Profiling of substrate-specificity and rational design of broad-spectrum peptidomimetic inhibitors for main proteases of coronaviruses

CP Chuck, ZH Ke, C Chen, DCC Wan, HF Chow, KB Wong *

KEY MESSAGES

1. Substrate specificities of the main proteases from group 1, 2a, 2b, and 3 coronaviruses were comprehensively profiled at P5 to P3' positions.
2. Despite subtle differences in substrate specificities identified at P1, P2, and P4 positions, the main proteases from different strains of coronaviruses share many similarities, suggesting the feasibility of developing a broad-spectrum inhibitor.
3. `Super-active' substrates with >4-fold increases in activity were created by combining multiple favourable substitutions.
4. Cbz-AVLQ-CN is a broad-spectrum inhibitor effective against six different strains of coronaviruses (HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV-HKU1, SARS-CoV, IBV) with IC_{50} values of 1.3 to 4.6 μM.

Hong Kong Med J 2014;20(Suppl 4):S22-5
RFCID project number: 09080282

[1] CP Chuck, [2] ZH Ke, [2] C Chen, [3] DCC Wan, [2] HF Chow, [1] KB Wong *

The Chinese University of Hong Kong:
[1] *School of Life Sciences*
[2] *Department of Chemistry*
[3] *School of Biomedical Sciences*

* Principal applicant and corresponding author: kbwong@cuhk.edu.hk

Introduction

Coronavirus (CoV) infection causes a number of respiratory tract diseases in human.[1] The most infamous was the outbreak of severe acute respiratory syndrome (SARS) in 2003. Coronaviruses can be classified into group 1, 2a, 2b, and 3 based on sequence analysis.[2] Most of the CoVs that infect humans belong to group 1 (eg HCoV-NL63, HCoV-229E), 2a (eg HCoV-OC43, HCoV-HKU1), and 2b (SARS-CoV). Because of the high frequency of mutations and the possibility of animal-to-human transmission, CoVs remain a potential threat to public health. Currently, there is no approved drug to combat CoV infection. The development of broad-spectrum inhibitors that can target all strains of CoV is preferable as first-line defence against CoV infection.

The main protease (M^{pro}) is an attractive target for such broad-spectrum inhibitors. The M^{pro} is responsible for proteolytic processing of polyproteins 1a and 1ab, which release at least 15 non-structural proteins that are essential for viral replication. The M^{pro} is a cysteine protease, which uses an invariant cysteine residue to attack the scissile peptide bond. Although the substrate specificities of SARS-CoV Mpro have been extensively studied,[3] there is no comprehensive study on substrate specificities on M^{pro} from other CoVs. We therefore profiled the substrate specificities of M^{pro} from group 1 (HCoV-NL63), 2a (HCoV-OC43), 2b (SARS-CoV), and 3 (IBV) using a 19x8 substrate library.[4] Guided by substrate specificities obtained, we created 'super-active' substrates by combining multiple favourable substitutions. Finally, we synthesised peptidomimetic inhibitors based on the nitrile warhead, and demonstrated broad-spectrum inhibition of six different strains of CoVs.

Methods

Profiling of substrate specificity

This study was conducted from January 2010 to December 2011. Cloning, expression, and purification of M^{pro} from HCoV-NL63, HCoV-OC43, SARS-CoV were performed as described.[3] The purification of the19x8 substrate library, and the assay of protease activity of different Mpro against the substrate library were as described.[3]

Synthesis and assay of inhibitors

Peptidomimetic inhibitors were synthesised by coupling a nitrile warhead to the C-terminus of a peptide using the mixed anhydride method, and were protected at the N-terminus by a carboxybenzyl (Cbz) group. The protease activity in the presence of 0.5 to 256 μM of inhibitors was measured using the FRET assay.[3] The IC_{50} values were obtained by fitting the protease activity to a four-parameter logistics curve.

Structure determination of M^{pro}-inhibitor complex

SARS-CoV M^{pro} without the inhibitor was crystallised in 50mM-(N-morpholino)ethane-sulfonic acid, pH 5.5, 8.5% (w/v) of polyethylene glycol 6000, 10% (v/v) glycerol, 3% (v/v) DMSO, 1 mM ethylenediaminetetraacetic acid and 1 mM dithiothreitol at 16°C using the hanging-drop-vapour-diffusion method. Inhibitor (600 μM) was added to 5 ml of mother liquor containing single crystals of SARS-CoV M^{pro}, and was incubated overnight at 16°C. The crystals were cryoprotected by 20% (v/v) glycerol, and diffraction data were collected in an in-house Rigaku FRE+ X-ray source, and were processed by the CCP4 program suite. The structure of the M^{pro}-inhibitor complex was solved by molecular replacement, built interactively by the program COOT, and refined by the program PHENIX.

Results

Substrate-specificities of M^{pro} from group 1, 2a, 2b, and 3 coronaviruses

The substrate specificities of M^{pro} were profiled from different groups of CoVs using a 19x8 substrate library.[4] All of the M^{pro} could cleave the sequence SAVLQ↓SGF with specific activities of 443±11, 124±13, 180±5, and 174±19 min^{-1} mM^{-1} for HCoV-NL63 (group 1), HCoV-OC43 (group 2a), SARS-CoV (group 2b), and IBV (group 3), respectively (Fig 1).[3]

Engineering of a 'super-active' substrate sequence

Multiple favourable substitutions were combined to determine if a 'super-active' substrate sequence could be engineered. Based on the substrate specificities profiled, all M^{pro} favoured a Val at P5 and a Arg at P3 positions, with relative activities of 1.23 to 1.80 and 0.97 to 1.73, respectively (Table).[3] By combining both substitutions, the activity was increased to 1.70 to 3.24 (Table). The doubly substituted sequence, VARLQ↓SGF, appeared to be a good broad-spectrum substrate for all M^{pro}.

Substitution of P4-Val and P4-Pro resulted in higher than wild-type activity for SARS-CoV and IBV M^{pro}, respectively (Fig 1).[3] Based on this observation, the P4-Val or P4-Pro substitutions were introduced to VARLQ↓SGF, and two triply substituted substrates were generated. The protease activity for the triply substituted substrate (VVRLQ↓SGF) was further increased to 2.5 for SARS-CoV M^{pro} (Table).[3] This increase in activity for SARS-CoV M^{pro} was achieved at the expense of reduced activity for other M^{pro}. Similarly, >4-fold increase for IBV M^{pro} was achieved by combining all favourable substitutions (P3-Arg, P4-Pro, and P5-Val). The triply substituted sequence (VPRLQ↓SGF) represented the best 'super-active' sequence for IBV M^{pro}.

Design, synthesis, and characterisation of broad-spectrum peptidomimetic inhibitors

The autocleavage sequence (SAVLQ↓SGF) could

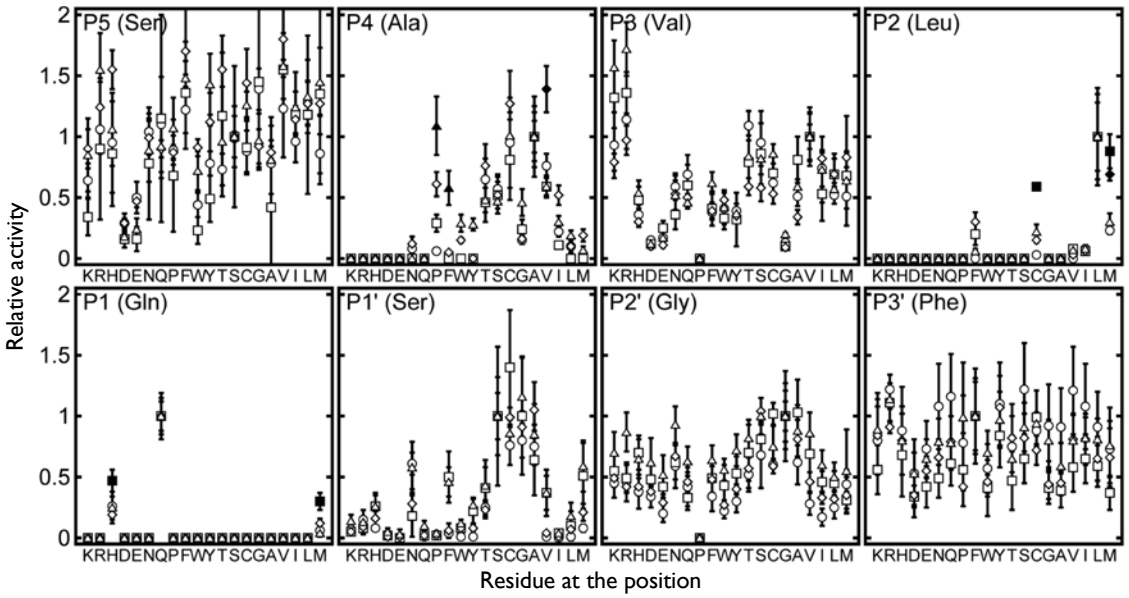

FIG 1. Substrate specificity of Mpro at P5 to P3' positions. The relative activities of Mpro from HCoV-NL63, HCoV-OC43, SARS-CoV, IBV are represented by circles, squares, diamond, and triangles, respectively. Significantly higher relative activities (P<0.001) are represented by filled symbols. This figure was adapted from: Chuck CP, Chow HF, Wan DC, Wong KB. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. PLoS One 2011;6:e27228

TABLE. Relative protease activity of doubly and triply substituted substrate variants. This table was adapted from: Chuck CP, Chow HF, Wan DC, Wong KB. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. *PLoS* One 2011;6:e27228

Substrate sequence	HCoV-NL63 (group 1)	HCoV-OC43 (group 2a)	SARS-CoV (group 2b)	IBV (group 3)
VAVLQ↓SGF	1.23±0.40	1.55±0.30	1.80±0.31	1.58±0.27
SARLQ↓SGF	1.14±0.24	1.36±0.17	0.97±0.12	1.72±0.22
VARLQ↓SGF	1.70±0.07	1.87±0.17	1.70±0.17	3.24±0.37
SPVLQ↓SGF	0.06±0.01	0.29±0.12	0.61±0.10	1.09±0.24
VPRLQ↓SGF	0.15±0.04	0.91±0.12	0.99±0.12	4.33±0.98
SVVLQ↓SGF	0.76±0.10	0.59±0.07	1.39±0.19	0.59±0.09
VVVLQ↓SGF	1.23±0.06	0.60±0.05	1.97±0.19	0.86±0.05
VVRLQ↓SGF	1.63±0.07	0.55±0.04	2.50±0.51	2.19±0.13

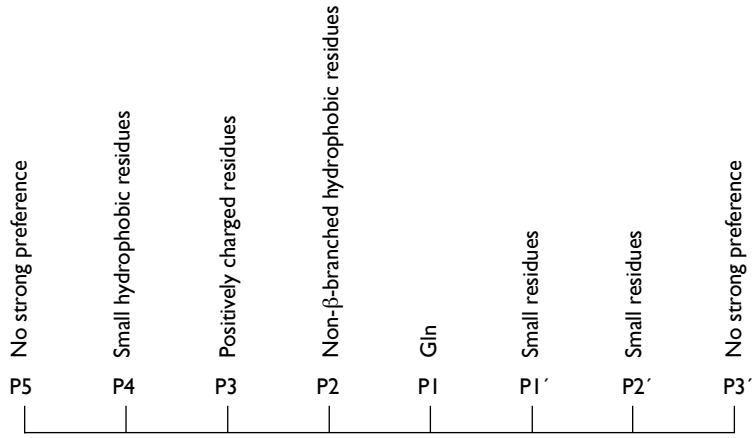

FIG 2. A summary of common substrate specificities for Mpro. This figure was modified based on Table 2 of: Chuck CP, Chow HF, Wan DC, Wong KB. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. PLoS One 2011;6:e27228

be cleaved efficiently by all M^{pro} tested. Based on a tetrapeptide-based inhibitor (Cbz-AVLQ-CN), the nitrile group was an efficient warhead in inhibiting M^{pro} with IC_{50} values of 4.6±0.2 μM.[5] The roles of P5 and P6 residues were then tested by extending the length of the peptide sequence and creating a hexapeptide inhibitor (Cbz-TSAVLQ-CN). Notably, inclusion of P5-Ser and P6-Thr residues did not improve the efficacy, as the IC_{50} value of the hexapeptide inhibitor was 39±1 μM.[5]

To understand the structural basis of the enzyme-inhibition interaction, the crystal structure of Cbz-TSAVLQ-CN in complex with SARS-CoV M^{pro} (PDB code: 3VB6) was compared to the structure of Cbz-AVLQ-CN in complex with SARS-CoV M^{pro} (PDB code: 3VB5).[5] The nitrile warhead of both inhibitors was covalently attached to the thiol group of the active site cysteine residue (Cys145). Residues at the P1 to P4 positions were well defined and fitted nicely into the S1-S4 substrate binding pockets of M^{pro}. In the crystal structure of Cbz-AVLQ-CN in complex with M^{pro}, the benzene group of the Cbz-AVLQ-CN was ordered and located in a pocket formed by Glu166-Pro168. In contrast, in the crystal structure of Cbz-TSAVLQ-CN in complex with M^{pro}, no electron density was observed for the Cbz protective group, P5-Ser, and P6-Thr residues. This suggested that they were disordered, probably due to the lack of a defined interaction with the M^{pro}, which may explain why the hexapeptide inhibitor had a higher IC50 value than the tetrapeptide inhibitor.

For tetrapeptide-based inhibitors, measurement of IC_{50} values of Cbz-AVLQ-CN was extended to other M^{pro}. The IC_{50} values were 2.8±0.1, 2.3±0.1, 1.6±0.1, 1.3±0.1, 4.6±0.2, 3.7±0.2 μM for HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV-HKU1, SARS-CoV, and IBV, respectively.[5] The inhibition of Cbz-AVLQ-CN was specific to M^{pro}, as there was no detectable inhibition at 0.5 to 256 μM against the control cysteine protease (Caspase 3).

Discussion

In this study, we comprehensively profiled the substrate specificity of M^{pro} from HCoV-NL63 (group 1), HCoV-OC43 (group 2a), SARS-CoV (group 2b), and IBV (group 3) using a 19x8 substrate library. Despite subtle differences, all M^{pro} share many similarities in their substrate preferences. The common substrate specificities of M^{pro} are summarised in Figure 2. The substrate preference at different positions was in general additive, as combining several favourable substitutions further increased the protease activities of M^{pro}.

That all M^{pro} shares similarity in their substrate specificities implies that it is feasible to design a broad-spectrum peptidomimetic inhibitor for M^{pro} from various strains of CoVs. The autocleavage sequence (SAVLQ↓SGF) is a very good broad-spectrum substrate, which was cleaved efficiently by all M^{pro} with specific activity ranging from 124 to 443 min^{-1} mM^{-1}.[3] This sequence is the starting point for further iteration in the design of broad-spectrum inhibitors for M^{pro}. By comparing the IC_{50} values of Cbz-TSAVLQ-CN and that of Cbz-AVLQ-CN, increasing the length of the peptide from tetrapeptide to hexapeptide did not improve inhibition.[5] This observation was justified because the P5-Ser and P6-Thr residues were disordered in crystal structure of Cbz-TSAVLQ-CN in complex with M^{pro}, as they did not have a significant interaction with the active site of the enzyme.[5] In contrast, the protective carbobenzyloxy group located at the C-terminus of P4 residues helped stabilise the enzyme-inhibitor complex. As P5 and P6 residues did not contribute

to an improved inhibition, the tetrapeptide-based inhibitor (Cbz-AVLQ-CN) was probably a good candidate for broad-spectrum inhibition against M^{pro}. In fact, this inhibitor could efficiently inhibit M^{pro} from six different strain of CoV, with IC_{50} values ranging from 1.3 to 4.6 μM.

Conclusions

We comprehensively profiled the substrate specificities of M^{pro} from group 1, 2a, 2b, and 3 CoVs at P5 to P3' positions. Differences and similarities in substrate specificities were identified. 'Super-active' substrates with >4-fold increase in activity were engineered by combining multiple favourable substitutions. Nitrile-based peptidomimetic inhibitors were effective against M^{pro}. Inclusion of P5 and P6 residues did not improve inhibitor efficacy, but rather, it was the N-terminal Cbz protective group that provided better enzyme-inhibitor interactions. The Cbz-AVLQ-CN inhibitor was a broad-spectrum inhibitor, as it could inhibit M^{pro} from six different strains of CoVs with IC_{50} values of 1.3 to 4.6 μM.

Acknowledgements

This study was supported by the Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#09080282). Results on substrate specificities were published in full in: Chuck CP, Chow HF, Wan DC, Wong KB. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. *PLoS One* 2011;6:e27228. Results on peptidomimetic inhibitors were published in full in: Chuck CP, Chen C, Ke Z, Wan DC, Chow HF, Wong KB. Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. *Eur J Med Chem* 2013;59:1-6.

References

1. Sloots TP, Whiley DM, Lambert SB, Nissen MD. Emerging respiratory agents: new viruses for old diseases? J Clin Virol 2008;42:233-43.
2. Woo PC, Lau SK, Huang Y, Yuen KY. Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med (Maywood) 2009;234:1117-27.
3. Chuck CP, Chow HF, Wan DC, Wong KB. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. PLoS One 2011;6:e27228.
4. 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.
5. Chuck CP, Chen C, Ke Z, Wan DC, Chow HF, Wong KB. Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. Eur J Med Chem 2013;59:1-6.