Identification of SARS-CoV2 Main Protease Coldspots Suitable for Drug Targeting

Navaneethakrishnan Krishnamoorthy[1,*] and Khalid Fakhro[1,2,3]

[1]Department of Human Genetics, Sidra Medicine, Doha, Qatar

[2]Department of Genetic Medicine, Weill Cornell Medical College, Doha, Qatar

[3]College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar

*Correspondence: nkrishnamoorthy2@sidra.org

Most attempts to target the novel coronavirus SARS-CoV2 are focusing on the main protease (M^{pro}) [1-9]. However, >19,000 missense mutations in the M^{pro} have already been reported [10]. The mutations encompassing 282 amino acid positions and these "hotspots" might change the M^{pro} structure and activity, potentially rendering novel antivirals and vaccines ineffective. Here we identified 24 mutational "coldspots" that have resisted mutation since the virus was first detected. We compared the structure-function relationship of these coldspots with several SARS-CoV2 M^{pro} X-ray crystal structures. We found that three coldspot residues (Leu141, Phe185 and Gln192) help to form the active site, while six (Gly2, Arg4, Tyr126, Lys137, Leu141 and Leu286) contribute to dimer formation that is required for M^{pro} activity. The surface of the dimer interface is more resistant to mutations compared to the active site. Interestingly, 16 coldspots are found in conserved patterns when compared with other coronaviruses. Importantly, several conserved coldpots are available on the surface of the active site and at the dimer interface for targeting. The identification and short list of these coldspots offers a new perspective to target the SARS-CoV2 M^{pro} while avoiding mutation-based drug resistance.

Keywords: Mutation Hotspot, X-ray Structure, Missense Mutation, Structure-function Relationship, Dimer interface, Mutation-based Drug Resistance

Mutational hotspots and coldspots

The SARS-CoV2 main protease (M^{pro}) or 3CL-protease ($3CL^{pro}$) is essential for its proteolytic activity, structural protein production and host cell infection[11]. Early in the COVID-19 pandemic, mutational hotspots were reported within SARS-CoV2 genomic sequences [12,13]. Mutational coldspots within viral genomes and/or proteins can indicate the location of appropriate targets for therapeutics that can evade mutation-based drug-resistance. However,

mutational coldspots with no known mutations have not been examined in the amino acid sequence and 3D structure of SARS-CoV2 M^{pro}. To identify these coldspots in SARS-CoV2 M^{pro}, we aggregated the circulating missense mutations reported in Global Initiative on Sharing All Influenza Data (GISAID) until November 2, 2020. This was approximately 11 months since the start of the COVID-19 outbreak, which should have provided enough time for the virus to accumulate some key mutations for survival [13,14]. The dataset contained 19,154 mutations (see supplementary Table S1 for the details) covering total of 282 out of 306 residue positions of SARS-CoV2 M^{pro} which are referred here as mutational hotspots (Fig. 1a and 1b). These hotspots showed a minimum of one mutation (Fig. 1a). In particular, the data (top 13 with >200 mutations) showed the following hotspot positions were the most frequently mutated: Gly15 (6,297 reported mutations), Leu89 (2,392), Gly71 (1,615), Lys90 (1,108), and Asp248 (744) (Fig. 1b). The remaining 24 positions had no reported mutations and were considered mutational coldspots (Fig. 1c), as they have shown a degree of mutation resistance up to this stage of the pandemic. Although the 24 mutational coldspots are shortlisted from 306 residues, we studied structures of SARS-CoV2 M^{pro} to understand their structure-functional relevance.

Coldspots at the active/inhibitor site

To analyze the coldspots in and around the active site, we selected five 3D-structures with high resolution (Protein Data Bank (PDB) codes: 6LU7, 6Y2F, 6LZE, 6M0K, and 7BUY), out of several SARS-CoV2 M^{pro} structures, that had been co-crystalized with antiviral drug candidates in recently published studies [3,5,6,9]. However, the inhibitors were not optimal for SARS-CoV2 [15]. We believe the non-mutational residues (coldspots) could be appropriate target regions for designing effective inhibitors of SARS-CoV2 M^{pro}. In the SARS-CoV2 M^{pro} structures, domains I (8-101) and II (102-185) play major roles in the formation of the active

site and provide binding sites for inhibitors; while domain III (202-306) is important in the regulation of protease activity [3,9]. The catalytic dyad His41 and Cys145 is located at the active site that forms in a cleft between domain I and II. Most efforts to design anti-viral inhibitors using drug repurposing approaches are focused on targeting this active site [3,11]. We found 15 coldspots to be from domains I and II, and the remaining nine were in domain III (Fig. 1c). The inhibitor-binding sites in the five SARS-CoV2 M^{pro} structures were superimposed and are represented as a surface model in the 3D structures (Fig. 1d, 1e and 1f), which show that a total of 25 residues (Fig. 1g) form the binding sites for the reported inhibitors (6LU7-N3 (M^{pro}-inhibitor name), 6Y2F-13b, 6LZE-11a, 6M0K-11b, and 7BUY-carmofur). In these 25 positions, 22 were affected by a total of 525 mutations. In particular, residue positions 46, 49, 142, 190, and 191 showed more than 15 mutations each. This result suggests that most of the active site residues are mutated and challenging to target.

Interestingly, we mapped three coldspots, Leu141, Phe185, and Gln192, in the 6LU7-N3 complex (Fig. 1f). The structural importance of these coldspots was emphasized by the recent X-ray crystallographic studies of SARS-CoV2 M^{pro} [4,5,7] demonstrating the involvement of the spots in the formation of substrate-binding sites and Phe185 and Gln192 in the stability of the active site. We found coldspots Asn133 and Lys137 beneath the surface formed by the binding-site residues (Fig. 1e), specifically, Leu27, Asn119, and Gly146 are near the catalytic dyad (His41 and Cys145). They may provide some support to the catalytic center, as evidenced by a recent study, in which Leu27 was found to play a key role in the activity of the M^{pro} structure of SARS-CoV2 [8]. Whereas, Leu27 and Asn119 are involved in the formation of the binding site in SARS-CoV M^{pro} [16] (Table 1). However, based on our data analysis, the other pocket-forming residues in the structures undergo mutations, which may modify the shape of the binding pocket. This prediction is supported by a recent study [17], in which the structures of the

mutants Met49Ile, Pro184Leu/Ser, and Ala191Val were shown to substantially deviate from the wildtype. Thus, the residues were assumed to belong within the mobile regions of the active site, which control the conformational changes that may be required for catalysis. This indicates that coldspots are required at the active site to maintain effective targeting.

Most importantly, in SARS-CoV2 M^{pro}, the key active site residues His41 (3 mutations), Phe140 (1 mutation), Cys145 (3 mutations), Glu166 (3 mutations), and His172 (1 mutation) showed low mutation frequencies (a total of eleven out of 525 mutations at the active site) (Fig. 1g). This suggests that the residues involved in critical functions at the active site are mutated less frequently than other residues.

Coldspots at the dimer interface

An alternate therapeutic strategy is to design antiviral agents to target the dimerization of the SARS-CoV M^{pro}, as the dimeric form is essential for activity [18,19] and, with 98% identity, is also applicable to SARS-CoV2 M^{pro} [4,7]. Here, we examined the functional relevance of coldspots on the surface of the dimer interface in SARS-CoV2 (PDB code: 6LU7), as they could provide mutation-resistant drug and vaccine target sites (Fig. 2a). Half of the 24 coldspot positions are on the surface of the protease (Fig. 2a-2b), and the rest are buried. We discovered seven cold spot positions (Gly2, Arg4, Tyr126, Lys137, Leu141, Leu286, and Leu287) on the surface that are involved in the formation of the dimer interface in the SARS-CoV2 M^{pro} (Fig. 2c and 2d). They form two sites: the first is based on the positions Gly2, Arg4, Tyr126, Lys137, and Leu141 (Fig. 2c), and the second site includes the positions Arg4, Lys137, Leu286, and Leu287 (Fig. 2d). In the SARS-CoV M^{pro}, these sites include several key interactions, Arg4-Lys137-Glu290 [20], Gly2-Arg4-Tyr126 [21], Ser284-Tyr285-L286 [22], and Ser1-Glu166-His163-

His172 [23,24], that have been experimentally proven to be vital for maintaining the dimer interface and the active site (Table 1).

In SARS-CoV2 M^{pro}, we observed a hydrogen bond between Arg4 and Lys137 (Fig.2c). As both are coldspots (with three other coldspots nearby, Gly2, Tyr126, and Leu141), this appears to be a potential site for inhibition. It also appears slightly similar to the one recently proposed as a potential allosteric site in SARS-CoV2 M^{pro} [4]. Residue 141 plays a dual role by forming the active site and dimer interface with Val303 (Fig.1e, 1f and 2c). Leu286 forms two hydrogen bonds with Ser284 in protomer A and hydrophobic interactions with Tyr280, Gly283, and Ala285 in protomer B. Moreover, eight other coldspot residues (Asn119, Asn133, Tyr154, Phe185, Gln192, Gln256, Gly258, and Asp295) on the protease surface (Fig. 2a and 2b) and do not contribute to the dimer interface.

The other structures of SARS-CoV2 M^{pro} also confirms the functional relevance of the coldspot residues Gly2, Arg4, Tyr126, Lys137, Leu141, and Leu286 that are directly involved in dimer formation through various interactions [7,9] (Table 1). A recent electrophilic screening of 1,250 fragments provided three hits (Z1849009686, Z264347221, and POB0073) that bind to the dimer interface, and it was suggested these fragments might be used as quasi-allosteric inhibitors to disrupt the dimerization of the active M^{pro} [4]. The coldspots Gly2 and Leu141 are involved in the reported binding sites for fragment Z264347221, while Arg4 and Lys137 are involved in POB0073 binding. Furthermore, Glu286 is conserved at the interface of M^{pro} in Porcine epidemic diarrhea virus (PEDV), Transmissible gastroenteritis virus, Infectious bronchitis virus (TGEV), and Human coronavirus 229E (HCOV-229E) [25-27]. We noted that Leu286 makes vital hydrophobic contacts at the dimer interface in SARS-CoV2 M^{pro}. Tyr126

is another critical residue for dimer formation of M^{pro} in SARS-CoV, PEDV [27], TGEV [25], and HCoV-229E [26] and, together with 141, is involved in the regulation of catalytic activity [21,28]. These correlate with our hypothesis that the observed coldspots may serve as mutation-resistant allosteric sites.

There are 21 hotspots at the interface covering 296 mutations; out of 21, only 10 hotspots had more than eight mutations (Fig. 2e). The frequency of mutations was very high at residue positions Gly283, Ala285, and Arg298 (65, 51, and 52 mutations, respectively), compared with the hotspots at the N-finger (residue 1-8) region. Mutational frequency (296) of surface residues at the dimer interface is relatively less than the active site residues (525 mutations) (Fig.1g). This indicates that the dimer interface at SARS-CoV M^{pro} seems to be more resistant to mutations.

Coldspots conserved among coronaviruses

Furthering our understanding of those residues conserved among coronaviruses (CoVs) might provide optimum target regions for designing improved therapeutic agents [29]. We compared the coldspots of SARS-CoV2 with other CoVs to analyze their degree of conservation. The structure-based sequence alignment of 12 different CoVs shows that the majority of the coldspots are arranged as three clusters: four coldspots at the N-terminal, six near the C-terminal and, surprisingly, nine near the active site in domain II (Fig. 3). We found 16 coldspots are distributed in eight conserved blocks including a block GxcGSvGxn based on motif GSCGS that is essential for the initiation of the catalysis in MERS-CoV and SARS-CoV [15,30]. Similarly, the other conserved blocks might have some functional role as they are found in the key structural regions. Interestingly, 14 out of the 24 coldspots were conserved among all the

CoVs. Moreover, most of the 14 conserved coldspots of SARS-CoV2 M^{pro} have critical roles in the formation of the active site (Leu27 and Gln192) and at the dimer interface (Gly2, Cys16, Lys137, Leu287, and Asp295), and Leu141 has both roles (Table 1, Fig. 2 & Fig. 3). The significance of the other conserved coldspots (Asn133, Gly146, Asn203, Phe219, Asn231, and Gly258) in SARS-CoV2 M^{pro} are unclear. Overall, this sequence alignment suggests that not all the highly conserved residues in SARS-CoV2 M^{pro} are resistant to mutation. Although only certain coldspots are conserved among CoVs, most of the conserved sites contribute to the formation of critical interactions (Table 1).

Biological relevance

It is understood that the SARS-CoV2 M^{pro} is undergoing or accumulating mutations at many hotspots, thus it is essential to identify consistent mutational coldspots that can be targeted with antiviral drugs. In addition, the data of nearly 20,000 global mutations used in this study were collected at the end of the first wave of COVID-19, are minimal. However, the identified mutational coldspots have biological relevance, according to the high-resolution X-ray structures of SARS-CoV2, sequence conservation among CoVs, and experimental evidence provided by the published X-ray structures of other CoV proteases [18,22,23,25-28,31,32] (Table 1). We now have a short list of promising targets that might be considered before embarking on time-consuming translational research underlying antiviral design.

The observed mutational frequencies in the hotspots at the active site and dimer interface indicate that the virus may be developing protective strategies against inhibitors. This correlates with the findings described in recent reports, which show the positions are changing the shape of the sites via mutations and plasticity [17,33]. However, coldspots that are identified here might

be good areas to target. Although some of these coldspots may convert to hotspots in the future, the frequency of the new mutations is likely to be minimal, as the data in this study show the sites play critical structural roles and are mutation-resistant. This is evident from their ability to avoid mutations over 11 months since the virus was first detected. However, further research is warranted for a deeper understanding of the phenomenon.

In conclusion, mutations in SARS-CoV2 M^{pro} delay the structure-based design of antivirals. This study highlighted the existence of various mutation-resistant coldspots and investigated their significance using evidence from structural studies. We have conducted initial filtering of the mutations of structural importance and reduced the data from 306 residues to 24 mutation coldspots. Finally, we pinpointed several conserved coldspots at the surface of the active site and dimer interface that could be optimum targets for the design of mutation-resistance antivirals.

Methods

Mutation dataset

The missense mutation dataset consisting of the reported circulating non-synonymous replacement mutations of SARS-CoV2 M^{pro} was retrieved from GISAID[10] by searching (CoVsurver) the database against the reference protein sequence Wuhan-Hu-1 (NC_045512.2, 10,055-10,977) with 306 amino acid positions. The dataset was collected until November 2, 2020. There were 19,154 mutations (see supplementary Table S1 for the details) at 282 residue positions, which were considered as mutational hotspots. The data were consolidated based on

the residue positions to calculate their mutational frequencies. This led to the discovery of 24 positions with no reported mutations at the time of collection, which were considered SARS-CoV2 M^{pro} mutational coldspots. The 24 coldspots were used for structural and sequence analyses.

Structural analyses

Recently published high-resolution dimeric (protomer A and B) X-ray 3D structures of SARS-CoV2 M^{pro} were obtained from the PDB (codes: 6LU7, 6Y2F, 6LZE, 6M0K, and 7BUY) [3,5,6,9], and the M^{pro} structures of other CoVs determined in previous studies were used for structural investigations. The 3D-locations of the hotspot and coldspot residues within key functional regions were mapped in the structures of SARS-CoV2 M^{pro}. To examine the substrate-binding sites, we used the structures of SARS-CoV2 M^{pro} and inhibitor co-crystals (6LU7-N3, 6Y2F-13b, 6LZE-11a, 6M0K-11b, and 7BUY-carmofur) and generated a pocket surface model with all pocket-forming residues. Pymol was used for structural analyses and to represent the molecular structures (The PyMOL Molecular Graphics System, Schrödinger, LLC.).

Sequence alignment

In the structure-based sequence alignment, to represent the coronavirus superfamily[34], we used 11 structural-sequences of M^{pro} from different families. The following M^{pro} sequences of CoVs with X-ray structures (PDB codes: 2HOB Severe acute respiratory syndrome coronavirus (SARS-CoV), 4YOI Tylonycteris bat coronavirus HKU4 (HKU4), 4ZUH Porcine epidemic diarrhea virus (PEDV), 2ZU2 Human coronavirus 229E (HCOV-229E), 4WME Middle East respiratory syndrome-related coronavirus (MERS), 6JIJ Murine hepatitis virus strain A59 (MURINE), 3D23 Human coronavirus HKU1 (isolate N1) (HKU1-N1), 6FV2 Human

coronavirus NL63 (HCOV-NL63), 4ZRO Feline infectious peritonitis virus (strain 79-1146) (FIPV), 2AMP Transmissible gastroenteritis virus, Infectious bronchitis virus (TGEV), 2Q6F Infectious bronchitis virus (IBV)) were used in a multiple sequence alignment against M^{pro} of SARS-CoV2 (6LU7) as a reference. We used Multalign for the multiple sequence alignment[35].

References

1 Liu, C. *et al.* Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Central Science*, doi:10.1021/acscentsci.0c00272 (2020).

2 Luan, B., Huynh, T., Cheng, X., Lan, G. & Wang, H. R. Targeting Proteases for Treating COVID-19. *J Proteome Res* **19**, 4316-4326, doi:10.1021/acs.jproteome.0c00430 (2020).

3 Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science (New York, N.Y.)* **368**, 1331-1335, doi:10.1126/science.abb4489 (2020).

4 Douangamath, A. *et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nature communications* **11**, 5047, doi:10.1038/s41467-020-18709-w (2020).

5 Jin, Z. *et al.* Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293, doi:10.1038/s41586-020-2223-y (2020).

6 Jin, Z. *et al.* Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat Struct Mol Biol* **27**, 529-532, doi:10.1038/s41594-020-0440-6 (2020).

7 Kneller, D. W. *et al.* Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *The Journal of biological chemistry*, doi:10.1074/jbc.AC120.016154 (2020).

8 Rut, W. *et al.* SARS-CoV-2 M(pro) inhibitors and activity-based probes for patient-sample imaging. *Nature chemical biology*, doi:10.1038/s41589-020-00689-z (2020).

9 Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science (New York, N.Y.)* **368**, 409-412, doi:10.1126/science.abb3405 (2020).

10 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **22**, doi:10.2807/1560-7917.ES.2017.22.13.30494 (2017).

11 Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic & medicinal chemistry letters* **30**, 127377, doi:10.1016/j.bmcl.2020.127377 (2020).

12 Pachetti, M. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of translational medicine* **18**, 179, doi:10.1186/s12967-020-02344-6 (2020).

13 Badua, C., Baldo, K. A. T. & Medina, P. M. B. Genomic and proteomic mutation landscapes of SARS-CoV-2. *Journal of medical virology*, doi:10.1002/jmv.26548 (2020).

14 Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827 e819, doi:10.1016/j.cell.2020.06.043 (2020).

15 Ionescu, M. I. An Overview of the Crystallized Structures of the SARS-CoV-2. *Protein J*, doi:10.1007/s10930-020-09933-w (2020).

16 Xue, X. *et al.* Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *Journal of virology* **82**, 2515-2527, doi:10.1128/JVI.02114-07 (2008).

17 Sheik Amamuddy, O., Verkhivker, G. M. & Tastan Bishop, O. Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 M(pro). *Journal*

of chemical information and modeling **60**, 5080-5102, doi:10.1021/acs.jcim.0c00634 (2020).
18 Hsu, W. C. *et al.* Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. *The Journal of biological chemistry* **280**, 22741-22748, doi:10.1074/jbc.M502556200 (2005).
19 Goyal, B. & Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS combinatorial science* **22**, 297-305, doi:10.1021/acscombsci.0c00058 (2020).
20 Ghosh, A. K. *et al.* Design and synthesis of peptidomimetic severe acute respiratory syndrome chymotrypsin-like protease inhibitors. *Journal of medicinal chemistry* **48**, 6767-6771, doi:10.1021/jm050548m (2005).
21 Wei, P. *et al.* The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochemical and biophysical research communications* **339**, 865-872, doi:10.1016/j.bbrc.2005.11.102 (2006).
22 Lim, L., Shi, J., Mu, Y. & Song, J. Dynamically-driven enhancement of the catalytic machinery of the SARS 3C-like protease by the S284-T285-I286/A mutations on the extra domain. *PloS one* **9**, e101941, doi:10.1371/journal.pone.0101941 (2014).
23 Xue, X. *et al.* Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *Journal of molecular biology* **366**, 965-975, doi:10.1016/j.jmb.2006.11.073 (2007).
24 Tan, J. *et al.* pH-dependent conformational flexibility of the SARS-CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *Journal of molecular biology* **354**, 25-40, doi:10.1016/j.jmb.2005.09.012 (2005).
25 Anand, K. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO journal* **21**, 3213-3224, doi:10.1093/emboj/cdf327 (2002).
26 Lee, C. C. *et al.* Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. *The Journal of biological chemistry* **284**, 7646-7655, doi:10.1074/jbc.M807947200 (2009).
27 Ye, G. *et al.* Structural basis for the dimerization and substrate recognition specificity of porcine epidemic diarrhea virus 3C-like protease. *Virology* **494**, 225-235, doi:10.1016/j.virol.2016.04.018 (2016).
28 Shi, J., Sivaraman, J. & Song, J. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. *Journal of virology* **82**, 4620-4629, doi:10.1128/jvi.02680-07 (2008).
29 Naqvi, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et biophysica acta. Molecular basis of disease* **1866**, 165878, doi:10.1016/j.bbadis.2020.165878 (2020).
30 Wang, H. *et al.* Comprehensive Insights into the Catalytic Mechanism of Middle East Respiratory Syndrome 3C-Like Protease and Severe Acute Respiratory Syndrome 3C-Like Protease. *ACS Catal* **10**, 5871-5890, doi:10.1021/acscatal.0c00110 (2020).
31 Chen, S. *et al.* Residues on the dimer interface of SARS coronavirus 3C-like protease: dimer stability characterization and enzyme catalytic activity analysis. *Journal of biochemistry* **143**, 525-536, doi:10.1093/jb/mvm246 (2008).
32 Tomar, S. *et al.* Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CLpro): IMPLICATIONS FOR nsp5

REGULATION AND THE DEVELOPMENT OF ANTIVIRALS. *The Journal of biological chemistry* **290**, 19403-19422, doi:10.1074/jbc.M115.651463 (2015).
33 Kneller, D. W. *et al.* Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room temperature X-ray crystallography. *Nature communications* **11**, 3202, doi:10.1038/s41467-020-16954-7 (2020).
34 Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of medical virology* **92**, 418-423, doi:10.1002/jmv.25681 (2020).
35 Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881-10890, doi:10.1093/nar/16.22.10881 (1988).
36 Rathnayake, A. D. *et al.* 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV-infected mice. *Science translational medicine* **12**, doi:10.1126/scitranslmed.abc5332 (2020).
37 Shi, J. & Song, J. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *The FEBS journal* **273**, 1035-1045, doi:10.1111/j.1742-4658.2006.05130.x (2006).

Acknowledgement

We acknowledge GISAID for disseminating SARS-CoV2 data. We would like to thank all the communities worldwide involved and supported in the response to the COVID- 19 pandemic.

Competing interests

The authors declare no competing interests.

353 Table

354 Table 1. Structural and functional importance of mutational coldspots in SARS-CoV2

S. no	Coldspot residue position	Structural and/or functional role	Publication	Experimental method
1	2, 4, 137 141	Gly2, Arg4 and Lys137 in the binding pocket at the dimer interface of SARS-CoV2 M^{pro} for the inhibitor POB0073. 141 is involved in the formation of the binding site at the dimer interface of SARS-CoV2 M^{pro} for the fragment/inhibitor Z264347221	Douangamath et al.,[4]	Crystallographic and electrophilic fragment screening
2	16	Cys16 Forms hydrogen bonds with Ser10, Gly11, and Glu14 in SARS-CoV2 M^{pro} at the dimer interface	Kneller et al.,[7]	Neutron X-ray crystallography
3	27	Leu27 is involved in the activity of SARS-CoV2 M^{pro}. It interacts with active binding probe, Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS (B-QS1-VS)	Rut et al.,[8]	Synthesis and X-ray crystallography
4	192, 185, 141, 2, 4, 126, 137, 141, 286	Leu141, Phe185, Gln192 are involved in formation of inhibitor N3-binding pocket, and Gly2, Arg4, Tyr126, Lys137, Leu141, Leu286, and Leu287 function in the dimerization of SARS-CoV2 M^{pro}	Jin et al., [5]	X-ray crystallography
5	286	In SARS-CoV2 M^{pro}, Ala285 and Leu286 induced slightly closer packing at the dimer interface and increased catalytic activity (compared to SARS-CoV M^{pro})	Zhang et al.,[9]	X-ray crystallography
6	192	In MERS-CoV infected mice, residue 192 in M^{pro} interacts with compound 6h and has a potential specificity role in selection of inhibitors 7j and 7i	Rathnayake et al.,[36]	Therapeutic treatment in a mouse model, X-ray crystallography
7	185, 192	In SARS-CoV2 M^{pro}, a long loop Phe185-194 is stabilized by the interactions of 189-192-186. Phe185 and Leu167 form a deep hydrophobic pocket for substrate binding	Kneller et al.,[7]	Neutron X-ray crystallography
8	2, 4, 16, 119, 137, 141	In SARS-CoV2 M^{pro} at the dimer interface, N-terminal residues 1-16 (N-finger) interact with 118 (in protomer B)-125 (B) and loop 137 (B) -142 (B)	Kneller et al.,[7]	Neutron X-ray crystallography
9	4, 27, 141, 185, 295	Glu185 (in PEDV M^{pro}) and Val185 (in SARS-CoV M^{pro}) located in motif 3 of the substrate binding pocket. In PEDV M^{pro}, Leu27 and Asn141 are two key residues involved in enzyme catalytic activity. Arg4Ala and Gln295Ala mutations at the dimer interface reduced the catalytic activity	Ye et al.,[27]	X-ray crystallography, size-exclusion chromatography, and ultra-centrifugation
10	128, 137, 185,	In MERS CoV M^{pro}, Lys137 and Glu290 form hydrophobic contacts with 185, and 128 forms van der Waals interactions with His8-Lys155. These residues are involved in control of the dimer interface	Tomar et al.,[32]	X-ray crystallography
11	286	In SARS-CoV M^{pro}, Ser284-Thr285-Ile286/Ala mutations enhanced the catalytic machinery by enzyme dynamics	Lim et al.,[22]	X-ray crystallography
12	27, 119	Leu27 and Asn119 are involved in the formation of substrate binding site of SARS-CoV M^{pro}	Xue et al.,[16]	X-ray crystallography
13	4	In SARS-CoV M^{pro}, salt bridge between Arg4-Glu290 stabilizes the dimer	Shi et al.,[28]	X-ray crystallography
14	126	In SARS-CoV M^{pro}, Tyr126 is essential for dimer stability and for substrate catalytic machinery via aromatic and hydrophobic interactions between Tyr126 and Met6 and an aromatic interaction between Tyr126 with Phe140	Shi et al.,[28]	X-ray crystallography
15	126, 141	In SARS-CoV M^{pro}, 141-140-139 and 126 are directly involved in the dimer formation and regulation, which are required for catalysis	Shi et al.,[28]	X-ray crystallography, mutagenesis, and ultra-centrifugation
16	141, 185	In IBV M^{pro}, Gly141 and Asp185 form a substrate binding pocket for the inhibitor N3	Xue et al.,[16] 2008	X-ray crystallography
17	141, 192, 2, 4	Leu141 and Gln192 are in the active site of SARS-CoV M^{pro} interact with authentic N-terminal residues (Gly2, Arg4, Cys16), enhancing the activity	Xue et al.,[23]	X-ray crystallography

18	4, 126	In SARS-CoV M^{pro}, Arg4Glu mutation produces weak dimer with no activity. Tyr126-M6 hydrophobic interaction stabilizes the dimer conformation	Wei et al.,[21]	Mutagenesis, enzyme assay, and analytical ultracentrifugation
19	286	In SARS-CoV M^{pro}, Ile286 with other critical dimerization residues form a channel to the catalytic center, which may play a role in regulating catalytic machinery	Shi and Song,[37]	Mutagenesis, dynamic light scattering, CD and NMR spectroscopy
20	4, 137	In SARS-CoV M^{pro}, Arg4 forms dimer interface with Lys137-Gln127 and Glu290	Ghosh et al.,[20]	X-ray crystallography
21	2, 4	N-terminal (residues 1-4) truncation of SARS-CoV M^{pro} affects dimer and enzymatic activity	Hsu et al., [18]	Mutagenesis, enzyme assay, and analytical ultracentrifugation
22	2, 4, 126, 286, 295	Gly2, Arg4, Gly126, Glu286, and Gln295 are conserved and play key roles at the dimer interface in the M^{pro} structures of PEDV, TGEV, and HCoV-229E	Ye et al.,[27]; Anand et al.,[25]; Lee et al.,[26]	X-ray crystallography, size-exclusion chromatography, and ultra-centrifugation

SARS-CoV: Severe acute respiratory syndrome coronavirus, HKU4: Tylonycteris bat coronavirus HKU4, PEDV: Porcine epidemic diarrhea virus, HCOV-229E: Human coronavirus 229E, MERS: Middle East respiratory syndrome-related coronavirus, TGEV: Transmissible gastroenteritis virus, IBV: Infectious bronchitis virus

Figures

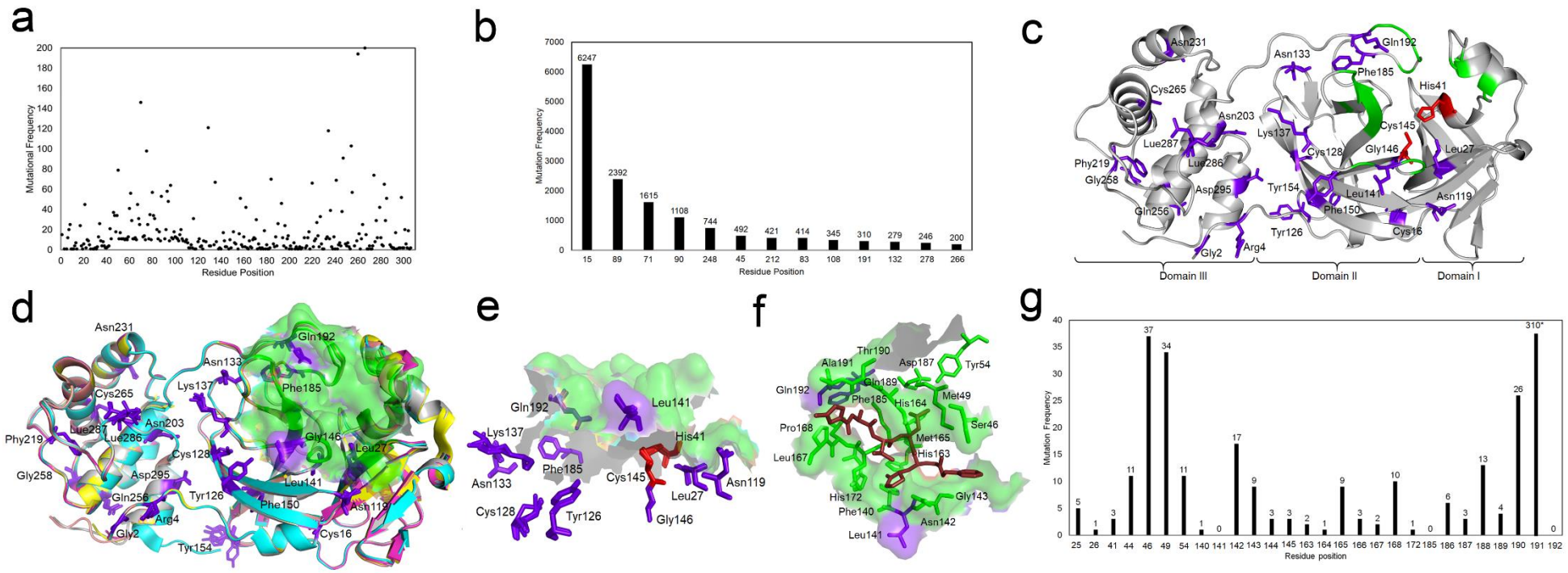

Figure 1: Hotspots and coldspots in SARS-Cov2 M^{pro}. (a) Residues with less than 200 mutations and (b) residues with more than 200 mutations plotted against mutation frequency. (c) Structural mapping of coldspots (PDB code: 6LU7); (d) Superimposition of high-resolution structures of SARS-CoV2 M^{pro} : PDB codes are 6LU7 (grey), 6Y2F (cyan), 6LZE (magenta), 6M0K (yellow), 7BUY (salmon); (e) Coldspots in and around the active sites of the superimposed structures of SARS-CoV2 M^{pro} ; (f) Active site pocket in 6LU7 with inhibitor N3 (ruby). (g) Mutational frequency of active site residues. 191 Coldspots: purple sticks; active sites: green surface/cartoon; catalytic dyads: red sticks.

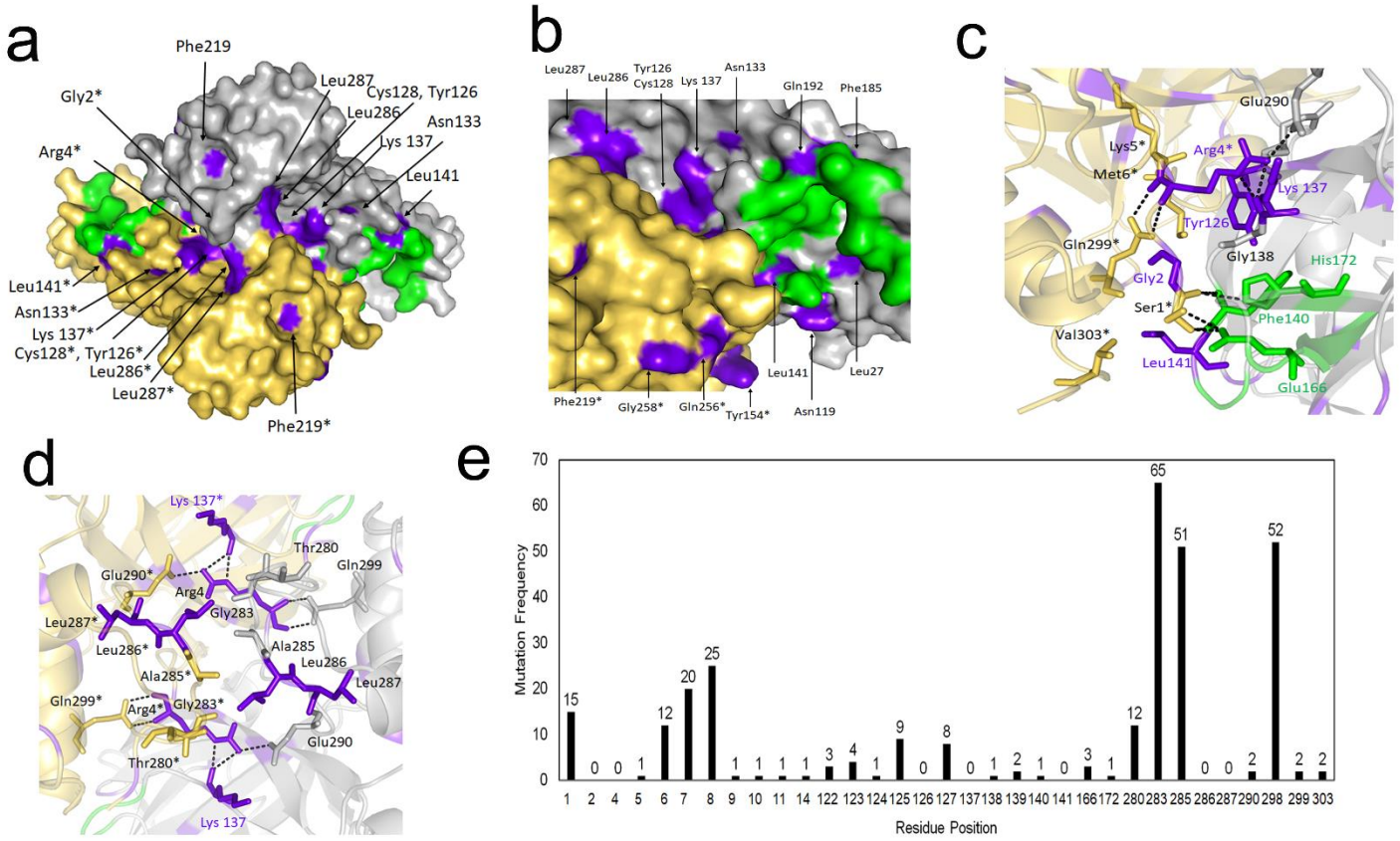

Figure 2: Dimer interface of SARS-CoV2 M^{pro}. (a) Surface model of the dimer (PDB code: 6LU7); (b) Extended and detailed view of panel a showing the dimer interface, with coldspot residues on the surface model; (c) Site1 and (d) Site2 with coldspot residues and interactions involved in dimerization; (e) Mutation frequency of dimer interface residues. Grey cartoon/surface: protomer A; Yellow surface/cartoon: protomer B; residues with * are from protomer B; purple: coldspot residues; green: active site; dashes: hydrogen bonds.

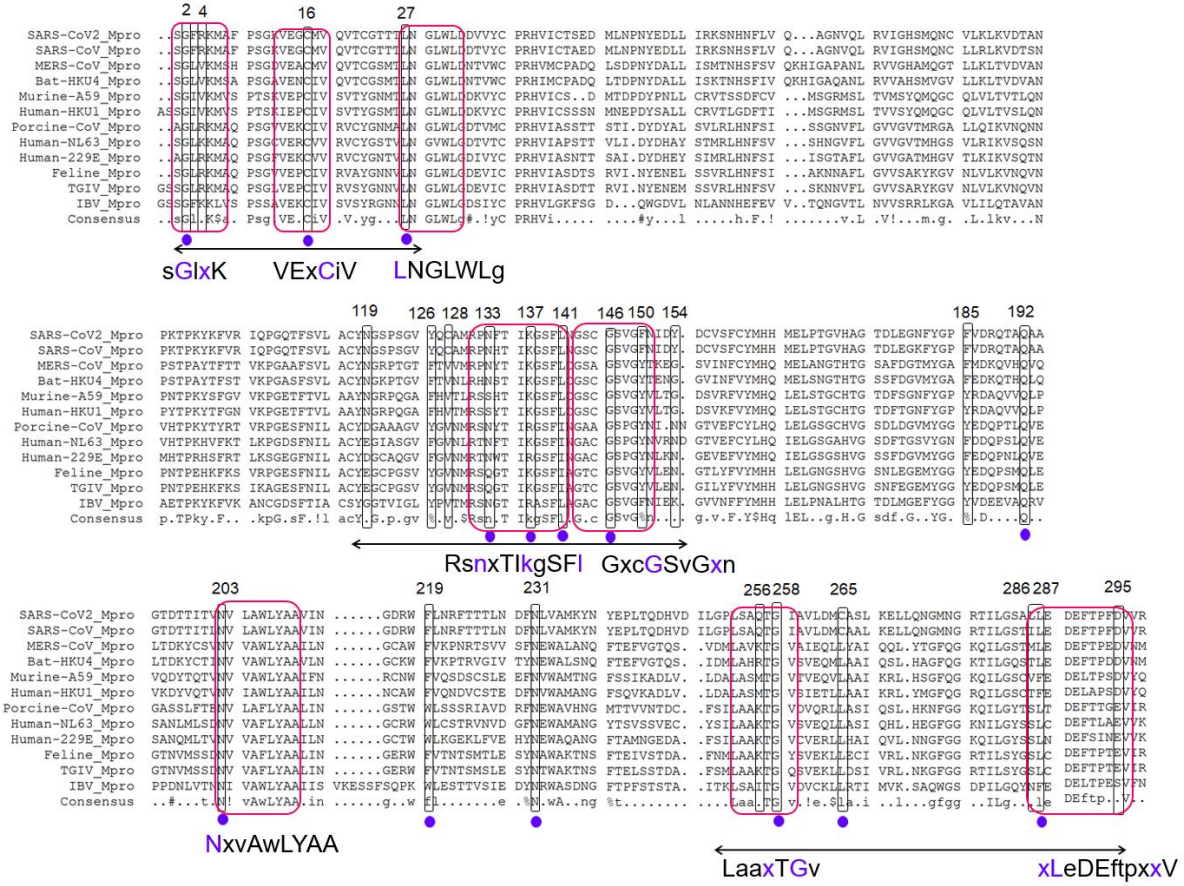

Figure 3: Multiple sequence alignment of M^{pro} from various coronaviruses. The mutational coldspots of SARS-CoV2 M^{pro} are shown in boxed regions and red boxed regions are conserved patterns. The purple dots indicate conserved coldspots among all the aligned coronavirus M^{pro}s, and double arrows indicate the coldspot clusters.