Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease

Elizabeth A. MacDonald,1 Gary Frey,2,3 Mark N. Namchuk,4 Stephen C. Harrison,1,4,5 Stephen M. Hinshaw,4,* and Ian W. Windsor,1,4,6,*

*Correspondence: hinshaw@crystal.harvard.edu and windsor@crystal.harvard.edu

1Laboratory of Molecular Medicine, Boston Children’s Hospital, Boston, MA, USA
2Department of Pediatrics, 3ICCB-Longwood Screening Facility, and 4Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA, USA
5Howard Hughes Medical Institute, Boston, MA, USA
6Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA

KEYWORDS: SARS-CoV-2, Mpro, protease, virology

ABSTRACT: The main protease (Mpro) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the cause of coronavirus disease (COVID-19), is an ideal target for pharmaceutical inhibition. It is required for infection, it cleaves the viral polyprotein at multiple sites, and it is conserved among coronaviruses and distinct from human proteases. We present crystal structures of SARS-CoV-2 Mpro bound to two viral substrate peptides. The structures show how Mpro recognizes substrates and how the peptide sequence can dictate catalytic efficiency by influencing the position of the scissile bond. One peptide, constituting the junction between viral non-structural proteins 8 and 9 (nsp8/9), has P1′ and P2′ residues that are unique among SARS-CoV-2 cleavage sites but conserved among nsp8/9 junctions in coronaviruses. Mpro cleaves nsp8/9 inefficiently, and amino acid substitutions at P1′ or P2′ can enhance catalysis. Visualization of Mpro with intact substrates provides new templates for antiviral drug design and suggests that the coronavirus lifecycle selects for finely tuned substrate-dependent catalytic parameters.

TEXT: Developing and stockpiling pan-coronavirus antiviral drugs for pandemic prevention has been a goal since the SARS outbreak of 2003.1,2 The coronavirus main protease (nsp5 or Mpro) is a conserved drug target and an important focus of these efforts. Hundreds of Mpro inhibitors have been reported. Most of these drugs occupy the active site cleft responsible for recognizing the N-terminal half of substrate peptides, and many form covalent bonds to the active site cysteine of Mpro (Cys145).3,4 A recent crystal structure of the nsp5/6 acyl-enzyme intermediate provides one template for chemical mimicry of this essential catalytic step.4 We provide evidence that contacts on both sides of the Mpro catalytic site can affect the rate of formation of the covalent complex, a characteristic that could be exploited by new protease inhibitors.

The nsp8/9 junction is a conserved Mpro substrate (Figure 1A-B). The nearly invariant Asn residues at P1′ and P2′ are unique within a given coronavirus polyprotein; all other cleavage sites have Gly, Ser, or Ala at P1′, P2′, or both sites.5 Cleavage of nsp8/9 is slow but required for replication of the closely-related Murine Hepatitis Virus.6 Indeed, a recently-determined cryo-EM structure shows that the N-terminus of nsp9 contacts nsp12, a core component of the viral RNA polymerase.7 In this context, the nsp9/8 P1′ to P3′ residues contribute to a binding site for a nucleotide that is transferred to the amino terminus of the P1′ residue.8 Therefore, the nsp9/8 junction has evolved to satisfy two evolutionary constraints required for viral replication: it must be cleaved in the Mpro active site, and it must serve as a substrate in a nucleotide monophosphate transfer reaction catalyzed by the nsp12 protein. We have used X-ray crystallography to study nsp8/9 and nsp4/5 recognition by Mpro. The structures show unique features of the Mpro-nsp8/9 complex constrained by natural selection and highlight the importance of P1′-P3′ residues in catalysis.

To study Mpro activity, we monitored cleavage of labeled substrate peptides in vitro and derived Michaelis-Menten parameters describing the reactions. Mpro cleavage of nsp8/9 is less efficient than cleavage of nsp4/5 (36-fold decrease in kcat/Km; Table 1).9 We also sought to understand the influence of the Asn residues at the P1′ and P2′ sites of the nsp8/9 substrate (Table 1). Alanine substitutions at either position approximately doubled the catalytic efficiency. The P1′ Asn-to-Ala substitution lowered K_m and raised k_cat, while the P2′ substitution only raised.

Figure 1. Viral substrates and structures of substrate-bound Mpro. (A) Protein sequence alignment of the eleven SARS-CoV-2 Mpro cleavage sites required for maturation of SARS-CoV2. (B) Protein sequence alignment of nsp8/9 Mpro cleavage sites from representative coronaviruses. (C-D) Crystal structures showing nsp5/5 (C) or nsp8/9 (D) substrate engagement by Mpro. Protease protomers are colored dark or light gray. A single peptide is colored for each, though peptide occupancy all active sites in both determined structures. The vertical arrow marks the two-fold axis of the Mpro dimer.
Installation of an isosteric Asp residue at the P1’ position completely abrogated activity, while the analogous Asn-to-Asp substitution at P2’ diminished but did not abrogate activity. We suspect that placing additional negative charge near the active site raises the energetic barrier to attaining the oxanion transition states.14, 15

Differences in $k_{cat}$ for the peptide substrates tested dominated the small changes in $K_M$ and drove the observed changes in $k_{cat}/K_M$. The P5-P1 residues were constant for nsp8/9 and its derivatives, ruling out acyl enzyme hydrolysis as the step that determines $k_{cat}$. Therefore, either formation of the enzyme-substrate complex or conversion to the acyl enzyme intermediate must limit $k_{cat}$ for the nsp8/9 substrate, and similar $K_M$ values imply the latter is true. Unlabeled nsp8/9 did not inhibit cleavage of labeled nsp4/5 at concentrations up to 250 μM, while the control inhibitor MG-132 produced a $K_c$ of 740 nM (Figure S2). Its inability to interfere with nsp4/5 cleavage confirms that nsp8/9 is a poor M<sup>pro</sup> substrate relative to nsp4/5. Similar competition among the 11 viral M<sup>pro</sup> substrates occurs during virus replication.

To understand the structural basis for differential cleavage efficiency, we determined crystal structures of M<sup>pro</sup> bound to the nsp4/5 and nsp8/9 substrates (Figure 1C-D). The active site Cys145Ala mutation trapped the intact substrates and enabled visualization of the P′ residues (Figure S3). A principal difference between the nsp4/5 and nsp8/9 substrates is the position of the scissile bond. The nsp8/9 P1 Ca is translated 0.4 Å along the C–O axis relative to nsp4/5 (Figure 2). Nevertheless, 11 conserved hydrogen bonds occur between M<sup>pro</sup> and each of the substrates (Figure 3, white dotted lines). Eight contacts between the peptide backbones of enzyme and substrate are shared among SARS-CoV nsp4/5, PEDV nsp4/5,16, 17 and the two SARS-CoV-2 peptides reported here. M<sup>pro</sup> Gly143 and Ala145 mainchain amides form the oxanion hole by donating a pair of hydrogen bonds to the scissile P1 carbonyl oxygen, which stabilizes the developing negative charge during covalent catalysis. His163 and the mainchain carbonyl of Phe140 make hydrogen bonds with the invariant sidechain of the P1 Gln, and Asn142 contacts the P1 Gln through a conserved water bridge. Neither SARS-CoV nor PEDV M<sup>pro</sup>/nsp4/5 complexes show the hydrogen bonds observed with Asn142 in the SARS-CoV2 substrate complexes.16, 17

Substrate interactions with the M<sup>pro</sup> Asn142 and Gln189 sidechains distinguish nsp4/5 and nsp8/9 recognition (Figure 3, green and magenta dotted lines). M<sup>pro</sup> Asn142 forms a hydrogen bond with the nsp4/5 P1′ backbone carbonyl oxygen, and M<sup>pro</sup> Gln189 forms a water bridge with the nsp4/5 P2 amide nitrogen. In contrast, M<sup>pro</sup> Gln189 engages the nsp8/9 P3 and P1′ side chains via an ordered water molecule. In addition to these contacts, the ordered waters found in the nsp8/9-bound structure could donate hydrogen bonds to the P1′ and P2 mainchain carbonyl oxygens. Finally, the nsp8/9 P3 Lys forms a hydrogen bond with the P2 carbonyl. This interaction and the ordered waters described above likely account for subtle bending of the nsp8/9 substrate P-fragment away from the active site cleft relative to the nsp4/5 peptide.

The peptide recognition described above produces the different catalytic efficiencies associated with cleavage of the nsp4/5 and nsp8/9 substrates. The near-invariant nsp8/9 P1’ and P2’ Asn side chains are bulkier than the P1′ (Ser/Ala) and P2’ (Gly/Ala) side chains of other M<sup>pro</sup> substrates, though there is greater tolerance for P2’ diversity.18 The nsp8/9 P1′ Asn reaches more deeply into the S′ subsite than the nsp4/5 P1′ Ser and therefore likely restrains the P′ peptide to a greater degree. M<sup>pro</sup> Asn142 and Gly143 coordinate the nsp8/9 P2′ residue through peptide backbone interactions, and similar interactions position the nsp4/5 P2′ Gly (Figure 4A-B). Overall, the bulkier nsp8/9 Asn side chains in the S1′ and S2′ subsites shift nsp8/9 relative to nsp4/5 (Figure 2). The resulting alignment with the M<sup>pro</sup> cysteine nucleophile differs from that of nsp4/5 and provides an explanation for reduced catalytic efficiency that is partially restored by alanine substitution. That the restoration is incomplete relative to nsp4/5 suggests that both P- and P′- substrate fragments align the scissile bond.

Hydrophobic interactions dictate recognition of N-terminal substrate fragments (P residues, excluding the invariant P1 Gln). M<sup>pro</sup> Met49 and Met165 define the S4, S2, and S1′ subsites (Figure 4). The nsp4/5 P3 Ala is smaller than the nsp8/9 P3 Val, allowing the P4 Ala of nsp4/5 to sit more deeply in the S4 subsite. This produces a 1.0 Å shift between the nsp4/5 and nsp8/9 P3 residues (measured from corresponding Ca atoms). A recent crystal structure shows that the intact nsp5/6 substrate is also shifted relative to nsp4/5 due to a bulky Phe at the P2 position (Figure 4C).8 Indeed, M<sup>pro</sup> cleavage is most efficient for peptides bearing P2 Leu and less efficient for those bearing

### Table 1. Catalytic efficiencies for M<sup>pro</sup> substrates and analogs

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Sequence&lt;sup&gt;a&lt;/sup&gt;</th>
<th>$k_{cat}$ (s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>$K_M$ (μM)</th>
<th>$k_{cat}/K_M$ (M&lt;sup&gt;−1&lt;/sup&gt;s&lt;sup&gt;−1&lt;/sup&gt;)</th>
<th>Fold change&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>nsp4/5</td>
<td>TSAVLQ/GFGRKM</td>
<td>0.52 ± 0.07</td>
<td>41 ± 9</td>
<td>1.3 ± 0.3 × 10&lt;sup&gt;4&lt;/sup&gt;</td>
<td>–</td>
</tr>
<tr>
<td>nsp8/9</td>
<td>RVVKLQ/NNELMP</td>
<td>0.013 ± 0.001</td>
<td>36 ± 6</td>
<td>3.6 ± 0.7 × 10&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1.0</td>
</tr>
<tr>
<td>nsp8/9 N1′A</td>
<td>RVVKLQ/aNELMP</td>
<td>0.022 ± 0.001</td>
<td>22 ± 3</td>
<td>1.0 ± 0.1 × 10&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2.9</td>
</tr>
<tr>
<td>nsp8/9 N2′A</td>
<td>RVVKLQ/NaELMP</td>
<td>0.034 ± 0.002</td>
<td>46 ± 5</td>
<td>7.5 ± 0.8 × 10&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2.1</td>
</tr>
<tr>
<td>nsp8/9 N1′D</td>
<td>RVVKLQ/dNELMP</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>nsp8/9 N2′D</td>
<td>RVVKLQ/NdELMP</td>
<td>0.0029 ± 0.0001</td>
<td>19 ± 1</td>
<td>1.6 ± 1.2 × 10&lt;sup&gt;2&lt;/sup&gt;</td>
<td>0.4</td>
</tr>
</tbody>
</table>

<sup>a</sup>Lys-DABCYL and Glu-EDANS and are appended to the N- and C-termini. Residues that differ from the wild-type sequence are lower case.

<sup>b</sup>Fold change = ($k_{cat}/K_M$) nsp8/9 analog / ($k_{cat}/K_M$) nsp8/9

---

MacDonald et al.
P2 Phe. Like nsp8/9, cleavage of SARS nsp5/6 depends more heavily on P' recognition than does nsp4/5. The nsp8/9 P3 Lys might also limit catalysis. Water bridges connect its terminal nitrogen (Nζ) with the nsp8/9 P1' Asn (mentioned above), and the resulting conformation could slow peptide accommodation to the Mpro active site. Therefore, the diverse Mpro-substrate interactions can fine-tune substrate geometry and position the scissile bond to adjust substrate-specific activity.

The structures we have resolved show how Mpro active site plasticity and substrate evolution can tune catalysis. Slow cleavage of the nsp8/9 junction, which is observed among disparate coronaviruses, might be a selected trait required for coordinated assembly of the RNA replication machinery. Inability of the nsp8/9 substrate to inhibit nsp4/5 cleavage is likely important for maturation of the viral polyprotein. The need for the nsp8/9 junction to support both Mpro cleavage and nsp12 binding (and subsequent nucleotide monophosphate acceptance) accounts for the near-invariance of the P1-P2' residues. The sequence is therefore a compromise that satisfies the requirements of two unrelated catalytic mechanisms, and mimicry of the nsp8/9 junction presents a unique opportunity to chemically inhibit both Mpro and the viral polymerase.

The structures also present templates for new protease inhibitor scaffolds. In particular, that the nsp8/9 P3 sidechain can fold back to contact P1' suggests macrocyclic inhibitors could mimic this interaction. Similar strategies have been pursued for Hepatitis C NS3, HIV-1, and Rhinovirus 3C proteases. Kinetic analyses of nsp8/9 and its variants suggests that inhibitor P1' and P2' site contacts could influence formation of covalent inhibitor-enzyme adducts. α-ketoamide warheads, which have not been exhaustively explored as ligands for Mpro Cys145, are good candidates for this objective.

ASSOCIATED CONTENT
Supporting Information
Supporting Information document contains methods, inhibition experiments, and crystallographic information. Crystallographic data has been deposited at the PDB with associated codes 7MGR (Mpro-nsp8/9) and 7MGS (Mpro-nsp4/5).

AUTHOR INFORMATION
Corresponding Author
* Stephen M. Hinshaw – hinshaw@crystal.harvard.edu and Ian W. Windsor – windsor@crystal.harvard.edu

Present Address
† Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources
We acknowledge funding support from the Massachusetts Consortium for Pathogen Readiness (M.N.N.), the Howard Hughes Medical Institute (S.C.H.), and the Nancy Lurie Marks Family Foundation (S.C.H.). S.M.H. is an HHMI Fellow of the Helen Hay Whitney Foundation. NE-CAT is supported by NIH grant P30 GM 124165.

Figure 3. Differential recognition of nsp4/5 and nsp8/9 substrates by Mpro. Identical views of nsp4/5 (A) and nsp8/9 (B) substrates in the Mpro·Cys145Ala active site. Substrate peptide P and P’ residues are labeled with colored numbers. Key Mpro residues mentioned in the text are labeled. Conserved hydrogen bonds enabling Mpro recognition of substrate mainchain and P1 Gln side chain atoms are shown as white dotted lines. The hydrogen bonds between Mpro and nsp4/5 (green) differ from those with nsp8/9 (magenta). Mpro Asn142 and Glu189 contact the nsp4/5 substrate through bound water molecules, while the same Mpro residues contact the nsp8/9 peptide at distinct sites through networked water molecules.

Figure 4. Steric effects that influence Mpro recognition and activity. Key atoms dictating the shape complementarity between Mpro subsites and corresponding side chains from: (A) nsp4/5, (B) nsp8/9, and (C) nsp5 (acyl-enzyme intermediate; PDB 7KHP).
ACKNOWLEDGMENTS

Enzyme assays were performed at the Institute for Chemistry and Chemical Biology (ICCB-Longwood) at Harvard Medical School. We thank Linlin Zhang (Hilgenfeld Laboratory, University of Lübeck) for providing the codon-optimized Mrp expression plasmid. X-ray diffraction data were collected at the Advanced Photon Source (APS) on NE-CAT beamline 24-IDC.

REFERENCES

