Engineering a novel endopeptidase based on SARS 3CL\textsuperscript{pro}

Chih-Jung Kuo,\textsuperscript{1,2} Yan-Ping Shih,\textsuperscript{1} Daphne Kan,\textsuperscript{1} and Po-Huang Liang\textsuperscript{1,2}

\textsuperscript{1}Institute of Biological Chemistry, Academia Sinica and Core Facility of Recombinant Protein Production, Taipei, Taiwan and \textsuperscript{2}Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan

Short Technical Report

**Introduction**

Severe acute respiratory syndrome–coronavirus (SARS-CoV) caused an outbreak in 2003 that killed approximately 800 patients worldwide (1). A 3C-like protease from the virus, 3CL\textsuperscript{pro}, is required for viral replication, cleaving the polyproteins at 11 sites with the conserved Gln↓(Ser, Ala, Gly) sequences. In this study, we developed a mutant 3CL\textsuperscript{pro} (T25G) with an expanded S1′ space that demonstrates 43.5-fold better $k_{cat}/K_m$ compared with wild-type in cleaving substrates with a larger Met at P1′ and is suitable for tag removal from recombinant fusion proteins. Two vectors for expressing fusion proteins with the T25G recognition site (Ala-Val-Leu-Gln↓Met) in *Escherichia coli* and yeast were constructed. Identical recognition sites were used in these vectors for parallel cloning. *Pst*I was chosen as a 5′ cloning site because it overlapped the nucleotide sequence encoding the protease site and avoided addition of extra amino acids at the N terminus of recombinant proteins. 3CL\textsuperscript{pro} (T25G) was found to have a 3-fold improvement over TEV\textsuperscript{pro} in tag cleavage at each respective preferred cleavage site.

**Materials and methods**

Expression and purification of mutant 3CL\textsuperscript{pro}

Expression and purification of wild-type and mutant SARS 3CL\textsuperscript{pro} in *Escherichia coli* was accomplished according to reported procedures (12). T25G and T25S mutants were prepared from the wild-type by using the QuickChange site-directed mutagenesis kit (Cat. no. 200518; Stratagene, La Jolla, CA, USA). C-terminally His-tagged T25G was expressed using pET16b vector (Cat. no. 69662; Novagen, Darmstadt, Germany).

Construction of the expression vectors for producing tag-cleavable fusion proteins in *E. coli* and yeast

The UPPs-encoding gene (13) was employed as a template for PCR using primers containing the nucleotides encoding the T25G 3CL\textsuperscript{pro} recognition site AVLQ, and the TEV\textsuperscript{pro} recognition site EDLYFQ, respectively. The PCR products were purified from an agarose gel following electrophoresis and cloned into the pET32Xa/Lic vector (Novagen). To serve as a control, the UPPs fusion protein with AAAQ instead of AVLQ was also expressed.

For expressing EGFP fusion proteins in yeast, primers were used to generate a PCR product that was ligated into pHTPY7, which was modified from pPICZαA.

**Keywords:** SARS-CoV; 3CL protease; endopeptidase; tag cleavage; parallel cloning

---

**Figure 1. Structural basis for T25 mutation.** Predicted structural model of SARS-CoV 3CL\textsuperscript{pro} with a modified peptide containing Met at P1′ (Thr-Ser-Ala-Val-Leu-Gln-Met-Phe-Arg-Lys), based on the crystal structure of SARS-CoV 3CL\textsuperscript{pro} H41A mutant in complex with a peptide (PDB entry: 2Q6G).
(Invitrogen) by incorporating nucleotides encoding a starch binding domain (SBD) (14) and AVLO cleavage site.

Evaluation of tag removal by the proteases
The UPPs and EGFP fusion proteins were purified using NiNTA columns. To examine the tag cleavage reactions, the purified fusion proteins (5.4 μM each) were treated with 0.1 μM wild-type and two mutant (T25G and T25S) 3CLpro for 90 min at 37°C. For time course measurements, the fusion proteins (5.4 μM each) were treated with T25G (0.1 μM) at 37°C. The reactions were stopped by 2% trifluoroacetic acid after appropriate time periods and analyzed by SDS-PAGE. For comparing the tag cleavage efficiency of T25G and TEVpro (Invitrogen), the fusion proteins (Tag-AVLO-UPPS and Tags-ENLYFO-UPPS, 5.4 μM each) were treated with 0.1 μM T25G and TEVpro at 37°C, respectively and then analyzed by SDS-PAGE.

Substrate specificity and kinetic parameters of the mutant SARS 3CLpro
The peptides used as substrates for the T25G protease were synthesized via solid phase, using a 433A peptide synthesizer (Applied Biosystems, Foster City, CA, USA). Each peptide (100 μM) was incubated with 0.1 μM T25G for 1, 2, and 6 h, and the subsequent mixtures were analyzed by HPLC on a C-18 reverse-phase analytic column. Cleavage products were resolved using a 30-min, 2–90% linear gradient of acetonitrile plus 0.1% TFA. The product peak areas were integrated to calculate the reaction rates for each peptide substrate. For $K_m$ and $k_{cat}$ measurements, 0.1 μM T25G and 10–200 μM SAVLQMGFRK substrate were used, and the plot of initial rates within 10% substrate consumption versus different substrate concentrations was fitted to the Michaeli-Menten equation using the KaleidaGraph computer program (Synergy Software, Reading, PA, USA).

Results and discussion
Design, preparation, and characterization of T25G and T25S 3CLpro
Based on the crystal structure of SARS-CoV 3CLpro in complex with a peptide (Protein Data Bank entry 2Q6G; www.rcsb.org/pdb), we have generated a structural model of the protease binding with a modified peptide (Thr-Ser-Ala-Val-Leu-Gln-Met*-Phe-Arg-Lys), where the Ser at P1' was changed to Met (indicated by the asterisk). We found that Thr25/Cys2 of the 3CLpro is within a short distance of 1.32 Å of Met/SD of the peptide (Figure 1) as determined by the COOT program (15). Thus, Thr25 may be replaced by a smaller Gly or Ser (maintaining an -OH group) for better accommodation of Met at P1'.

To test the above hypothesis, T25G and T25S 3CLpro were expressed in E. coli and purified using NiNTA chromatography. Yields were approximately 20 mg/L medium, which were similar to yields from wild-type preparations. T25G, with a C-terminal His-tag, was also prepared for removing the protease using NiNTA after tag cleavage. The enzymatic activities of these protease forms were measured using the fluorogenic substrate Dabcyl-KTSAVLQMGFRKME-Edans, as described previously (12). Compared with the activity of wild-type, no significant difference was observed for T25G and C-terminal His-tagged T25S, but T25S showed almost complete loss of activity (Figure 2A, upper panel). However, for the peptide substrate SAVLQMGFRK containing Met at P1', T25S showed significantly higher specific activity than the wild-type (83.5 μM/min versus 6.8 μM/min) (Figure 2A, lower panel), indicating that T25G can tolerate the larger residue Met at P1'. In comparison with the $k_{cat}$ of 1.6 ± 0.2
min$^{-1}$ and the $K_m$ of 76.6 ± 3.5 μM for the wild-type, the T25G mutant displayed the $k_{cat}$ of 16.2 ± 0.5 min$^{-1}$ and the $K_m$ of 18.6 ± 2.4 μM (43.5-fold higher $k_{cat}/K_m$) against the SAVLQ-4MGFRK substrate.

Substrate specificity and kinetic parameters of the mutant SARS 3CL$^{pro}$

Next, peptides corresponding to the N-terminal maturation site of SARS 3CL$^{pro}$ with 10 selected variations (Glu, Phe, Gly, His, Lys, Leu, Met, Pro, Ser, and Trp) at P1′ were prepared and used to evaluate the substrate specificity of T25G. As shown in Figure 2B, T25G showed a 12-fold and 8-fold higher activity against the substrates with Met and Leu at P1′, respectively. For the optimal substrate SAVLQ-4MGFRK of the wild-type, T25G mutant showed about equal activity, indicating that T25G still holds the P1′ residue of the small side chain. Similar to the wild-type, T25G did not tolerate peptides with bulky amino acids such as Trp and Phe or charged amino acids such as Glu and Lys. The peptide with P1′-Pro showed no activity.

Construction of E. coli and yeast vectors to express tag-cleavable fusion proteins by T25G 3CL$^{pro}$

We constructed two vectors for use with E. coli and yeast to express fusion proteins with an AVLQ recognition site to test tag cleavage by T25G. The Pro site CTGAG, which is part of the AVLQ-encoding sequence CGCGGTGCTGACAG, was used as a 5′ cloning site, in conjunction with the 3′ XhoI site, for sticky-end ligation with the PCR product of the target gene (see the strategy illustrated in Figure 3A). As shown in Figure 3B, the purified 5.4 μM fusion UPPs (lane 1) was incubated with 0.1 μM wild-type and two mutant 3CL$^{pro}$ (T25G and T25S), and the final products are shown in lanes 2, 3, and 4, respectively. Only T25G mutant efficiently cleaved the fusion protein, yielding tag-free UPPs (28.3 kDa) and the tags (17.6 kDa) as shown in lane 3. Under these conditions, the tag cleavage reaction was completed <90 min (data not shown). As a control experiment, a UPP's fusion protein with an AAAQ sequence was incubated with the protease, but was not cleaved (data not shown), indicating that T25G specifically recognized the AVLQ cleavage site.

Using a yeast Pichia expression system, the EGFP fusion protein with SBD, His-tag, and AVLQ site was overexpressed. SBD was included for the purpose of using starch as an affinity matrix for protein purification, which would lower the associated costs. With 0.1 μM T25G, the cleavage of the fusion protein (5.4 μM) was completed <120 min (data not shown).

Comparison of tag cleavage using TEV protease and T25G 3CL$^{pro}$

Since TEV$^{pro}$ is one of the most commonly used endopeptidases for tag cleavage and shares a similar substrate specificity with SARS 3CL$^{pro}$, we compared the efficiency of tag cleavage using T25G 3CL$^{pro}$ to that of TEV$^{pro}$ against the E. coli-expressed UPPs fusion proteins containing their preferred recognition sites. Compared with TEV$^{pro}$, T25G showed a 3-fold higher cleavage rate (0.106 μM/min versus 0.035 μM/min; data not shown). However, compared with that of FXa, another commonly used endopeptidase, the cleavage rate of T25G was 1.7 times lower (0.106 μM/min versus 0.178 μM/min; data not shown). TEV$^{pro}$ generally accepts any amino acid at P1′ except Pro (16). However, besides Met, T25G prefers small residues that are actually very common N-terminal residues of “native proteins” due to the post-translational action of Met amino peptidase. This suggests a great advantage of using T25G as a novel endopeptidase for tag removal. These engineered vectors and T25G can be assembled as a kit for the maximal production of soluble and functional proteins with authentic sequences.

Acknowledgments

The authors thank the National Science Council’s NRPGM project for financial support.

Competing interests

The authors declare no competing interests.

References

1. Liang, P.H. 2006. Characterization and inhibition of SARS-Coronavirus main protease. Curr. Top. Med. Chem. 6:361-376.
2. Ruan, Y.J., C.L. Wei, A.L. Ee, V.B. Vega, H. Thoreau, S.T. Su, J.M. Chia, P. Ng, et al. 2003. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet 361:1779-1785.
3. Anand, K., J. Ziebuhr, P. Wadhwani, J.R. Mesters, and R. Hilgenfeld. 2003. Coronavirus main proteinase (3CL$^{pro}$) structure: basis for design of anti-SARS drugs. Science 300:1763-1767.
4. Long, A.C., D.C. Orr, J.M. Cameron, B.M. Dunn, and J. Kay. 1989. A consensus sequence for substrate hydrolysis by rhinovirus 3C proteinase. FEBS Lett. 258:75-78.
5. Thiel, V., K.A. Ivanov, A. Putes, T. Hertzig, B. Schelle, S. Bayer, B. Weisbuch, E.J. Snijder, et al. 2003. Mechanism and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. 84:2305-2315.
6. Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, et al. 2003. The crystal structure of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. USA 100:13190-13195.
7. Hsu, M.F., C.J. Kuo, K.T. Chang, H.C. Chang, C.C. Chou, T.P. Ko, H.L. Shyr, G.G. Chang, et al. 2005. Mechanism of the maturation process of SARS-CoV 3CL protease. J. Biol. Chem. 280:31257-31266.
8. Xue, X., H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, et al. 2007. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J. Mol. Biol. 366:965-976.
9. Zeng, G. 1998. Sticky-end PCR: new method for subcloning. Biotechniques 25:206-208.
10. Shih, Y.P., W.M. Kung, J.C. Chen, C.H. Yeh, A.H. Wang, and T.F. Wang. 2002. High-throughput screening of soluble recombinant proteins. Protein Sci. 11:1714-1719.
11. Wang, H.M., Y.P. Shih, S.M. Hu, W.T. Lo, H.M. Lin, S.S. Ding, H.C. Liao, and P.H. Liang. Parallel gene cloning and protein production in multiple expression systems. Biotechnol. Prog. [In press].
12. Kuo, C.J., Y.H. Chi, J.T. Hsu, and P.H. Liang. 2004. Characterization of SARS main protease and inhibitor assay using a fluorelogenic substrate. Biochem. Biophys. Res. Commun. 318:862-867.
13. Pan, J.J., S.T. Chiou, and P.H. Liang. 2000. Product distribution and pre-steady-state kinetic analysis of Escherichia coli undecaprenyl pyrophosphate synthase reaction. Biochemistry 39:10936-10942.
14. Tung, J.Y., M.D. Chang, W.I. Chou, Y.Y. Liu, Y.H. Yeh, F.Y. Chang, S.C. Lin, Z.L. Qiu, and Y.J. Sun. 2008. Crystal structures of the starch-binding domain from Rhizopus oryzae glucoamylase reveal a polysaccharide-binding path. Biochem. J. 416:27-36.
15. Emsley, P. and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60:2126-2132.
16. Kapust, R.B., J. Tózsér, T.D. Copeland, and D.S. Waugh. 2002. The P1′ specificity of tobacco etch virus protease. Biochem. Biophys. Res. Commun. 294:949-955.

Received 5 August 2009; accepted 13 October 2009.

Address correspondence to Po-Huang Liang, Institute of Biological Chemistry, Academia Sinica and Core Facility of Recombinant Protein Production, Taipei 11529, Taiwan. email: phliang@gate.sinica.edu.tw