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.
The SARS-CoV-2 main protease as drug target

Sven Ullrich, Christoph Nitsche
Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

ABSTRACT

The unprecedented pandemic of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is threatening global health. The virus emerged in late 2019 and can cause a severe disease associated with significant mortality. Several vaccine development and drug discovery campaigns are underway. The SARS-CoV-2 main protease is considered a promising drug target, as it is dissimilar to human proteases. Sequence and structure of the main protease are closely related to those from other betacoronaviruses, facilitating drug discovery attempts based on previous lead compounds. Covalently binding peptidomimetics and small molecules are investigated. Various compounds show antiviral activity in infected human cells.

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the ongoing COVID-19 (coronavirus disease 2019) pandemic. Globally, 10 million infections have been confirmed with 500,000 fatalities.¹ The novel coronavirus had first been reported in Hubei province of China in late 2019, where it caused a major cluster of atypical pneumonia.²–⁴ Despite major efforts to contain the original outbreak, SARS-CoV-2 has since spread worldwide.⁵,⁶ Following the 2002/2003 SARS and 2012 MERS (Middle East respiratory syndrome) epidemics, this marks the third notable coronavirus outbreak in the 21st century.⁶,⁹

Four additional coronaviruses can infect humans, HCoV-229E, HCoV-HKU1, HCoV-NL63 and HCoV-OC43,¹⁰ which in stark contrast to the highly contagious and pathogenic SARS-CoV, MERS-CoV and SARS-CoV-2,⁷,¹¹,¹² cause only mild respiratory illness like the common cold.¹³ The case fatality rate (CFR) of COVID-19 is estimated to be lower than for SARS (~2–3) and MERS (~35%). However, its basic reproduction number (R₀) is potentially higher than for SARS (~10%) and MERS (~35%). However, its basic reproduction number (R₀) is potentially higher than for SARS (~2–3) and MERS (< 1). The values for CFR and R₀ of SARS-CoV-2 are still under controversial debate.¹⁴,¹⁵ Undetected and asymptomatic infections can challenge the accuracy of these parameters.¹⁶–¹⁸ The ongoing COVID-19 pandemic has had an unprecedented impact on individuals and the economy, as travel restrictions, social distancing and quarantine measures were implemented by many countries.¹⁹,²⁰

SARS-CoV, MERS-CoV and SARS-CoV-2 belong to the family of Coronavirusidae and the genus Betacoronavirus. Coronavirus infections are enveloped, positive-sense, single-stranded RNA viruses that feature the largest known RNA virus genomes ranging approximately from 26 to 32 kb,²²–²⁶ containing at least 6 (14 in case of SARS-CoV-2) open reading frames (ORFs).²⁷–²⁸ The major reading frame ORF 1ab encodes for two overlapping polyproteins (pp1a, pp1ab), which are cleaved into 16 non-structural proteins (nsp1-16) by the main protease Mpro (also referred to as 3CLpro) and the papain-like protease PLpro (Fig. 1).²⁷–²⁹,³² In addition, the papain-like protease is also a deubiquitinase.³³ The remainder of the genome encodes for accessory and structural proteins such as the spike glycoprotein (S), envelope protein (E), membrane protein (M) and the nucleocapsid phosphoprotein (N).²⁷,²⁹–³² The first genome sequence of SARS-CoV-2 deposited in Genbank was from the ~30 kb isolate Wuhan-Hu-1 (MN908947),³ which is used for sequence-related analyses in this article.

SARS-CoV-2 is closely related to other viruses of the Betacoronavirus genus such as the bat coronavirus BatCoV RaTG13 (~96% sequence identity) and the SARS-CoV (~80% sequence identity).³⁴,³⁵ SARS-CoV and MERS-CoV are both of zoonotic origin, with bats being their natural reservoir. Transmission to humans can occur via their intermediate hosts palm civets (SARS) and dromedary camels (MERS).³⁶ SARS-CoV-2 is thought to have followed a similar evolutionary transmission cascade.³⁷,³⁸ The spike glycoprotein (S) plays an important role in host range (tropism) and ‘host jumps’.³⁹,⁴⁰ In the case of SARS-CoV and SARS-CoV-2, it recognises the receptor angiotensin-converting enzyme 2 (ACE 2), which both viruses employ for cell entry.¹¹–¹³ Comparison of structural proteins of SARS-CoV-2, such as the spike protein (S), with those from animal coronaviruses indicate the involvement of bats as natural reservoir with a possibility for pangolins as intermediate hosts.³⁵,⁴⁴–⁴⁹ As the introduction of coronaviruses into human population has been observed on multiple occasions, a better understanding of the naturally circulating viruses is of high interest for pandemic control.

Keywords:
COVID-19
Coronavirus
Protease inhibitors
Peptidomimetics
Antivirals
preclinical studies that were based on previous compounds targeting the SARS-CoV Mpro. In contrast, sequence alignments reveal ~99% identity with BatCoV RaTG13 Mpro, which might prove valuable for the development of pan-coronaviral drugs and restore the catalytic dyad.

The main protease as a drug target

The current COVID-19 pandemic has triggered global efforts for the rapid identification of vaccines and specific antiviral treatments. Amongst the coronaviral targets that have been studied in the past, the main protease (Mpro, 3CLpro, nsp5) received major attention, particularly following the first SARS-CoV outbreak in the early 2000s. Alternative coronavirus targets include the spike protein (S), RNA-dependent RNA-polymerase (RdRp, nsp12), NTPase/helicase (nsp13) and cell deubiquitinases, making drug-discovery campaigns against PLpro challenging. In stark contrast, the main protease Mpro exclusively cleaves polypeptide sequences after a glutamine residue, positioning the main protease as an ideal drug target because, to the best of our knowledge, no human host-cell proteases are known with this substrate specificity.

Viral proteases are well validated drug targets that have led to various approved drugs, for example, against chronic infections with human immunodeficiency virus (HIV) or hepatitis C virus (HCV), which employ aspartyl and serine proteases, respectively. The SARS-CoV-2 Mpro proteolytically cleaves the overlapping pp1a and pp1ab polyproteins to functional proteins (Fig. 1), which is a critical step during viral replication. Replication-essential enzymes such as RdRp or nsp13 cannot fully function without prior proteolytic release, positioning Mpro as a key enzyme in the viral replication cycle. Consequently, its inhibition can stall the production of infectious viral particles and thus alleviate disease symptoms.

Capitalising on knowledge gained on structure and inhibitors of Mpro from previous coronaviral targets, the main protease as drug target was the focus of SARS-CoV-2 drug discovery research. Preclinical studies that were based on previous compounds targeting the SARS-CoV or MERS-CoV main proteases.

Mpro is a cysteine protease with a catalytic dyad (cysteine and histidine) in its active centre (Fig. 3). While other cysteine and serine proteases contain a third catalytic residue, a buried water molecule occupies this place in the active site of Mpro. The proteolytic process is believed to follow a multi-step mechanism. After the cysteine side chain proton is abstracted by the histidine’s imidazole, the resulting thiolate nucleophile attacks the amide bond of the substrate. The N-terminus of S is subsequently broken from the histidine before the thioester is hydrolysed to release the C-terminal peptide product and restore the catalytic dyad.

Mpro (nsp5) autopeptidolyses itself between pp4 and pp6, before processing the overlapping polyproteins pp1a and pp1ab at 11 cleavage sites (Fig. 1). While the Mpro monomer is basically inactive, the homodimer is the primary active species with both protomers almost orthogonal to each other. Each protomer consists of three domains. In case of SARS-CoV and SARS-CoV-2, domains I and II comprise residues 8–101 and 102–184, respectively, and include an antiparallel β-barrel with similarities to trypsin-like serine proteases. Domain II is connected to Domain III (residues 201–306) via a longer loop region (residues 185–200). Domain III is characterised by a cluster of five α-helices.

The proteases bind to each other via an N-terminal finger (residues 1–7) located between domains II and III, which is involved in the formation of the substrate-binding site in a cleft between domains I and II. It is known that the mutations Ser284Ala, Thr285Ala and Ile286Leu are present in SARS-CoV-2 Mpro, potentially explaining higher activity observed for SARS-CoV-2 Mpro compared to SARS-CoV Mpro. The Thr285Ala mutation is believed to bring domain III of both proteases closer to each other.

Substrate specificity and inhibitor design

According to the nomenclature introduced by Schechter and Berger, Mpro mainly recognises substrate residues ranging from P4 to P1’, Prime site recognition beyond P1’ is not conserved (Fig. 5). Specificity is mostly determined by P1, P2 and P1’, which show the highest degree of conservation amongst the cleavage sites. Glutamine in P1 is highly conserved in all polyprotein cleavage sites of SARS-CoV, MERS-CoV and SARS-CoV-2. In P4 more hydrophobic amino acids are tolerated with a clear preference for leucine. P1 tolerates small residues like serine or alanine.

Analysis of all polyprotein cleavage sites processed by Mpro for SARS-CoV, MERS-CoV and SARS-CoV-2 illustrates very similar substrate recognition profiles amongst these viruses (Fig. 5). Particularly important is the pronounced preference for glutamine in P1, strongly informing inhibitor design. Since no human
host-cell proteases with similar specificity are reported, reduced off-target effects are assumed for peptidomimetic inhibitors.60,61

The catalytic dyad of Mpro is located in a cleft between domains I and II (Fig. 4b).23,71 Despite the minor mutation S46A in close proximity, the active sites of SARS-CoV and SARS-CoV-2 Mpro are highly conserved. The influence of the S46A mutation on shape, size, flexibility and plasticity of the active site and its relevance for inhibitor design is under debate.87,88

As the monomer of Mpro is principally regarded as inactive, the dimerization interface offers an alternative target site for drug discovery.68,72 Although strong dimerization inhibitors are yet unavailable, the principle has been proven with the N-terminal octapeptide of SARS-CoV Mpro.89

Inhibitors

Inhibitors of SARS-CoV Mpro have been reviewed comprehensively by Pillaiyar et al. in 2016.72 MERS-CoV inhibitors have been reviewed by Liang et al. in 2018.90 Peptidomimetics and small molecules have been reported with affinities in the micro- to nanomolar range. They often depend on warhead-based design strategies, employing different reactive groups to covalently attack the catalytic cysteine residue. Warheads utilised include Michael acceptors, aldehydes, epoxy ketones and other ketones.72

Although SARS-CoV-2 emerged only very recently, several inhibitors have already been identified and successfully co-crystallised with Mpro (Fig. 6).59,79,91 They are often derived from previous campaigns which targeted the main proteases of SARS-CoV or MERS-CoV and contain cysteine-reactive warheads (Fig. 6).

The first reported inhibitors were covalently binding peptidomimetics (1–3) addressing the major substrate-recognition motif from P1′ to P3.59 They all comprise an α-ketoamide functionality that forms a hemithioacetal with Cys145. α-Ketoamides are already used as viral serine protease warheads in the approved HCV drugs telaprevir and boceprevir.62 Compound 1 has previously been investigated as a broad-spectrum corona- and enteroviral protease inhibitor.60 Like many other Mpro inhibitors, the P1 side chains of 1–3 employ a γ-lactam as a glutamine mimetic. P2 comprises hydrophobic cyclohexyl (1, 2) or smaller cyclopropyl (3) groups as leucine mimetics and P1′ contains cyclopropyl (2) or benzyl (1, 3) residues. Compounds 2 and 3 feature pyridone rings between P2 and P3 as well as N-terminal Boc groups, which were associated with increased plasma half-life, kinetic solubility and thermodynamic solubility. Pharmacokinetic profiling of 2 and 3 in mice also indicated favourable lung tropism. Compounds 1 and 3 displayed sub-micromolar Mpro inhibition (Fig. 7). Compound 3 is similarly active against the SARS-CoV and MERS-CoV main proteases and inhibits
SARS-CoV-2 replication in human Calu3 lung cells.

Compound 4 is another peptidomimetic Mpro inhibitor co-crystallised in complex with SARS-CoV-2 Mpro (Fig. 6b).79 It originated from previous campaigns targeting SARS-CoV Mpro.92 Its Michael acceptor irreversibly modifies Cys145. Compound 4 shows anti-SARS-CoV-2 activity in Vero cells.

Compounds 5 and 6 are the strongest known SARS-CoV-2 Mpro inhibitors with inhibition constants in the two-digit nanomolar range. These small peptidomimetics feature an indole moiety at the N-terminus (P₃) and a C-terminal aldehyde warhead which binds covalently to Cys145, as proven by X-ray crystallography (Fig. 6c).93 Similar peptide-aldehydes have previously been explored as inhibitors of SARS-CoV Mpro.93,94 Compound 5 and the previously investigated α-ketoamides 1 and 2 are structurally identical in P₃ and P₅; however, 5 inhibits one order of magnitude stronger, which is likely caused by the increased electrophilicity of the aldehyde warhead compared to the more drug-like α-ketoamide. Despite, the high reactivity of the aldehyde function, compounds 5 and 6 displayed sub-micromolar antiviral activity in Vero cells (Fig. 7). In agreement with compound 3, the in vitro activity of 5 and 6 was one order of magnitude weaker than the direct Mpro inhibition in the enzymatic assay. Notably, 5 exhibited only low toxicity in animal models despite its aldehyde warhead.91

A high-throughput screening campaign of a library of approved drugs and clinical candidates revealed six small molecules, ebselen, disulfiram, carmofur, tideglibus, shikonin and PX-12, as inhibitors of SARS-CoV-2 Mpro (Fig. 7).79 Mass spectrometry experiments showed that ebselen, carmofur and PX-12 covalently modify Cys145. Small covalent modifiers like these may bind unspecifically, a characteristic associated with pan-assay interference compounds (PAINS).95 Ebselen showed antiviral activity in Vero cells in the low micromolar range. In case of carmofur, a crystal structure of SARS-CoV-2 Mpro revealed transfer of the hexylurea side chain to Cys145, forming a hexylcarbamothioate interacting with the S₂ subsite (PDB: 7BUY).96

Crystal structures of SARS-CoV-2 Mpro in complex with X77 and baicalein (Fig. 7) have been deposited in the protein data bank under the accession codes 6W63 and 6M2N, respectively. Notably, both compounds bind non-covalently to the active site. X77 and derivatives have previously been investigated as low-micromolar inhibitors of SARS-CoV Mpro, where a strong stereochemical bias for the (R) enantiomer of the pyridyl side chain has been observed.97,98 In the X-ray co-crystal structure of X77, the aforementioned pyridyl side chain acts as a P₃ mimic. Baicalein is a flavonoid found in Scutellaria baicalensis, a plant used in traditional Chinese medicine.99 Several flavonoids and derivatives had previously been reported to inhibit the activity of SARS-CoV Mpro.100

A fragment screening has produced several crystal structures of fragments bound to SARS-CoV-2 Mpro, including covalent modifiers of Cys145.101 While the majority of fragments bind to the active site, some...
bind near the dimer interface of SARS-CoV-2 Mpro. These fragments may inform the development of small-molecule inhibitors that are not substrate-derived.

While this manuscript was under peer-review, a study assessing known protease inhibitors for their anti-SARS-CoV-2 Mpro activity was published.102 Amongst the compounds that displayed Mpro inhibition and reduction of cytopathic effect were the α-ketoamides boceprevir ($K_i = 1.18 \mu M$, $EC_{50} = 1.31 \mu M$) and calpain inhibitor XII ($K_i = 0.13 \mu M$, $EC_{50} = 0.49 \mu M$), the peptide-aldehyde calpain inhibitor II ($K_i = 0.40 \mu M$, $EC_{50} = 2.07 \mu M$) and the sulfonate-featured peptide GC-376 ($k_2/K_i = 40,800 M^{-1} s^{-1}$, $EC_{50} = 3.37 \mu M$). A crystal structure of SARS-CoV-2 Mpro in complex with GC-376 was also reported (PDB: 6WTT).

**Conclusion and outlook**

The COVID-19 pandemic poses a major challenge to mankind. In view of the magnitude of the current global crisis, numerous attempts to develop vaccines and antiviral treatments are obviously underway. With respect to drug development, the main protease of SARS-CoV-2 stands out as a promising viral target, as it differs significantly from human proteases. Given the conserved structure and specificity of Mpro amongst SARS-CoV, MERS-CoV and SARS-CoV-2, pan-coronaviral main protease inhibitors might become available. However, in line with previously successful examples like HIV or HCV, the development of novel specific protease inhibitors and their approval will take several years. Although this process will likely take too long to impact on the current COVID-19 crisis, protease inhibitors would be worth pursuing as they may provide specific drugs for upcoming coronavirus outbreaks.

Pharmacodynamic and pharmacokinetic properties of peptidomimetic Mpro inhibitors like 2, 3, 5 or 6 already point into the right direction.59,91 Although peptide-aldehydes have entered clinical trials before (e.g. efegatran),103,104 5 and 6 are likely to face challenges during further drug development. The α-ketoamide in 1–3 or Michael acceptor in 4 are covalent modifiers with precedents in approved drugs (e.g. telaprevir or afatinib).62,105 Potential problems associated with limited drug-likeness of peptidomimetics could be circumvented by pursuing alternative small molecules, which might, for example, be accessible from fragment-based drug discovery campaigns.101,106 Repurposing of known drugs can provide an accelerated route to approval, which is likely the only option to address the current COVID-19 crisis. Small molecules like ebelsen or carmofur are Mpro inhibitors with anti-SARS-CoV-2 activity in cells;79 however, their thiol reactivity might prove challenging. Repurposing approved protease inhibitors is an alternative approach.102 Attempts to repurpose the approved combination of HIV protease inhibitors, ritonavir and lopinavir, was unsuccessful in clinical studies, which is not entirely unexpected, given the differences between the proteases of HIV and SARS-CoV-2.107 It is likely that SARS-CoV-2 is not the last human coronavirus emerging from animals. It is therefore important to closely monitor virus populations to understand their replication mechanism early on and investigate druggable targets. A sharp decline in research funding had been noted a few years after the first SARS epidemic.61 Given the long-term nature of drug discovery projects, this has proven disastrous with respect to the current crisis. Now is the best time to progress protease inhibitors to anti-coronaviral drugs.

**Fig. 6.** X-ray co-crystal structures of SARS-CoV-2 Mpro with covalently binding peptidomimetic inhibitors. Mpro is shown as hydrophobicity surface (red indicating hydrophobic and blue hydrophilic surface areas) and grey ribbon with amino acid side chains (UCSF Chimera).70 Inhibitor groups binding to protease subsites are indicated according to the Schechter Berger nomenclature.61 Electrophilic warheads covalently binding to Cys145 are circled. (a) Compound 3 with an α-ketoamide warhead (PDB: 6Y2F). (b) Compound 4 with a Michael acceptor warhead (PDB: 7BQY). (c) Compound 5 with a C-terminal aldehyde warhead (PDB: 6LZE).
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

C.N. thanks the Australian Research Council for a Discovery Early Career Research Award (DE190100015). S.U. acknowledges a PROMOS scholarship from the German Academic Exchange Service.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20(5):533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020;395(10223):514–523. https://doi.org/10.1016/S0140-6736(20)30154-9.
3. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382(8):727–733. https://doi.org/10.1056/NEJMoa2001017.
4. Zhang D, Wang W, et al. An interactive web-based dashboard to track COVID-19 in real-time. Lancet Infect Dis. 2020;20(5):533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.
5. Ma Q, Guan W, et al. A family cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020;395(10223):514–523. https://doi.org/10.1016/S0140-6736(20)30154-9.
6. Yang Y, Peng F, Wang R, et al. The deadly coronaviruses: the 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J Autoimmun. 2020;109:102434. https://doi.org/10.1016/j.jaut.2020.102434.
7. Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. 2020;382(18):1708–1720. https://doi.org/10.1056/NEJMoa2002032.
8. Lau H, Khosrawipour V, Kocbach P, et al. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. J Travel Med. 2020;27(3):taaa037. https://doi.org/10.1093/jtm/taaa037.
9. Da Costa VG, Morel ML, Saivish MV. The emergence of SARS, MERS and novel SARS-CoV-2 coronaviruses in the 21st century. Arch Virol. 2020;165(7):1517–1526. https://doi.org/10.1007/s00705-020-04628-0.
10. Fauci AS. Coronavirus infections—more than just the common cold. JAMA. 2020;323(8):707–708. https://doi.org/10.1001/jama.2020.0757.
11. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Nat Med. 2020;26(5):533–534. https://doi.org/10.1038/s41591-020-0843-1.
Petterse EF, Goddard TD, Huang CC, et al. UCSF Chimera - a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605–1612. https://doi.org/10.1002/jcc.20084.

Ho B-L, Cheng S-C, Shi L, Wang T-Y, Ho K-I, Chou C-Y. Critical assessment of the important residues involved in the dimerization and catalysis of MERS coronavirus main protease. PLoS ONE. 2015;10(12):e014465. https://doi.org/10.1371/journal.pone.014465.

Pillayar T, Manickam M, Namavizayam V, Hayashi Y, Jung S-H. An overview of severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy. J Med Chem. 2016;59(14):6595–6628. https://doi.org/10.1021/acs.jmedchem.5b01461.

Huang C, Wei P, Fan K, Liu Y, Lai L. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry. 2004;43(15):4568–4574. https://doi.org/10.1021/bi036022q.

Xia B, Kang X. Activation and maturation of SARS-CoV main protease. Protein Cell. 2011;2(4):282–289. https://doi.org/10.1007/s13238-011-1004-1.

Muramatsu T, Kim Y-T, Nishii W, Terada T, Shirouzu M, Yokoyama S. Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CLpro) from its polyproteins. PLoS ONE. 2013;8(19):2002–2012. https://doi.org/10.1111/feb3.12222.

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(35):31257–31266. https://doi.org/10.1074/jbc.M502772020.

Chou C-Y, Chang H-C, Hsu W-C, Lin T-Z, Lin C-H, Chang G-G. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. Biochemistry. 2004;43(47):14958–14970. https://doi.org/10.1021/bi0490237.

Ul Qamar MT, Alqahtani SM, Alamri MA, Chen L-L. Structural basis of SARS-CoV-2 main protease (Mpro) for small-molecule inhibitor design. Int J Mol Sci. 2020;21(9):3099. https://doi.org/10.3390/ijms21093099.

S. Ullrich and C. Nitsche