Substrate specificity and rational design of peptidomimetic inhibitors for SARS coronavirus main protease

KB Wong *, DCC Wan, HF Chow

KEY MESSAGES

1. Substrate-specificity of the main protease of SARS coronavirus was systematically profiled at P5 to P3' positions, which provided insights into a rational design of peptidomimetic inhibitors.
2. Leu and Gln were most favoured at P2 and P1 positions, respectively. Substrate preferences at P5 to P3 positions were important in enhancing the main protease activity. 'Super-reactive' substrate sequences were engineered, with more than a 2-fold increase in activity, by combining the best residue choices at P5 to P3 positions.
3. A novel class of peptidomimetic inhibitor against the main protease was developed using the nitrile warhead. The most potent inhibitor synthesised was Cbz-AVLQ-CN, with an IC50 value of 5 μM.
4. The crystal structure of the main protease in complex with Cbz-AVLQ-CN was determined, which provided structural insights into protease-inhibitor interactions for future structured-basis design of inhibitors.

Hong Kong Med J 2014;20(Suppl 4):S18-21
RFCID project number: 06060432

[1] **KB Wong *, [2] DCC Wan, [3] HF Chow**

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

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

The main protease (M^{pro}) of severe acute respiratory syndrome coronavirus (SARS CoV) is a key enzyme for viral replication, and is thus an attractive target for anti-SARS CoV drug development. M^{pro} belongs to the family of 3C-like cysteine proteases. The basic design of peptidomimetic inhibitors involved a warhead that can form covalent modifications to the –SH group of the active site residue Cys145 and a substrate peptide sequence that forms favourable interactions with the protease.

We systematically profiled the substrate specificity of M^{pro}, which forms the basis of a rational design of peptidomimetic inhibitors.[1] First, we created a library of protein-based substrates, and profiled the preference of amino acid residues at each of the P5 to P3' positions. Based on the substrate-specificity profile, we created 'super-reactive' substrate sequences. In addition, a novel peptidomimetic inhibitor was synthesised using nitrile as the warhead.[2] A number of inhibitors were synthesised to test the role of N-terminal protective groups and the substrate peptide sequences. Finally, the crystal structure of M^{pro} in complex with the best inhibitor was determined to provide a better understanding of protease-inhibitor interactions.

The results on substrate specificity profile of M^{pro} have been reported.[1] To profile the substrate-specificity of the SARS-CoV M^{pro}, saturated mutagenesis was performed at each of the P5 to P3' positions of the WT auto-cleavage sequences (SAVLQ↓SGF) [Fig 1]. A library of 19x8 variant substrate sequences was created and their relative activity was measured. At the P5 position, many substitutions exhibited higher activity than the WT substrate. The most preferred residue at the P5 position was Val. At the P4 position, small residues (Ala, Cys, Ser, Val, and Thr) were favoured. Substitutions with a bulky residue (Phe, Trp, Tyr, Leu, Met), or large polar residues (Lys, Arg, His, Asp, Glu, Asn) resulted in substrate sequences that had low relative activity. The best residue at the P4 position was Val. At the P3 position, positively charged residues (Lys, Arg) were favoured, but negatively charged residues (Glu, Asp) were repelled. The only non-cleavable substitution was V3P. The best residue this position was Arg. At the P2 position, only hydrophobic residues (Ala, Ile, Leu, Met, Phe, Pro, and Val) could be cleaved. The best residue was Leu. At the P1 position, the substrate sequence required a Gln residue to enable cleavage.[3-5] Surprisingly, M^{pro} was able to cleave substrate sequences containing a His or a Met. This finding challenges the established view that Gln is required at the P1 position. If M^{pro} can recognise His/Met at the P1 position, there may be more cleavage sites for M^{pro} along the SARS polyprotein. At the P1' position, small residues (Ala, Cys, Gly, and Ser) were favoured. Substitutions with large residues resulted in significant reduction in the relative activity. At the P2' position, small residues (Gly, Ala, Ser, and Thr) were also favoured, although

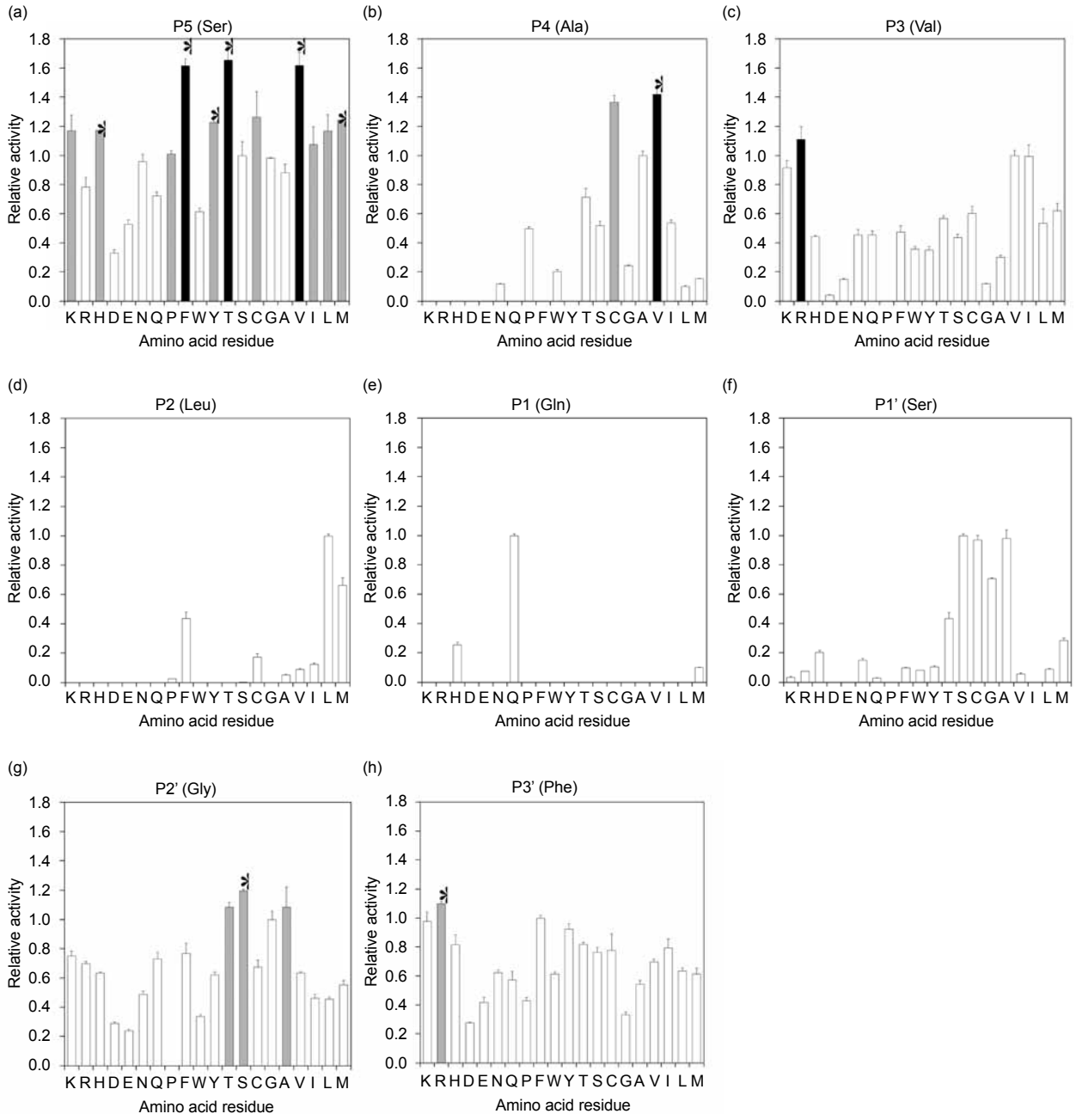


FIG 1. Substrate-specificity profiling of the P5 to P3' positions of SARS-CoV M^{pro}

Saturation mutagenesis is performed at the P5 to P3' positions of the autocleavage sequence of M^{pro} to create 19×8 variant substrate sequences. The k_{cat}/K_m values of M^{pro} on these variant substrate sequences are determined, and their activity relative to the wild-type sequence is shown. Substitutions resulted in higher activities are labelled in grey, whereas those selected for the design of super-active substrate sequence are labelled in black

* Significantly higher relative activity than that of the WT substrate

larger residues were also allowed. Only substitution with Pro resulted in a substrate sequence that could not be cleaved. Substitutions with Thr, Ser, and Ala resulted in a better substrate than the wild-type sequence. At the P3' position, although the substrate preference was loose, positively charged residues (Lys, Arg) were consistently better than negatively charged residues (Glu, Asp).

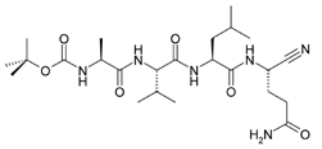
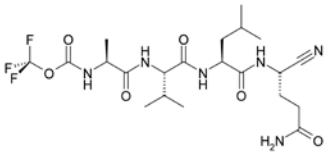
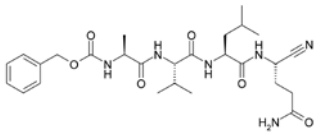
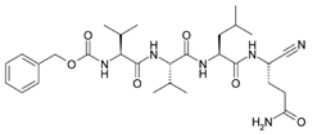
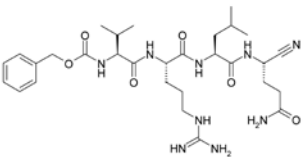
Substrate-specificity profiling suggested that single substitutions at the P5 to P3 positions affected the relative activities of the substrate sequences. We then combined the best substitutions at these positions to determine whether we could generate an even better substrate sequence.[1] Most of the substitutions could significantly improve relative activity. Among these 'super-reactive' variant,

'TVRLQ' and 'VVRLQ' were the best with relative activities of >2.5.

The M^{pro} activity on the 19x8 variant substrate sequences was determined by profiling the substrate specificity at each of the P5 to P3' positions (Fig 1). This systematic profiling study provided a rationale basis for peptidomimetic inhibitor design. Leu and Gln were the best choice for the P2 and P1 positions in the design of a peptidomimetic inhibitor. Residues at the P5 to P3 positions were important in the best substrate for M^{pro}. The trend observed was further confirmed by creating 'super-reactive' substrate via double and triple substitutions.[1]

The design and synthesis of peptidomimetic inhibitors have been reported.[2] First, we tried two warheads, nitrile and propargylamide groups, and found that only the nitrile group could serve as an efficient warhead. Next, we tested if the N-terminal protective group and the substrate sequence would influence the inhibitory effect (Table).[2] Our data suggested that the best protecting group was Cbz, and the best substrate sequence was AVLQ. Noteworthy, Cbz-VRLQ-CN had no observable inhibition on M^{pro} (Table). The nitrile warhead was not stable and was hydrolysed to amide (confirmed by mass spectrometry) in the inhibitor Cbz-VRLQ-CN, rendering it ineffective. The best inhibitor synthesised was Cbz-AVLQ-CN, with an M^{pro} value of 5 μM (Table).[2] To better understand how this inhibitor interacts with the M^{pro}, we determined the crystal structure of M^{pro} in complex with the inhibitor Cbz-AVLQ-CN. In the crystal structure, the nitrile warhead of the Cbz-AVLQ-CN inhibitor was attacked by the –SH group of the active site residue Cys145 to form a covalent-linked structure analogous to the acyl-enzyme intermediates (Fig 2). As a result, the side-chains of the inhibitor could form optimal interactions with the M^{pro}. Our structure of the M^{pro}-inhibitor complex also explained why Cbz was a better protecting group. The benzene ring of Cbz is inserted in a pocket to form favourable hydrophobic interactions with the Pro168, aliphatic chain of Glu166, and Val at the P3 position of the inhibitor. With the crystal structure of M^{pro}-inhibitor complex, we are in a much better position to improve the inhibitor activity against M^{pro} in the future.

Table. IC50 values of the peptidomimetic inhibitors synthesised

Inhibitor	Chemical structure	IC50
Boc-AVLQ-CN		52±12 uM
TFA-AVLQ-CN		>64 uM
Cbz-AVLQ-CN		5±1 uM
Cbz-VVLQ-CN		19±6 uM
Cbz-VRLQ-CN		No inhibition

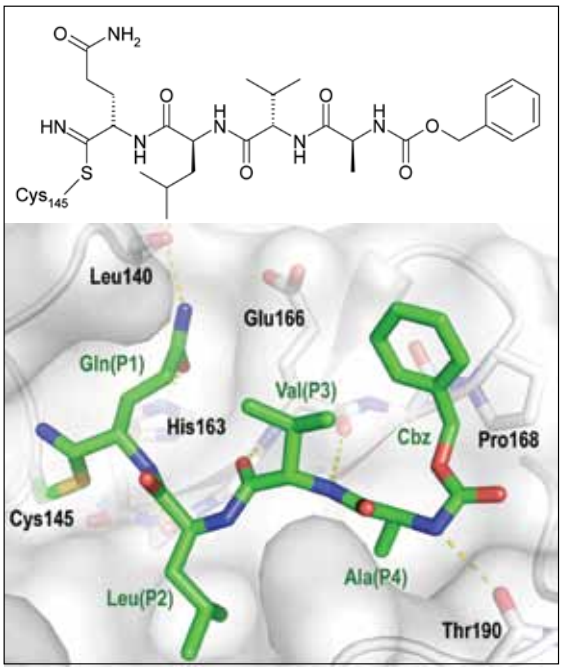


FIG 2. Structure of SARS-CoV Mpro in complex with inhibitor Cbz-AVLQ-CN
The inhibitor is covalently linked to the SG atom of Cys145. Hydrogen bonds are indicated as dotted lines

Acknowledgements

This study was supported by the Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#06060432). Results of this study were published in full in: 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. 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. 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.
2. 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.
3. Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol 2000;81:853-79.
4. Hegyi A, Ziebuhr J. Conservation of substrate specificities among coronavirus main proteases. J Gen Virol 2002;83:595-9.
5. Ziebuhr J, Heusipp G, Siddell SG. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. J Virol 1997;71:3992-7.