



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Optimization of the expression of the main protease from SARS-CoV-2

Yi Rong, Chaofeng Zhang^{**}, Wen-Chao Gao, Cheng Zhao^{*}

College of Biomedical Engineering, Taiyuan University of Technology, Taiyuan, 030024, People's Republic of China

ARTICLE INFO

Keywords:

SARS-CoV-2
Main protease
Enzyme activity
Additional residues
Kinetic parameters

ABSTRACT

The main protease (M^{pro}) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) plays a vital role in viral replication. To study the function of M^{pro} and screen inhibitors targeting M^{pro} , it is necessary to prepare high-purity and high-activity M^{pro} . In this study, four types of SARS-CoV-2 M^{pro} s containing different termini were prepared, and their activities were determined successfully. The results showed that the activity of wild-type (WT) M^{pro} was the highest, and the additional residues at the N-terminus but not at the C-terminus had a major effect on the enzyme activity. To explain this, the alignment of structures of different forms of M^{pro} was determined, and the additional residues at the N-terminus were found to interfere with the formation of the substrate binding pocket. This study confirms the importance of the natural N-terminus to the activity of M^{pro} and suggests that WT-GPH₆ (M^{pro} with eight additional residues at the C-terminus) can be used as a substitute for authentic M^{pro} to screen inhibitors. In short, this study provides a reference for the expression and purification of new coronaviruses confronted in the future.

1. Introduction

In 2019, coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) appeared in Wuhan, China [1,2]. The main clinical manifestations of mild patients were fever, cough, and dyspnea. A few patients had diarrhea, nausea and other gastrointestinal symptoms [3,4]. Severe patients presented with acute respiratory dysfunction syndrome, multiple organ failure, and even death [5]. The worldwide prevalence of SARS-CoV-2 poses a serious threat to human health and life because of its strong infectivity.

Similar to the two coronaviruses that caused severe acute respiratory syndrome (SARS) in 2003 [6] and Middle East respiratory syndrome (MERS) in 2012 [7], SARS-CoV-2 causing COVID-19 belongs to the genus *Betacoronavirus*, including SARS-CoV and MERS-CoV [1,8]. Furthermore, SARS-CoV-2 has 79.6% genome sequence identity with SARS-CoV [9] and nearly 50% genome sequence identity with MERS-CoV [10].

SARS-CoV-2 is a single-stranded RNA virus. Its genome is composed of 14 functional open reading frames (ORFs), which can encode at least 29 kinds of proteins. Four structural proteins are included: spike protein, nucleocapsid protein, membrane glycoprotein, and envelope protein [11,12]. SARS-CoV-2 ORF1a and ORF1b encode two different polyproteins, pp1a and pp1ab. These polyproteins are cleaved to generate

the nonstructural protein nsp1-16 by pivotal enzymes, including papain-like protease (PL^{pro}) and main protease (M^{pro}) [13]. M^{pro} can cleave polyproteins at 11 sites and initially release itself by autocleavage [14,15]. M^{pro} can recognize substrates as long as 10 residues but generally have specific selectivity for four subsites. Interestingly, substrate recognition pockets of M^{pro} from different coronaviruses are highly conserved due to sequence homology [16,17]. There is up to a remarkable 96% sequence identity of M^{pro} derived from SARS-CoV-2 and SARS-CoV [18]. In summary, highly conserved M^{pro} plays a pivotal role in virus life and has no homologous enzyme in humans, so it can be regarded as a key target for screening inhibitors [19–21].

M^{pro} , with an approximate molecular weight of 33 kDa, is demonstrated to have a dimeric structure, and each protomer has three domains. The catalytic active sites and active pockets are situated in domain I and domain II, and domain III mainly participates in the formation of dimer structures. The catalytic activity center is a catalytic dimer composed of cysteine (Cys145) and histidine (His41) [17,22–25].

Although SARS-CoV-2 M^{pro} has been characterized sufficiently, one particular aspect draws our attention. The kinetic parameters of SARS-CoV-2 M^{pro} were obviously inconsistent from different studies, in which k_{cat}/K_m ranged from 370.4 to 28500 $M^{-1} s^{-1}$. This is difficult to explain by the differences in experimental procedures and reagents [20, 23,26–28]. However, studies on SARS-CoV M^{pro} showed that additional residues at the terminus could lead to the depression of enzyme activity

* Corresponding author.

** Corresponding author.

E-mail addresses: zhangchaofeng@tyut.edu.cn (C. Zhang), zhaocheng@tyut.edu.cn (C. Zhao).

Abbreviations

| | |
|---------------------|---|
| M ^{Pro} | main protease |
| SARS-CoV-2 | severe acute respiratory syndrome coronavirus 2 |
| COVID-19 | coronavirus disease 2019 |
| SARS | severe acute respiratory syndrome |
| MERS | Middle East respiratory syndrome |
| ORFs | open reading frames |
| PL ^{Pro} | papain-like protease |
| WT-GPH ₆ | SARS-CoV-2 M ^{Pro} with eight additional residues (GPHHHHHH) at the C-terminus |
| WT | wild-type SARS-CoV-2 M ^{Pro} with authentic termini |
| H ₆ -WT | SARS-CoV-2 M ^{Pro} with twenty-one additional residues (MGSSHHHHHHSSGRENLYFQG) at the N-terminus |
| G-WT | SARS-CoV-2 M ^{Pro} with one additional residue (G) at the N-terminus |
| PSP | PreScission protease |
| TEV | tobacco etch virus |

2. Materials and methods

2.1. Construction of recombinant plasmids

The full-length M^{Pro} gene 10055–10972 of SARS-CoV-2 (GenBank accession no. **NC_045512.2**) was optimized and synthesized for *Escherichia coli* (*E. coli*) expression. Then, the synthesized gene was cloned into the pGEX-6p-1 vector via the restriction sites *Bam*HI and *Xho*I to obtain the recombinant plasmid, which was designated pGEX-M^{Pro} (Genewiz). To obtain the authentic N-terminus, 12 nucleotides and 24 nucleotides corresponding to AVLQ and GPHHHHHH at N- and C-termini, respectively, were added to the M^{Pro} gene, which was performed by PCR via primers P1–P2 (Table 1). Subsequently, the amplified gene product was ligated into the pGEX-6p-1 vector via the restriction sites *Bam*HI and *Xho*I, and the constructed plasmid was designated pGEX-M^{Pro}-GPH₆. Similarly, to obtain the authentic C-terminus, the M^{Pro} gene was amplified with primers P3–P4 (Table 1) and ligated into the modified pET21a vector, which was digested with the restriction enzymes *Bam*HI and *Xho*I. The modified vector contained the His tag and tobacco etch virus (TEV) protease cleavage site upstream of restriction site *Bam*HI [30]. The constructed recombinant plasmid containing the

Table 1
Primers used in this study.

| Primer | Sequence |
|--------|--|
| P1 | TTCCAGGGGCCCTGGATCCGCGTACTGCAGAGCGGCTTTCG |
| P2 | GTCACGATGCGGCCGCTCGAGTTAGTGGTGGTGGTGGTGGGGTCCCTGAAAGGTCA |
| P3 | GAAAACTGTATTCCAGGGCAGCGGCTTTCGCAAAATGG |
| P4 | GTGGTGGTGGTGGTCTCGAGTTACTGAAAGGTACAGCCGC |

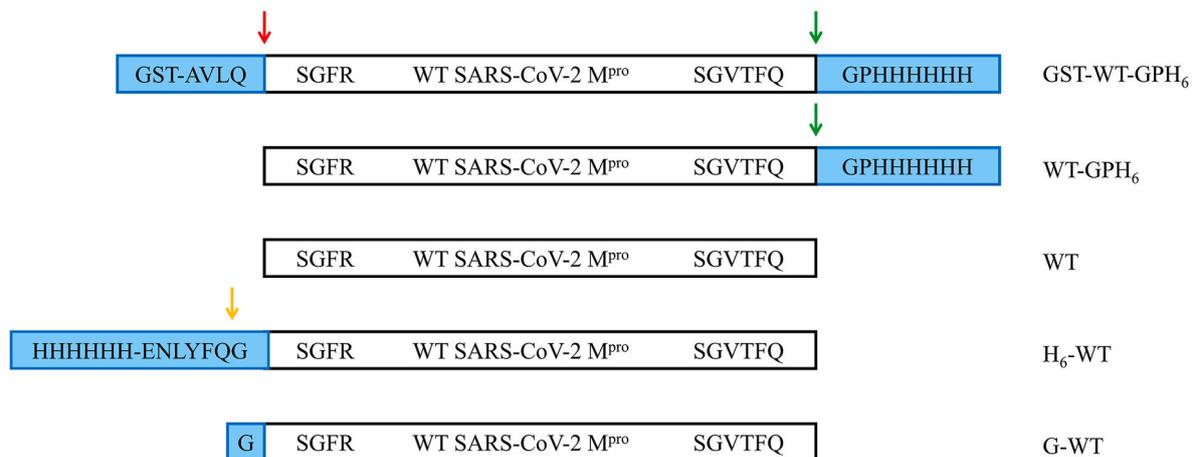


Fig. 1. Schematic diagram of four types of SARS-CoV-2 M^{Pro}s. The additional residues are shown in blue. The colored arrows represent the protease cleavage site. Red, M^{Pro}; green, PSP; yellow, TEV protease.

to varying degrees [29]. Considering the high homology of M^{Pro} from SARS-CoV and SARS-CoV-2, we speculated that additional amino acids would affect the activity of SARS-CoV-2 M^{Pro} analogously.

In this study, we expressed four types of SARS-CoV-2 M^{Pro}s with different termini and measured their activity. Similar to previous studies on SARS-CoV M^{Pro}, additional residues at the N-terminus significantly reduced the activity, but additional residues at the C-terminus had little effect on the activity. These results indicated that M^{Pro} in the WT-GPH₆ form can be used as a substitute for authentic M^{Pro} and provided references for the expression and purification of coronavirus M^{Pro} in the future.

pET21a and M^{Pro} genes was designated pET-H₆-M^{Pro}. The gene sequences of the constructed plasmids were verified by sequencing (Genewiz).

2.2. Expression and purification of four types of SARS-CoV-2 M^{Pro}s

The plasmid pGEX-M^{Pro}-GPH₆ was expressed to obtain M^{Pro} with eight additional residues (GPHHHHHH) at the C-terminus, and the generated M^{Pro} was designated WT-GPH₆. Purified WT-GPH₆ was cleaved with PreScission Protease (PSP) to produce authentic M^{Pro}, which was named WT. Similarly, plasmid pET-H₆-M^{Pro} was expressed to generate M^{Pro} with twenty-one additional residues (MGSSHHHHHHSSGRENLYFQG) at the N-terminus, and the resulting

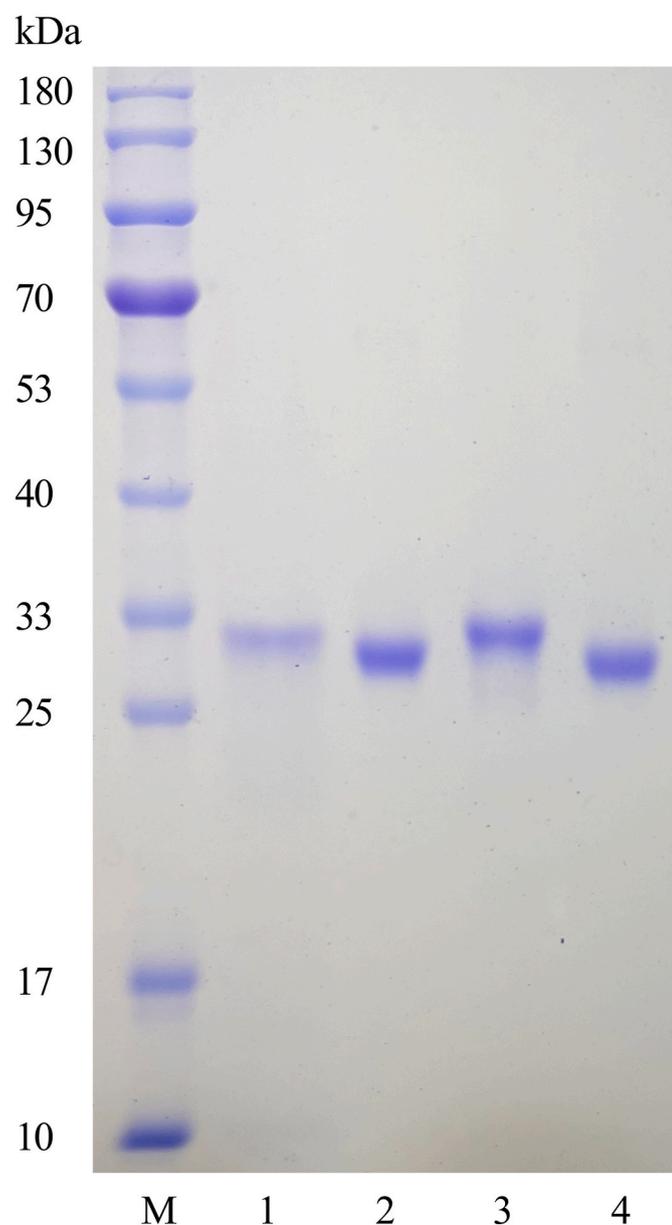


Fig. 2. Expression and purification of M^{Pro} . Lane M, protein molecular mass marker; Lane 1, WT-GPH₆; Lane 2, WT; Lane 3, H₆-WT; Lane 4, G-WT.

M^{Pro} was designated H₆-WT. Subsequently, extra residues of H₆-WT were removed by using TEV protease to obtain M^{Pro} designated G-WT, which included an extra glycine residue at the N-terminus. The expression and purification of four types of SARS-CoV-2 M^{Pro} s (WT-GPH₆, WT, H₆-WT, G-WT) are described in detail below.

For the expression and purification of WT-GPH₆, the plasmid pGEX- M^{Pro} -GPH₆ was transformed into *E. coli* BL21 (DE3). Positive clones harboring the recombinant plasmid were cultured in LB medium containing 100 µg/mL ampicillin at 37 °C. When the OD₆₀₀ reached 0.6–0.8, 0.5 mM isopropyl-D-thiogalactoside (IPTG) was added, and M^{Pro} was overexpressed at 16 °C for 14–16 h. The cells were harvested by centrifugation at 6000 rpm for 5 min, and then pellets were resuspended in buffer A (20 mM Tris-HCl, 150 mM NaCl, pH 7.5) and lysed by sonication on ice. The lysate was centrifuged at 10000 rpm for 60 min, and the precipitate was discarded. The supernatant was loaded onto a HisTrap column, which was subsequently successively washed with buffer A containing imidazole at different concentrations (20 mM, 50 mM, 100 mM, 300 mM, 500 mM, and 1000 mM). The fractions

containing proteins of interest were concentrated and then dialyzed in buffer B (50 mM Tris-HCl, 1 mM EDTA, pH 7.3). The purified WT-GPH₆ was analyzed by SDS-PAGE.

For the purification of WT, PSP was added to purified WT-GPH₆ to cleave additional residues at a mass ratio of 1:30, resulting in the target protein WT with an authentic C-terminus. The digestion mixture was loaded onto a GSTrap column, and the fraction that did not bind to the column was collected to remove PSP. The protein obtained in the previous step was subjected to a HisTrap column, and then the target protein WT was collected by eluting with buffer A to remove WT-GPH₆, which had not been digested with PSP. The purified WT was analyzed by SDS-PAGE.

For the expression and purification of H₆-WT, the plasmid pET-H₆- M^{Pro} was transformed into *E. coli* BL21 (DE3). The following procedure was similar to that used for the expression and purification of WT-GPH₆. The purified H₆-WT was analyzed by SDS-PAGE.

For the purification of G-WT, TEV protease was added to purified H₆-WT to remove the His tag at a mass ratio of 1:30, resulting in the target protein G-WT. A mixture of digestion was applied to a HisTrap column, and then the target protein G-WT was fragmented by washing with buffer A to exclude TEV protease and H₆-WT, which had not been digested with TEV protease. The purified G-WT was analyzed by SDS-PAGE.

2.3. Activity detection of four types of SARS-CoV-2 M^{Pro} s

Enzyme activity assays of four types of M^{Pro} s (WT-GPH₆, WT, H₆-WT, and G-WT) were performed based on fluorescence resonance energy transfer, and the substrate was MCA-AVLQSGFR-Lys(Dnp)-Lys-NH₂ (P9731-5 mg, Beyotime Biotechnology, China) according to previous research [20]. The fluorescence was measured at excitation and emission wavelengths of 320 nm and 395 nm, respectively, and the reaction was carried out in buffer B with a total volume of 500 µL. Initially, 440 µL buffer B and 50 µL M^{Pro} (final concentration of 0.2 µM for WT-GPH₆ and WT, 5 µM for H₆-WT, and 2.5 µM for G-WT) were added to the reaction system. As soon as 10 µL of substrate (final concentrations of 2.5 µM, 3.3 µM, 5 µM, 10 µM, 20 µM, and 40 µM) was added to the reaction system, the relative fluorescence unit (RFU) value was immediately monitored for 1200 s (F2700, Hitachi, Japan). Similarly, the RFU values of a series of different concentrations of MCA (0.625 µM, 1.25 µM, 2.5 µM, 5 µM, 10 µM, and 20 µM, M185662-1 g, Aladdin, China) were measured in the absence of M^{Pro} to calculate the amount of the cleaved substrate. The initial rate depended on the change in RFUs in the first 90 s. The kinetic constants K_m and k_{cat} were obtained from a double reciprocal plot, which was drawn by GraphPad Prism 8.0 software. Experiments were performed in triplicate, and the values were presented as the mean ± standard deviation (SD).

3. Results and discussion

3.1. Construction of the recombinant plasmids

To shed light on the effect of additional residues on the activity of SARS-CoV-2 M^{Pro} , two recombinant plasmids were constructed to produce four types of M^{Pro} s with different termini, which included WT-GPH₆, WT, H₆-WT, and G-WT (Fig. 1). Two types of M^{Pro} s can be obtained when one recombinant plasmid is expressed based on the hydrolysis strategy of protease by introducing a cleavage site. According to a previous study of SARS-CoV M^{Pro} [29], the plasmid pGEX- M^{Pro} -GPH₆ was constructed to produce WT-GPH₆ (with eight additional residues at the C-terminus) and WT (with authentic termini). When constructing pGEX- M^{Pro} -GPH₆, 12 nucleotides encoding AVLQ were added upstream of the M^{Pro} gene to generate the cleavage site of M^{Pro} , which led to the self-cleavage of the recombinant protein GST-WT-GPH₆ to produce WT-GPH₆. Meanwhile, 24 nucleotides encoding GPHHHHHH were added downstream of the M^{Pro} gene to create the cleavage site of PSP,

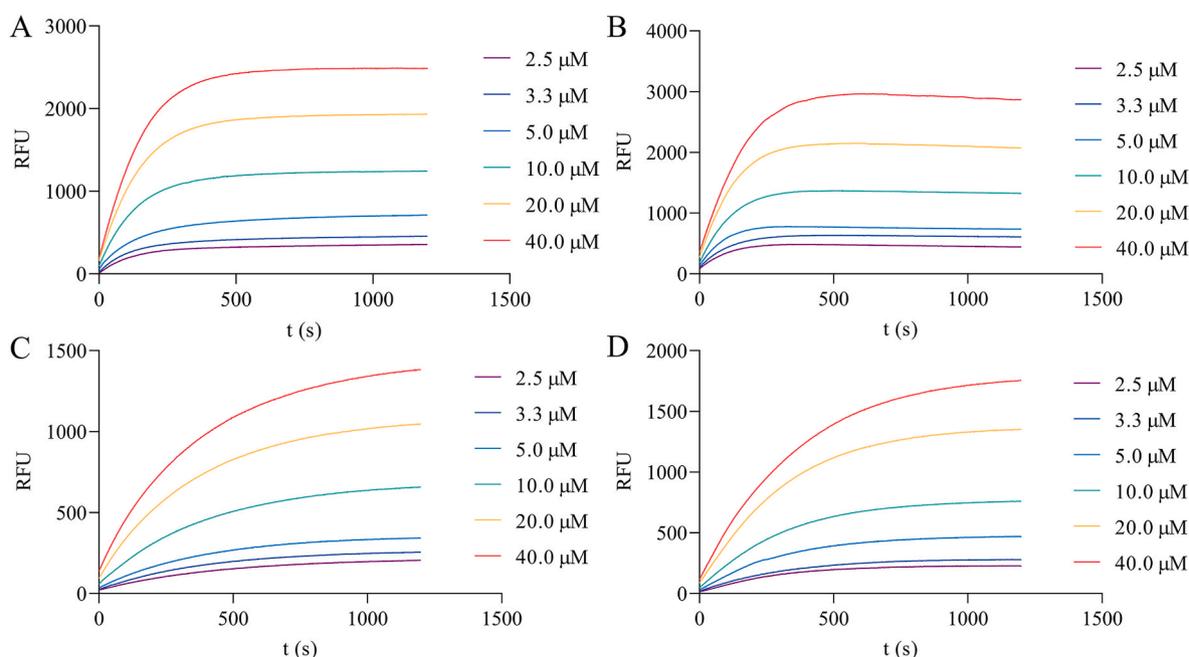


Fig. 3. The activities of four types of M^{pro} s. A, WT-GPH₆; B, WT; C, H₆-WT; D, G-WT.

Table 2
Comparison of enzyme activities of four types of SARS-CoV-2 M^{pro} s.

| SARS-CoV-2 M^{pro} | K_m (μM) | k_{cat} (s^{-1}) | k_{cat}/K_m ($\text{M}^{-1}\text{s}^{-1}$) |
|-----------------------------|-------------------------|--------------------------------------|---|
| H ₆ -WT | 41.81 ± 9.79 | 0.0058 ± 0.0013 | 139 |
| G-WT | 28.01 ± 4.77 | 0.0110 ± 0.0016 | 393 |
| WT-GPH ₆ | 26.48 ± 5.67 | 0.4167 ± 0.0807 | 15736 |
| WT | 13.31 ± 1.33 | 0.3259 ± 0.0277 | 24485 |

which digested WT-GPH₆ to WT. To obtain H₆-WT and G-WT, we constructed the recombinant plasmid pET-H₆- M^{pro} , which contained a His tag sequence and a TEV protease cleavage site upstream of the M^{pro} gene.

3.2. Preparations of WT-GPH₆ and WT

When the plasmid pGEX- M^{pro} -GPH₆ was expressed in *E. coli*, the GST tag-fused protein GST-WT-GPH₆ was initially generated. Since four residues (AVLQ) were introduced into the N-terminus of the first residue serine of M^{pro} , the recognition site of M^{pro} was formed. In this case, GST-

WT-GPH₆ was cleaved by itself to produce WT-GPH₆. This phenomenon was probably similar to the idea that M^{pro} released itself from poly-proteins in viruses [14]. After purification of the expression product by Ni-affinity chromatography, purified protein was analyzed by SDS-PAGE, which showed that a single band (Fig. 2, Lane 1) appeared near 33 kDa. This result was in accord with the predicted molecular mass of 34.77366 kDa, which revealed that SARS-CoV-2 M^{pro} in the WT-GPH₆ form was obtained.

It was necessary to obtain authentic M^{pro} to clarify the influence of additional residues. To remove the redundant residues, the cleavage site of PSP was generated by adding two residues, glycine and proline, at the C-terminus of M^{pro} . In this case, the six amino acids (SGVTFQ) at the C-terminus of M^{pro} and the two added amino acids (GP) corresponded to the recognition sites P6-P1 and P1'-P2' of PSP, respectively. When the mixture of WT-GPH₆ processed by PSP was purified by affinity chromatography, a slightly lower band appeared in Lane 2 than in Lane 1 (Fig. 2). The slight difference between Lane 1 and Lane 2 can be explained by the difference in molecular mass between WT-GPH₆ (34.77366 kDa) and WT (33.79664 kDa), which indicated that the excess eight amino acids at the C-terminus were successfully removed

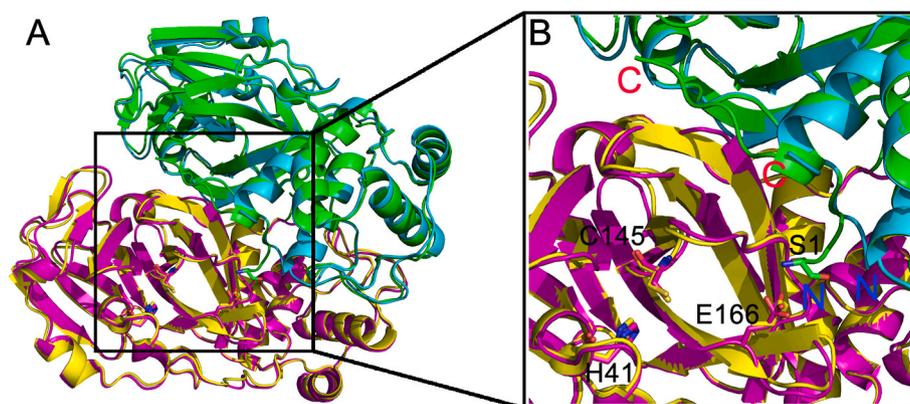


Fig. 4. Superposition of WT and G-WT. A, The overall superposition of WT and G-WT; B, the zoom-in view of the termini. Protomer A and B in the WT form are colored green and yellow, respectively. Protomer A and B in the G-WT are colored cyan and magenta, respectively. The key amino acids are presented by stick and labeled in black. “N” and “C” represent the N-terminus and C-terminus, respectively.

and M^{Pro} in the WT form was obtained.

Although two types of M^{Pro}s (WT-GPH₆ and WT) were obtained in our study, a special circumstance attracted our attention. In recent research, a series of structures of complexes of M^{Pro} binding with natural substrates were resolved. In the structure of the M^{Pro} H41A mutant containing the nsp5|6 substrate (PDB ID: 7VAH), the side chain of alanine at P2' was exposed to solvent, which indicated that the P2' site had less specificity for substrate recognition [14]. To isolate natural M^{Pro}, a His tag and two excess residues of GP were added. After affinity chromatography, WT-GPH₆ was successfully obtained, which indicated that the His tag was rarely removed. Interestingly, the His tag was cleaved by PSP, which revealed that the C-terminal residues (SGVTFQGP) of WT-GPH₆ could be recognized by PSP but not by M^{Pro}. Comparing the sequence (SGVTFQGP) with that of the natural nsp5|6 substrate (SGVTFQSAV), it was found that Ser at the P1' site and Ala at the P2' site were replaced by Gly and Pro, respectively. The S1' subsite was a shallow pocket and could accommodate small residues. Therefore, the reason that the sequence (SGVTFQGP) was not identified by M^{Pro} was because Pro was at the P2' site, not Gly at the P1' site, which was consistent with profiling of the substrate specificity of SARS-CoV M^{Pro} [31]. This might be due to the change in peptide chain structure caused by the special structure of proline, which resulted in the inactivation of the substrate.

3.3. Preparations of H₆-WT and G-WT

In a previous study, when there were two or five residues in the N-terminus of SARS-CoV M^{Pro}, the activity was only 4% or 1% of that of authentic M^{Pro}, respectively [29]. Considering that the sequences of M^{Pro} from SARS-CoV-2 and SARS-CoV are highly homologous and that their structures are highly similar [20], we speculated that the excess residues at the N-terminus would also significantly affect the activity of M^{Pro} derived from SARS-CoV-2. Therefore, M^{Pro} in the H₆-WT form with 21 additional residues (MGSSHHHHHSSGRENLYFQG) at the N-terminus was generated when the plasmid pET-H₆-M^{Pro} was expressed in *E. coli*. Similarly, the expression product was purified by Ni-affinity chromatography, and purified protein was detected by SDS-PAGE. The result showed that there was a single band (Fig. 2, Lane 3) at the upper position of the band at Lane 2, which was in accord with the theoretical molecular mass of H₆-M^{Pro} (36.22120 kDa).

H₆-M^{Pro} is a fusion protein composed of M^{Pro}, the N-terminal His tag and the TEV protease cleavage site. TEV protease can recognize the sequence ENLYFQG and cleave between glutamine and glycine. Therefore, when H₆-M^{Pro} was digested by TEV protease, one glycine was left on M^{Pro}. As expected, after purification by affinity chromatography, it was observed that the band at Lane 4 was slightly lower than that at Lane 3 (Fig. 2). The small gap of the bands may exactly correspond to the molecular mass (2.36751 kDa) of the cleaved residues, which demonstrated that the His tag at the N-terminus was successfully released and M^{Pro} in the G-WT form was obtained.

3.4. Activity determination of four types of M^{Pro}s

To ascertain whether the additional residues would affect the catalytic activity of SARS-CoV-2 M^{Pro}, the activities of four types of M^{Pro}s (WT-GPH₆, WT, H₆-WT, G-WT) were determined by measuring the kinetic parameters. When four types of M^{Pro}s reacted with substrates of different concentrations, the changes in RFUs over 1200 s were continuously monitored (Fig. 3). The data showed that the RFU value increased rapidly and that the reaction conformed to the first-order reaction in the first 90 s. This phenomenon indicated that M^{Pro} could recognize and cleave the fluorescent substrates, which suggested that the four types of M^{Pro}s were indeed active.

To calculate the kinetic parameters of M^{Pro}s, double reciprocal plots were performed according to the initial rate and substrate concentration of the reaction (Table 2, Fig. S1, S2). The catalytic efficiency of an

enzyme is best defined by k_{cat}/K_m [32]. Obviously, WT with a k_{cat}/K_m of $24485 \text{ M}^{-1}\text{s}^{-1}$ had the highest catalytic efficiency, while H₆-WT with a k_{cat}/K_m of $139 \text{ M}^{-1}\text{s}^{-1}$ had the lowest catalytic efficiency. The activity of WT was more than 170 times greater than that of H₆-WT and 60 times greater than that of G-WT. The results showed that additional residues at the N-terminus significantly led to a decrease in M^{Pro} activity. Moreover, the greater the number of additional residues at the N-terminus, the greater the decrease in M^{Pro} activity. Fortunately, the activity of WT-GPH₆ ($k_{cat}/K_m = 15736 \text{ M}^{-1}\text{s}^{-1}$) was approximately two-thirds of that of WT, which suggested that additional residues at the C-terminus had little effect on M^{Pro} activity.

In a previous study, SARS-CoV M^{Pro} in the WT form had the highest cleavage efficiency compared to that in other forms, and additional residues at the N-terminus, but not at the C-terminus, were detrimental to enzyme activity [29]. This was consistent with our results, likely because the M^{Pro}s from the two viruses were not only highly homologous in sequence but also highly consistent in structure. This confirmed that M^{Pro} could be used as a screening target for broad-spectrum inhibitors.

To date, a series of structures of SARS-CoV-2 M^{Pro} in different forms have been resolved. To determine how the additional residues affected the activity of M^{Pro}, we compared the structures of M^{Pro} in the WT form (PDB ID: 6M03) and G-WT form (PDB ID: 7BRO) (Fig. 4). In the WT structure, the first residue serine interacted with Glu166, stabilizing the conformation of the substrate binding pocket. In contrast, the N-terminus of G-WT slightly deviated from the orientation to the active site due to excess glycine [33]. Similarly, the N-terminus could not participate in the formation of a substrate binding pocket for SARS-CoV M^{Pro} in the GPGLS-WT form with five additional residues at the N-terminus, which likely resulted in a reduction in enzyme activity [29]. The structural difference would account for the higher activity of the WT protease compared with other proteases. Furthermore, our results also showed that the activity of H₆-WT was lower than that of G-WT. This might be due to the additional residues at the N-terminus, which were adjacent to the substrate binding pocket and would interfere with substrate recognition. In summary, our study once again emphasized the importance of authentic N-terminus for the activity of M^{Pro}.

In contrast to the N-terminus, the C-terminus of M^{Pro} in the WT form was oriented away from the substrate binding pocket. Therefore, the C-terminal extra residues had less effect on enzyme activity, which indicated that WT-GPH₆ can be used as a substitute for authentic M^{Pro} to screen inhibitors targeting the active site.

4. Conclusions

In this study, we described an ingenious strategy for producing four types of SARS-CoV-2 M^{Pro}s by expressing the two types of recombinant plasmids in *E. coli*. Activity assays showed that the WT had the highest cleavage efficiency, and the effect of the additional residues at the N-terminus on the enzyme activity was much greater than that at the C-terminus. This study confirms the importance of the natural N-terminus of M^{Pro} and provides a reference for the expression of M^{Pro} of the new coronavirus in the future.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Fundamental Research Program of Shanxi Province, China (grant 20210302124076, 202103021224067, 202103021224070) and the Shanxi Scholarship Council of China, China (grant 2021-036).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.pep.2022.106208>.

References

- [1] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature* 579 (2020) 265–269, <https://doi.org/10.1038/s41586-020-2008-3>.
- [2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, Y. Tan, Y.H. China Novel Coronavirus, T. Research, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (2020) 727–733, <https://doi.org/10.1056/NEJMoa2001017>.
- [3] F. Jiang, L. Deng, L. Zhang, Y. Cai, C.W. Cheung, Z. Xia, Review of the clinical characteristics of coronavirus disease 2019 (COVID-19), *J. Gen. Intern. Med.* 35 (2020) 1545–1549, <https://doi.org/10.1007/s11606-020-05762-w>.
- [4] W.J. Guan, Z.Y. Ni, Y. Hu, W.H. Liang, C.Q. Ou, J.X. He, L. Liu, H. Shan, C.L. Lei, D. S.C. Hui, B. Du, L.J. Li, G. Zeng, K.Y. Yuen, R.C. Chen, C.L. Tang, T. Wang, P. Y. Chen, J. Xiang, S.Y. Li, J.L. Wang, Z.J. Liang, Y.X. Peng, L. Wei, Y. Liu, Y.H. Hu, P. Peng, J.M. Wang, J.Y. Liu, Z. Chen, G. Li, Z.J. Zheng, S.Q. Qiu, J. Luo, C.J. Ye, S. Y. Zhu, N.S. Zhong, C. China Medical Treatment, Expert group for, clinical characteristics of coronavirus disease 2019 in China, *N. Engl. J. Med.* 382 (2020) 1708–1720, <https://doi.org/10.1056/NEJMoa2002032>.
- [5] Y. Du, L. Tu, P. Zhu, M. Mu, R. Wang, P. Yang, X. Wang, C. Hu, R. Ping, P. Hu, T. Li, F. Cao, C. Chang, Q. Hu, Y. Jin, G. Xu, Clinical features of 85 fatal cases of COVID-19 from Wuhan. A retrospective observational study, *Am. J. Respir. Crit. Care Med.* 201 (2020) 1372–1379, <https://doi.org/10.1164/rccm.202003-0543OC>.
- [6] C. Drosten, S. Günther, W. Preiser, S. Van Der Werf, H.-R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R.A. Fouchier, Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *N. Engl. J. Med.* 348 (2003) 1967–1976, <https://doi.org/10.1056/nejm03030747>.
- [7] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D. Osterhaus, R.A. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, *N. Engl. J. Med.* 367 (2012) 1814–1820, <https://doi.org/10.1056/NEJMoa1211721>.
- [8] J. Cui, F. Li, Z.L. Shi, Origin and evolution of pathogenic coronaviruses, *Nat. Rev. Microbiol.* 17 (2019) 181–192, <https://doi.org/10.1038/s41579-018-0118-9>.
- [9] P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang, H.R. Si, Y. Zhu, B. Li, C. L. Huang, H.D. Chen, J. Chen, Y. Luo, H. Guo, R.D. Jiang, M.Q. Liu, Y. Chen, X. R. Shen, X. Wang, X.S. Zheng, K. Zhao, Q.J. Chen, F. Deng, L.L. Liu, B. Yan, F. X. Zhan, Y.Y. Wang, G.F. Xiao, Z.L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin, *Nature* 579 (2020) 270–273, <https://doi.org/10.1038/s41586-020-2012-7>.
- [10] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W.J. Liu, D. Wang, W. Xu, E. C. Holmes, G.F. Gao, G. Wu, W. Chen, W. Shi, W. Tan, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet* 395 (2020) 565–574, [https://doi.org/10.1016/s0140-6736\(20\)30251-8](https://doi.org/10.1016/s0140-6736(20)30251-8).
- [11] T.M. Karpinski, M. Ozarowski, A. Seremak-Mrozikiewicz, H. Wolski, D. Wlodkovic, The 2020 race towards SARS-CoV-2 specific vaccines, *Theranostics* 11 (2021) 1690–1702, <https://doi.org/10.7150/thno.53691>.
- [12] H. Yao, Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, Z. Wu, L. Cheng, D. Shi, X. Lu, J. Lei, M. Crispin, Y. Shi, L. Li, S. Li, Molecular architecture of the SARS-CoV-2 virus, *Cell* 183 (2020) 730–738, <https://doi.org/10.1016/j.cell.2020.09.018>, e713.
- [13] C. Bai, Q. Zhong, G.F. Gao, Overview of SARS-CoV-2 genome-encoded proteins, *Sci. China Life Sci.* 65 (2022) 280–294, <https://doi.org/10.1007/s11427-021-1964-4>.
- [14] Y. Zhao, Y. Zhu, X. Liu, Z. Jin, Y. Duan, Q. Zhang, C. Wu, L. Feng, X. Du, J. Zhao, M. Shao, B. Zhang, X. Yang, L. Wu, X. Ji, L.W. Guddat, K. Yang, Z. Rao, H. Yang, Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2, *Proc. Natl. Acad. Sci. U. S. A.* 119 (2022), e2117142119, <https://doi.org/10.1073/pnas.2117142119>.
- [15] A. Hegyi, J. Ziebuhr, Conservation of substrate specificities among coronavirus main proteases, *J. Gen. Virol.* 83 (2002) 595–599, <https://doi.org/10.1099/0022-1317-83-3-595>.
- [16] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of wide-spectrum inhibitors targeting coronavirus main proteases, *PLoS Biol.* 3 (2005) e324, <https://doi.org/10.1371/journal.pbio.0030324>.
- [17] K. Anand, G.J. Palm, J.R. Mesters, S. Siddell, Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain, *EMBO J.* 21 (2002) 3213–3224, <https://doi.org/10.1093/emboj/cdf327>.
- [18] J.S. Morse, T. Lalonde, S. Xu, W. Liu, Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV, *ChemBiochem* 21 (2020) 730–738, <https://doi.org/10.26434/chemrxiv.11728983>.
- [19] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.H. Jung, An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy, *J. Med. Chem.* 59 (2016) 6595–6628, <https://doi.org/10.1021/acs.jmedchem.5b01461>.
- [20] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors, *Nature* 582 (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [21] R. Arya, S. Kumari, B. Pandey, H. Mistry, S.C. Bihani, A. Das, V. Prashar, G. D. Gupta, L. Panicker, M. Kumar, Structural insights into SARS-CoV-2 proteins, *J. Mol. Biol.* 433 (2021), 166725, <https://doi.org/10.1016/j.jmb.2020.11.024>.
- [22] W. Cui, K. Yang, H. Yang, Recent progress in the drug development targeting SARS-CoV-2 main protease as treatment for COVID-19, *Front. Mol. Biosci.* 7 (2020), 616341, <https://doi.org/10.3389/fmolb.2020.616341>.
- [23] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors, *Science* 368 (2020) 409–412, <https://doi.org/10.1126/science.abb3405>.
- [24] L. Lu, S. Su, H. Yang, S. Jiang, Antivirals with common targets against highly pathogenic viruses, *Cell* 184 (2021) 1604–1620, <https://doi.org/10.1016/j.cell.2021.02.013>.
- [25] M. Xiong, H. Su, W. Zhao, H. Xie, Q. Shao, Y. Xu, What coronavirus 3C-like protease tells us: from structure, substrate selectivity, to inhibitor design, *Med. Res. Rev.* 41 (2021) 1965–1998, <https://doi.org/10.1002/med.21783>.
- [26] T.T.H. Nguyen, J.H. Jung, M.K. Kim, S. Lim, J.M. Choi, B. Chung, D.W. Kim, D. Kim, The inhibitory effects of plant derivate polyphenols on the main protease of SARS coronavirus 2 and their structure-activity relationship, *Molecules* (2021) 26, <https://doi.org/10.3390/molecules26071924>.
- [27] T. Drazic, N. Kuhl, M.M. Leuthold, M.A.M. Behnam, C.D. Klein, Efficiency improvements and discovery of new substrates for a SARS-CoV-2 main protease FRET assay, *SLAS Discov* 26 (2021) 1189–1199, <https://doi.org/10.1177/24725552211020681>.
- [28] J.C. Ferreira, S. Fadl, W.M. Rabeh, Key dimer interface residues impact the catalytic activity of 3CLpro, the main protease of SARS-CoV-2, *J. Biol. Chem.* 298 (2022), 102023, <https://doi.org/10.1016/j.jbc.2022.102023>.
- [29] X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, Z. Rao, Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction, *J. Mol. Biol.* 366 (2007) 965–975, <https://doi.org/10.1016/j.jmb.2006.11.073>.
- [30] R. Peng, Z. Li, Y. Xu, S. He, Q. Peng, L.A. Wu, Y. Wu, J. Qi, P. Wang, Y. Shi, G. F. Gao, Structural insight into multistage inhibition of CRISPR-Cas12a by AcrVA4, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 18928–18936, <https://doi.org/10.1073/pnas.1909400116>.
- [31] C.P. Chuck, L.T. Chong, C. Chen, H.F. Chow, D.C. Wan, K.B. Wong, Profiling of substrate specificity of SARS-CoV 3CL, *PLoS One* 5 (2010), e13197, <https://doi.org/10.1371/journal.pone.0013197>.
- [32] R.A. Copeland, *Enzymes, A Practical Introduction to Structure, Mechanism, and Data Analysis*, second ed., Wiley VCH, New York, 2000.
- [33] L. Fu, F. Ye, Y. Feng, F. Yu, Q. Wang, Y. Wu, C. Zhao, H. Sun, B. Huang, P. Niu, H. Song, Y. Shi, X. Li, W. Tan, J. Qi, G.F. Gao, Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease, *Nat. Commun.* 11 (2020) 4417, <https://doi.org/10.1038/s41467-020-18233-x>.