus, their ORFs overlap at a single nt, 20206. In the graphics format of this FCoV genome, the mRNA for nsp16 is positioned at nt 19307—nt 20209, whereas the mRNA for S-protein is positioned at nt 20206—nt 24564, thereby, their ORFs overlap at three nucleotides, 20206-20209.

Table 2 lists 39 frequently referenced CoV genomes, where ORF1ab overlaps the mRNA for S-protein in 22 CoVs († in the last column in Table 2), that is, the mRNA for pp1ab overlaps the mRNA for S-protein in some groups of CoVs. This feature is more remarkable in CoVs from avian, bat, feline, swine as well as MERS-CoVs. The overlapping can be as large as 58 nts for MERS-CoV (the penultimate row in Table 2).

The overlap of ORF1ab on the subgenomic mRNA in CoVs is intriguing because the size of proteins translated from

subgenomic mRNA is theoretically that of non-overlapping coding regions.⁶⁴ Most subgenomic mRNAs are structurally polycistronic but functionally monocistronic, and thus their translation begins only from the 5' of most ORF in viral proteins.⁶⁵

In the 1990s, the S-protein gene in equine isolate Berne virus (BEV), a torovirus,⁶⁶ overlapped the replicase gene (ORF1b).⁶⁴ Therefore, it was proposed that a motif, UGUUUAGU, directs the synthesis of the S-protein gene.⁶⁷ Subsequently, this S-protein gene (mRNA 2) was found to have a short non-contiguous leader coming from the 5' terminus of the BEV genome.⁶⁸ Here, the overlapping results from a heterologous RNA recombination.⁶⁸

In contrast to torovirus, CoVs have a common 5' leader sequence,²⁹ which protects the capped mRNAs from nsp1-induced endonucleolytic cleavage, causing the accumulation of SARS-CoV mRNAs and proteins.⁶⁹ In CoV lifecycle, the production of pp1ab occurs in the first phase of translation, whereas S-protein production by subgenomic mRNA occurs in the second phase.⁷⁰ Between these two phases is the replication of minus- and plus-strand RNA in DMV. It is also unclear whether such overlapping affects nsp16 and S-protein functions, and the fusion of S-protein with a leader sequence, which includes the transcription-regulating sequence (TRS). This leads to the question of the selective advantage of CoV with the leader sequence in its subgenomic mRNAs.⁶⁵

As the RNA-dependent RNA polymerase (RdRp), nsp12,⁷¹ catalyzes leader-body fusion, we closely explore nsp12 in the next section.

In summary, this section focuses on the overlapping of the S-protein mRNA by the nsp16 mRNA because several mRNA vaccines are based on the S-protein mRNA,⁷²⁻⁷⁴ whereas the overlapping could be a potential source for heterologous RNA recombination.

4 | nsp12 IN CoVs

RdRp (nsp12) is a target for developing anti-CoV drugs.^{16,75} Usually, nsp7 and nsp8 act as nsp12 cofactors.⁷⁶ Remdesivir, currently authorized by the FDA for emergency use,⁷⁷ was originally designed to target the polymerases in HIV and hepatitis C virus (HCV).⁷⁸

The first high-resolution cryo-electron microscopy structure of SARS-CoV-2 full-length nsp12 has 932 residues,⁷⁹ which is the size of joined nts 13442-13468 and 13468-16236 (SARS-CoV-2, Table 1). This illustration again confirms that nsp11 does not exist as a single entity in pp1ab; nsp11 only appears when ribosomal frameshifting does not occur, whereas ribosomal frameshifting is the only way to generate the joined nsp12, RdRp.

Here, one may ask whether RdRp and nsp11 must merge to function and whether nsp12 alone is nonfunctional. ORF1b

TABLE 2 Overlap (†) and non-overlap of nsp16 mRNA over the S-protein mRNA in 39 frequently referenced CoVs

Accession no	Strain or isolate	pp1ab (nt)	S-Protein (nt)
NC_001451.1	IBV strain Beaudette	529-20417	20368-23856†
NC_048213.1	IBV isolate Ind-TN92-03	529-20423	20374-23835†
JF732903.1	IBV strain Sczy3	526-20414	20365-23862†
MN711790.1	IBV isolate GA/1472/2004	529-20408	20359-23865†
MT460496.1	IBV isolate CK/CH/LAH/1806	526-12387 12462-20420	20371-23898†
NC_048212.1	Bat CoV isolate CMR704-P12	210-20842	20814-24623†
NC_028824.1	BtRf-AlphaCoV/YN2012	135-20284	20281-23679†
MF370205.1	Rhinolophus bat CoV HKU2 isolate swine enteric alphacoronavirus CH/GD-01/2017/P2	297-20482	20479-23871†
AF220295.1	BCoV strain Quebec	211-13362 13332-21389	23655-27746
AF391542.1	BCoV isolate BCoV-LUN	211-21494	23641-27732
NC_003045.1	BCoV isolate BCoV-ENT	211-21494	23641-27732
U00735.2	Bovine CoV strain Mebus	211-21494	23641-27732
DQ010921	FCoV strain FIPV 79-1146	312-20209	20206-24564†
NC_002306.3	Feline infectious peritonitis virus isolate 79-1146	311-20439	20436-24794†
NC_001846.1	MHV A59 C12 mutant	210-21745	23929-27903
AF201929.1	MHV strain 2	210-13460 13382-21583	23755-27840
AF208066.1	MHV strain Penn 97-1	210-13460 13382-21580	23712-27677
AF208067.1	MHV strain ML-10	210-13613 13535-21736	23867-27841
NC_003436.1	PEDV strain CV777	297-20641	20638-24789†
NC_028806.1	Swine enteric CoV strain Italy/213306/2009	307-12354 12312-20354	20355-24503
KR610993.1	PEDV clone CBR1	1-20345	20342-24499†
MF769442.1	SADS-CoV isolate DCD5	304-20489	20486-23878†
MK994937.1	SADS-CoV isolate GDWT-P83	304-20489	20486-23878†
MT039231.1	Mutant SADS-CoV strain icSADS	312-20497	20494-23886†
MT747188.1	SADS-CoV isolate CN/GDST/2017	304-20489	20486-23878†
NC_002645.1	HCoV-229E	293-20568	20570-24091
NC_004718.3	SARS-CoV Tor2	265-21485	21492-25259
AY278488.2	SARS-CoV BJ01	246-21466	21473-25240
AY278554.2	SARS-CoV CUHK-W1	250-21470	21477-25244
AY278741.1	SARS-CoV Urbani	265-21485	21492-25259
AY282752.2	SARS-CoV CUHK-Su10	250-21470	21477-25244
AY291451.1	SARS-CoV TW1	265-21485	21492-25259
NC_019843.3	MERS-CoV isolate HCoV-EMC/2012	279-21514	21456-25517†
KT029139.1	MERS-CoV/KOR/KNIH/002_05_2015	279-21514	21456-25517†
MF598722.1	MERS-CoV strain camel/UAE_415915_W6_2015	279-21514	21456-25517†
MG596803.1	MERS-CoV/Bat-CoV/P.khulii/Italy/206645-63/2011	208-21437	21379-25416†
MK967708.1	MERS-CoV isolate Merscov/Egypt/Camel/AHRI-FAO-1/2018	268-21503	21445-25505†
MN120514.1	MERS-CoV isolate 013	279-21514	21456-25517†
NC_045512.2	SARS-CoV-2 isolate Wuhan-Hu-1	266-21555	21563-25384

can encode a polyprotein only if ribosomal frameshifting from ORF1a into ORF1b occurs^{80,81} because ORF1b does not have an independent site for translation initiation.⁸² Thus, pp1b does not exist alone, and consequently nsp12 does not exist without nsp11.

Hence, the key point for generating RdRp is the -1 ribosomal frameshifting,⁸² which is the focus of many studies.^{50,83-85} This ribosomal frameshifting overlaps two ORFs with 43 nucleotides in HCoV-229E51, with 86 nucleotides in IBV,⁸³ and with 130 nucleotides in the L-A double-stranded RNA virus of *Saccharomyces cerevisiae*.⁸⁶ The occurrence of ribosomal frameshifting is required for the production of RdRp,⁷⁰ which is needed in the second phase of translation.⁷⁰ Although the signal for ribosomal frameshifting has been well studied^{57,80,87-90} and the RNA pseudoknot is designed as a target for anti-SARS agents,^{91,92} regulation of the ribosomal frameshifting mechanism remains unclarified.

A simple explanation of the mechanism is that the expression of ORF1b is initiated at specific levels relative to pp1a in CoVs,⁸² assuming there is a threshold for the occurrence of ribosomal frameshifting based on the ratio of pp1ab to pp1a. This ratio ranges from 1.8% to 1.9% in the L-A double-stranded RNA virus of *S. cerevisiae*,^{86,93} and from 5% to 10% in HIV-1.^{94,95} For HCoV-229E, the ribosomal frameshifting frequency ranges between 18% and 30%, of which less than 1% can synthesize pp1ab.⁵⁵ Thus, the chance of generating pp1ab in CoVs is remarkably lower than that in the L-A double-stranded RNA virus of *S. cerevisiae*,^{86,93} and HIV-1.^{94,95} Indeed, both pp1a and pp1ab from CoVs are difficult to detect in vivo.⁹⁶

This implies that many copies of pp1a, but few copies of pp1ab occur in CoVs. Although 3CL^{pro} is derived from nsp5 in both pp1a and pp1ab, most of 3CL^{pro} is from pp1a-derived. Thus, not every 3CL^{pro} has a good chance of cleaving the sites between nsp12/13, 13/14, 14/15, and nsp15/16 in pp1b.

In summary, this section reveals that the 3CL^{pro} significantly outnumbers its cleavage sites in pp1b; thus, a considerable number of 3CL^{pro} inhibitors are required to completely inhibit 3CL^{pro} action.

5 | CONCLUSION

In this perspective, we attempted to address several controversial issues on the cleavage sites and cleaved nsps in CoV pp1a and pp1ab by 3CL^{pro}. Meanwhile, several questions are raised: (a) what function does nsp11 perform, (b) can the overlap of nsp16 mRNA over the S-protein mRNA affect mRNA vaccine, and (c) can a low ribosomal frameshifting frequency affect 3CL^{pro} activity? The answers to these questions will enrich our understanding of the mechanism of viral

replication and benefit the development of anti-CoV drugs with 3CL^{pro} as targets in the current COVID-19 pandemic.

ACKNOWLEDGMENTS

This study was supported by Scientific Development Fund of Guangxi Academy of Sciences and Key Project of Guangxi Scientific Research and Technology Development Plan (AB17190534).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

G. Wu designed this perspective and wrote the first draft. Both finalized this manuscript.

ORCID

Shaomin Yan  <https://orcid.org/0000-0001-7642-3972>

Guang Wu  <https://orcid.org/0000-0003-0775-5759>

REFERENCES

1. Yang H, Yang M, Ding Y, et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci USA*. 2003;100:13190-13195.
2. Yang H, Xie W, Xue X, et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol*. 2005;3:e324.
3. Zhang S, Zhong N, Xue F, et al. Three-dimensional domain swapping as a mechanism to lock the active conformation in a super-active octamer of SARS-CoV main protease. *Protein Cell*. 2010;1:371-383.
4. Abuhammad A, Al-Aqtash RA, Anson BJ, Mesecar AD, Taha MO. Computational modeling of the Bat HKU4 coronavirus 3CL(pro) inhibitors as a tool for the development of antivirals against the emerging Middle East Respiratory Syndrome (MERS) coronavirus. *J Mol Recognit*. 2017;30:e2644.
5. Galasiti Kankanamalage AC, Kim Y, Damalanka VC, et al. Structure-guided design of potent and permeable inhibitors of MERS coronavirus 3CL protease that utilize a piperidine moiety as a novel design element. *Eur J Med Chem*. 2018;150:334-346.
6. Kumar V, Tan KP, Wang YM, Lin SW, Liang PH. Identification, synthesis and evaluation of SARS-CoV and MERS-CoV 3C-like protease inhibitors. *Bioorg Med Chem*. 2016;24:3035-3042.
7. Jin Z, Du X, Xu Y, et al. Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature*. 2020;582:289-293.
8. Zhang L, Lin D, Sun X, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*. 2020;368:409-412.
9. Gordon DE, Jang GM, Bouhaddou M, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 2020;583:459-468.
10. Míczí M, Golda M, Kunkli B, Nagy T, Tózsér J, Mótóyán JA. Identification of host cellular protein substrates of SARS-COV-2 main protease. *Int J Mol Sci*. 2020;21:E9523.
11. Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*. 2003;300:1394-1399.

12. Marra MA, Jones SJ, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. *Science*. 2003;300:1399-1404.
13. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature*. 2020;579:265-269.
14. Ziebuhr J. The coronavirus replicase. *Curr Top Microbiol Immunol*. 2005;287:57-94.
15. Báez-Santos YM, St John SE, Mesecar AD. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antiviral Res*. 2015;115:21-38.
16. Cannalire R, Cerchia C, Beccari AR, Di Leva FS, Summa V. Targeting SARS-CoV-2 proteases and polymerase for COVID-19 treatment: state of the art and future opportunities. *J Med Chem*. 2020. <https://doi.org/10.1021/acs.jmedchem.0c01140>
17. Harcourt BH, Jukneliene D, Kanjanahaluethai A, et al. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J Virol*. 2004;78:13600-13612.
18. Snijder EJ, Bredendek PJ, Dobbe JC, et al. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol*. 2003;331:991-1004.
19. Kiemer L, Lund O, Brunak S, Blom N. Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. *BMC Bioinform*. 2004;5:72.
20. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res*. 1989;17:4847-4861.
21. Chen YW, Yiu CB, Wong KY. Prediction of the SARS-CoV-2. (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Res*. 2020;9:129.
22. Elshabrawy HA. SARS-CoV-2: an update on potential antivirals in light of SARS-CoV antiviral drug discoveries. *Vaccines (Basel)*. 2020;8:335.
23. Durojaiye AB, Clarke J-RD, Stamatiades GA, Wang C. Repurposing cefuroxime for treatment of COVID-19: a scoping review of *in silico* studies. *J Biomol Struct Dyn*. 2020;38:1-8.
24. Joshi S, Joshi M, Degani MS. Tackling SARS-CoV-2: proposed targets and repurposed drugs. *Future Med Chem*. 2020;12(17):1579-1601. <https://doi.org/10.4155/fmc-2020-0147>
25. Stadler K, Massignani V, Eickmann M, et al. SARS—beginning to understand a new virus. *Nat Rev Microbiol*. 2003;1:209-218.
26. Thiel V, Ivanov KA, Putics A, et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol*. 2003;84:2305-2315.
27. Sawicki SG, Sawicki DL, Younker D, et al. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. *PLoS Pathog*. 2005;1:e39.
28. Sawicki SG, Sawicki DL, Siddell SG. A contemporary view of coronavirus transcription. *J Virol*. 2007;81:20-29.
29. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense? *J Gen Virol*. 2006;87:1403-1421.
30. Masters PS. The molecular biology of coronaviruses. *Adv Virus Res*. 2006;66:193-292.
31. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*. 2009;7:439-450.
32. Wang F, Chen C, Tan W, Yang K, Yang H. Structure of main protease from human coronavirus NL63: insights for wide spectrum anti-coronavirus drug design. *Sci Rep*. 2016;6:22677.
33. Yang H, Bartlam M, Rao Z. Drug design targeting the main protease, the Achilles' heel of coronaviruses. *Curr Pharm Des*. 2006;12:4573-4590.
34. Anand K, Ziebuhr J, Wadhvani P, Mesters JR, Hilgenfeld R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*. 2003;300:1763-1767.
35. Xue X, Yang H, Shen W, et al. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J Mol Biol*. 2007;366:965-975.
36. Hsu M-F, Kuo C-J, Chang K-T, et al. Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem*. 2005;280:31257-31266.
37. Lin CW, Tsai CH, Tsai FJ, Chen PJ, Lai CC, Wan L. Characterization of trans- and cis-cleavage activity of the SARS coronavirus 3CLpro protease: basis for the in vitro screening of anti-SARS drugs. *FEBS Lett*. 2004;574:131-137.
38. Muramatsu T, Takemoto C, Kim YT, et al. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc Natl Acad Sci USA*. 2016;113:12997-13002.
39. Wang H, He S, Deng W, et al. Comprehensive insights into the catalytic mechanism of Middle East Respiratory Syndrome 3C-like protease and Severe Acute Respiratory Syndrome 3C-like protease. *ACS Catal*. 2020;10:5871-5890.
40. Chuck CP, Chong LT, Chen C, Chow HF, Wan DC, Wong KB. Profiling of substrate specificity of SARS-CoV 3CL. *PLoS ONE*. 2010;5:e13197.
41. Xia B, Kang X. Activation and maturation of SARS-CoV main protease. *Protein Cell*. 2011;2:282-290.
42. Chen S, Hu T, Zhang J, et al. Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. *J Biol Chem*. 2008;283:554-564.
43. Chen S, Zhang J, Hu T, Chen K, Jiang H, Shen X. Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis. *J Biochem*. 2008;143:525-536.
44. Shi J, Sivaraman J, Song J. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *J Virol*. 2008;82:4620-4629.
45. Hu T, Zhang Y, Li L, et al. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. *Virology*. 2009;388:324-334.
46. 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;279:1637-1642.
47. Kuo CJ, Chi YH, Hsu JTA, Liang PH. Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem Biophys Res Commun*. 2004;318:862-867.
48. Barrila J, Bacha U, Freire E. Long-range cooperative interactions modulate dimerization in SARS 3CL^{pro}. *Biochemistry*. 2006;45:14908-14916.
49. Gao F, Ou HY, Chen LL, Zheng WX, Zhang CT. Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Lett*. 2003;553:451-456.

50. Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. *J Gen Virol*. 2000;81:853-879.
51. Ziebuhr J, Herold J, Siddell SG. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *J Virol*. 1995;69:4331-4338.
52. Denison MR, Zoltick PW, Leibowitz JL, Pachuk CJ, Weiss SR. Identification of polypeptides encoded in open reading frame 1b of the putative polymerase gene of the murine coronavirus mouse hepatitis virus A59. *J Virol*. 1991;65:3076-3082.
53. Chan JF, Kok KH, Zhu Z, et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 2020;9:221-236.
54. Chen S, Jonas F, Shen C, Hilgenfeld R. Liberation of SARS-CoV main protease from the viral polyprotein: n-terminal autocleavage does not depend on the mature dimerization mode. *Protein Cell*. 2010;1:59-74.
55. Herold J, Raabe T, Schelle-Prinz B, Siddell SG. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology*. 1993;195:680-691.
56. Herold J, Siddell SG. An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res*. 1993;21:5838-5842.
57. Namy O, Moran SJ, Stuart DI, Gilbert RJC, Brierley I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature*. 2006;441:244-247.
58. Schechter I, Berger A. On the size of the active site in proteases. I. papain. *Biochem Biophys Res Commun*. 1967;27:157-162.
59. Stoddard SV, Stoddard SD, Oelkers BK, et al. Optimization rules for SARS-CoV-2 Mpro antivirals: ensemble docking and exploration of the coronavirus protease active site. *Viruses*. 2020;12:942.
60. Decroly E, Debarnot C, Ferron F, et al. Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog*. 2011;7:e1002059.
61. Chen Y, Su C, Ke M, et al. Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. *PLoS Pathog*. 2011;7:e1002294.
62. Hilgenfeld R, Peiris M. From SARS to MERS: 10 years of research on highly pathogenic human coronaviruses. *Antiviral Res*. 2013;100:286-295.
63. Decroly E, Imbert I, Coutard B, et al. Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J Virol*. 2008;82:8071-8084.
64. Snijder EJ, Horzinek MC, Spaan WJ. A 3'-coterminal nested set of independently transcribed mRNAs is generated during Berne virus replication. *J Virol*. 1990;64:331-338.
65. Sola I, Almazán F, Zúñiga S, Enjuanes L. Continuous and discontinuous RNA synthesis in coronaviruses. *Annu Rev Virol*. 2015;2:265-288.
66. Snijder EJ, Horzinek MC. Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. *J Gen Virol*. 1993;74:2305-2316.
67. Snijder EJ, Den Boon JA, Spaan WJ, Weiss M, Horzinek MC. Primary structure and post-translational processing of the Berne virus peplomer protein. *Virology*. 1990;178:355-363.
68. van Vliet AL, Smits SL, Rottier PJ, de Groot RJ. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J*. 2002;21:6571-6580.
69. Huang C, Lokugamage KG, Rozovics JM, Narayanan K, Semler BL, Makino S. SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog*. 2011;7:e1002433.
70. Plant EP, Dinman JD. The role of programmed -1 ribosomal frameshifting in coronavirus propagation. *Front Biosci*. 2008;13:4873-4881.
71. Cheng A, Zhang W, Xie Y, et al. Expression, purification, and characterization of SARS coronavirus RNA polymerase. *Virology*. 2005;335:165-176.
72. Jackson LA, Anderson EJ, Roupheal NG, et al. An mRNA vaccine against SARS-CoV-2 - preliminary report. *N Engl J Med*. 2020;383:1920-1931.
73. Zhang NN, Li XF, Deng YQ, et al. A thermostable mRNA vaccine against COVID-19. *Cell*. 2020;182:1271-1283.
74. Corbett KS, Edwards DK, Leist SR, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature*. 2020;586:567-571.
75. Zhang Y, Tang LV. Overview of targets and potential drugs of SARS-CoV-2 according to the viral replication. *J Proteome Res*. 2021;20:49-59.
76. Subissi L, Posthuma CC, Collet A, et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci USA*. 2014;111:E3900-E3909.
77. Remdesivir EUA Fact Sheet for Healthcare Providers - updated October 22, 2020
78. Cho A, Saunders OL, Butler T, et al. Synthesis and antiviral activity of a series of 1'-substituted 4-aza-7,9-dideazaadenosine C-nucleosides. *Bioorg Med Chem Lett*. 2012;22:2705-2707.
79. Gao Y, Yan L, Huang Y, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. 2020;368:779-782.
80. Brierley I, Bournsnel ME, Binns MM, et al. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J*. 1987;6:3779-3785.
81. Bredenbeek PJ, Pachuk CJ, Noten AF, et al. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59: a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res*. 1990;18:1825-1832.
82. Baranov PV, Henderson CM, Anderson CB, Gesteland RF, Atkins JF, Howard MT. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*. 2005;332:498-510.
83. Brierley I, Digard P, Inglis SC. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell*. 1989;57:537-547.
84. Brierley I, Jenner AJ, Inglis SC. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol*. 1992;227:463-479.
85. Nam SH, Copeland TD, Hatanaka M, Oroszlan S. Characterization of ribosomal frameshifting for expression of pol gene products of human T-cell leukemia virus type I. *J Virol*. 1993;67:196-203.
86. Dinman JD, Wickner RB. Ribosomal frameshifting efficiency and gag/gag-pol ratio are critical for yeast M1 double-stranded RNA virus propagation. *J Virol*. 1992;66:3669-3676.

87. Jacks T, Madhani HD, Masiarz FR, Varmus HE. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell*. 1988;55:447-458.
88. Plant EP, Pérez-Alvarado GC, Jacobs JL, Mukhopadhyay B, Hennig M, Dinman JD. A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol*. 2005;3:e172.
89. Giedroc DP, Cornish PV. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res*. 2009;139:193-208.
90. Peselis A, Serganov A. Structure and function of pseudoknots involved in gene expression control. *Wiley Interdiscip Rev RNA*. 2014;5:803-822.
91. Park SJ, Kim YG, Park HJ. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J Am Chem Soc*. 2011;133:10094-10100.
92. Chen Y, Tao H, Shen S, et al. A drug screening toolkit based on the -1 ribosomal frameshifting of SARS-CoV-2. *Heliyon*. 2020;6:e04793.
93. Dinman JD, Icho T, Wickner RB. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc Natl Acad Sci USA*. 1991;88:174-178.
94. Hung M, Patel P, Davis S, Green SR. Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J Virol*. 1998;72:4819-4824.
95. Biswas P, Jiang X, Pacchia AL, Dougherty JP, Peltz SW. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol*. 2004;78:2082-2087.
96. Ziebuhr J, Thiel V, Gorbalenya AE. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. *J Biol Chem*. 2001;276: 33220-33232.

How to cite this article: Yan S, Wu G. Potential 3-chymotrypsin-like cysteine protease cleavage sites in the coronavirus polyproteins pp1a and pp1ab and their possible relevance to COVID-19 vaccine and drug development. *The FASEB Journal*. 2021;35:e21573. <https://doi.org/10.1096/fj.202100280RR>