

---

# The structural landscape of SARS-CoV-2 main protease: hints for inhibitor search.

Luigi Leonardo Palese<sup>1</sup>,

**1 Department of Basic Medical Sciences, Neurosciences and Sense Organs  
P.zza G. Cesare 11, 70124, Bari, Italy  
University of Bari "Aldo Moro"**

\* [luigileonardo.palese@uniba.it](mailto:luigileonardo.palese@uniba.it)

## Abstract

In 2019, an outbreak occurred which resulted in a global pandemic. The causative agent of this serious global health threat was a coronavirus similar to the agent of SARS, referred to as SARS-CoV-2. In this work an analysis of the available structures of the SARS-CoV-2 main protease has been performed. From a data set of crystallographic structures the dynamics of the protease has been obtained. Furthermore, a comparative analysis of the structures of SARS-CoV-2 with those of the main protease of the coronavirus responsible of SARS (SARS-CoV) was carried out. The results of these studies suggest that, although main proteases of SARS-CoV and SARS-CoV-2 are similar at the backbone level, some plasticity at the substrate binding site can be observed. The consequences of these structural aspects on the search for effective inhibitors of these enzymes are discussed, with a focus on already known compounds. The results obtained show that compounds containing an oxirane ring could be considered as inhibitors of the main protease of SARS-CoV-2.

## Introduction

In late 2019 a new pneumonia illness was first reported in Wuhan, China [51]. It has rapidly spread over the world as pandemic threat with thousands of infected and deaths; it has been named COVID-19 by the World Health Organization. The causative agent of this pathology is a new betacoronavirus, related to the SARS coronavirus (SARS-CoV), designated as SARS-CoV-2 [16]. Thanks to the efforts of several structural biologists, a significant number of crystallographic structures of the proteins of this virus have begun to be present in the Protein Data Bank (PDB) [6].

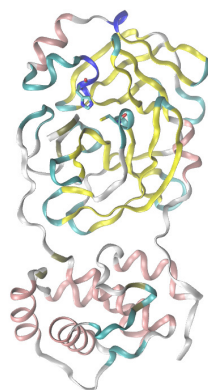
Crystal structures provide us with knowledge at atomic level of proteins or protein complexes. Although these static structures are extremely interesting and rich in information, proteins are dynamic entities and understanding their functioning mechanism requires the knowledge of their dynamics. To determine and analyze the dynamic behavior of proteins, a series of techniques have been developed, largely computational ones. Recently a technique for the reconstruction of the dynamics based on the presence of redundant structures of the same protein in the PDB has been developed [37]. These redundant sets of structures are an extremely interesting tool to gain insight into the protein dynamics, and previously this technique has been applied to the HIV protease [36,37].

Coronaviruses [12,23,31] have the largest known single-stranded, positive-sense RNA genomes (ranging from 25.5 to 32 kb). Several strains of coronaviruses are involved in pathological conditions in humans: particularly strains 229E, NL63, OC43, HKU1, MERS-CoV (Middle East Respiratory Syndrome), SARS-Cov (Severe Acute Respiratory Syndrome) and the, above mentioned, recently appeared, SARS-related SARS-CoV-2. Structural proteins of coronaviruses are the spike (S) protein,

membrane protein (M), envelop (E) protein and the nucleocapsid (N) protein. Some viruses contain also other structural proteins, such as the hemagglutinin esterase in beta-CoVs. The RNA genome contains several genes, whose order is generally preserved, coding for different proteins: PP1a, PP1ab, S, E, M, N. Two-third of the RNA genome is covered by the *ORF1a* and *ORF1b*, which produce two polyproteins, PP1a and PP1ab, whose processing leads to the formation of sixteen non-structural proteins (NSPs). NSPs participate in different viral functions, including the replicase-transcriptase complex.

Two cysteine proteases in CoVs act on these polyproteins to release the NSPs [1, 2, 18, 41, 52, 58]. One is the papain-like protease ( $PL^{pro}$ ), which performs three cleavage reaction. The other protease is a chymotrypsin-like cysteine protease, known as main protease ( $M^{pro}$ ) or 3C-like protease ( $3CL^{pro}$ ) because of its similarity to the picornavirus 3C protease.  $M^{pro}$  is responsible for the remaining 11 cuts leading to the formation of NSPs. The recognition sequence of  $M^{pro}$  X-(L/F/M)-Q | (G/A/S)-X (where X is any amino acid and | represents the cleavage site) is not recognized by any host protease, thus making this enzyme an optimal target for the search of possible inhibitors usable as drugs in the treatment of CoV infections.

The  $M^{pro}$  structure is remarkably similar in all CoVs [18]: the enzyme is a homodimer in which the N-terminus of one monomer participates to shape the substrate-specificity pocket and the oxyanion hole of the other monomer. Consequently dimerization is a requisite for activity. Each monomer



**Figure 1. Structure of the  $M^{pro}$  monomer.** Colors refer to the secondary structure motifs:  $\alpha$ -helix in pink,  $3_{10}$ -helix in blue, extended beta in yellow, turns in cyan. Residues of the active site dyad are reported in licorice. Atomic coordinates refer to the PDB entry 6LU7.

consists of two domains, I (residues 8-101 in 6LU7 [21]) and II (residues 102-184), which feature a chymotrypsin-like fold and harbor the catalytic dyad Cys-His (see Figure 1). A further  $\alpha$ -helical domain (domain III, residues 201-303) is connected by a long loop to domain II and is involved in the dimerization. Due to the global pandemic of COVID-19, in a short time numerous structures of this enzyme were deposited in the PDB; the analysis of these structures is the main topic of this communication.

## Materials and Methods

The analysis was conducted essentially as described [15, 36, 37]. Atomic coordinates of the SARS-CoV and SARS-Cov-2  $M^{pro}$ s were obtained from PDB [6]. The list is reported as Table 1 [20, 21, 24–26, 30, 33, 50, 55–57, 59]. A coarse-grained representation of the protein backbone was constructed considering the  $\alpha$ -carbon atoms. Multiple conformations of the protein backbone were removed, and only the most represented conformation was considered for the subsequent analysis. The correct number of  $\alpha$ -carbon atoms was set to 304 (residues 1-304) so as to include the greatest number of structures in the analysis. The pdb files were superposed by a Tcl script in a VMD [19] environment,

**Table 1. PDB IDs of SARS-CoV and SARS-CoV-2  $M^{pro}$ s.**

5R7Y	5R7Z	5R80	5R81	5R82	5R83	5R84	5RE4	5RE5	5RE6	5RE7
5RE8	5RE9	5REA	5REB	5REC	5RED	5REE	5REF	5REG	5REH	5REI
5REJ	5REK	5REL	5REM	5REN	5REO	5REP	5RER	5RES	5RET	5REU
5REV	5REW	5REX	5REY	5REZ	5RF0	5RF1	5RF2	5RF3	5RF4	5RF5
5RF6	5RF7	5RF8	5RF9	5RFB	5RFC	5RFD	5RFE	5RFF	5RFG	5RFH
5RFI	5RFJ	5RFK	5RFL	5RFM	5RFN	5RFO	5RFP	5RFQ	5RFR	5RFS
5RFT	5RFU	5RFV	5RFW	5RFX	5RFY	5RFZ	5RG0	6LU7	6M03	6W63
6Y84	2A5A	2DUC	2GX4	2GZ7	2GZ8	2GZ9	2H2Z	2HOB	2V6N	2Z3C
2Z3D	2Z3E	2Z94	2Z9G	2ZU4	2ZU5	3SNB	3SNC	3SND	3SNE	3SZN
3TIT	3TIU	3TNS	3TNT	3V3M						

as described [15, 36, 37]. The  $\alpha$ -carbon atom Cartesian coordinates were extracted from the updated pdb files and stored as a text file. This last file is the data matrix in which each row represents a  $M^{pro}$  structure in the database.

Since the number of components of the  $M^{pro}$  data sets is less than the number of degrees of freedom of the monomer, the exact calculation of the correlation matrix is not possible. Consequently, principal component analysis (PCA) was performed using the truncated SVD algorithm [17], which works even in the case of degenerate correlation matrices, as described in [15, 36, 37]. Principal modes of the protein dynamics obtained from PCA were visualized by means of the program NMWiz [5] in VMD. Random projections were performed as described [38]. Superposition of the enzymes belonging to different strains has been performed by protocols reported in [14, 42, 43].

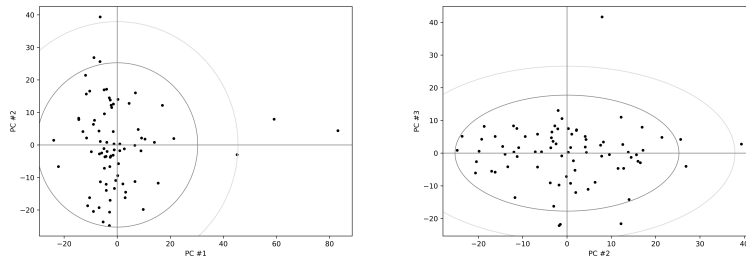
The atomic coordinates of proteins for which the crystallographic structure is not available were obtained from the SWISS-MODEL Repository [7].

Molecular docking was performed using the AutoDock Vina software; *pdbqt* files were obtained through the same software and / or through Open Babel toolbox [35, 48]. As a rule of thumb, binding affinity was considered significant only for values lower than -6 kcal / mol [45].

## Results and Discussion

### The SARS-CoV-2 main protease data set

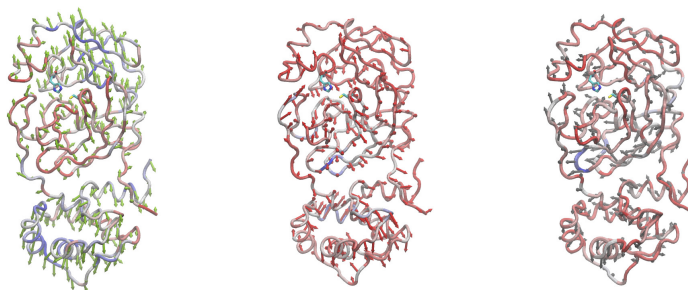
Results of PCA on the SARS-CoV-2  $M^{pro}$  data set are reported in Figure 2, left panel, which shows the second principal component (PC) *vs* the first PC, and the second *vs* third PC graph is reported in the right panel of the same Figure. These three components account for 24.9%, 17.3% and 8.5% of total variance respectively, and the first 25 PCs describe almost 91% of the total variance. This



**Figure 2. Principal component analysis.** Left panel reports the second *vs* first principal component, right panel reports the third *vs* second principal component of the  $M^{pro}$  data set of SARS-CoV-2. The 95% and 99% confidence ellipses are reported as dark and light gray respectively.

analysis shows that  $M^{pro}$  structures are distributed in a single cluster. Some outliers are detected by PCA: three structures are outside the 99% confidence level in the plane delimited by the first PC *vs* the second one, one structure is outside the 99% confidence level for the second PC *vs* the third PC plane. Similar results (i.e. a single cluster of structures and the same outliers) are obtained using the random projection method (not shown).

As for other proteins for which a sufficiently large number of redundant structures are available in the PDB [36,37], also for SARS-CoV-2  $M^{pro}$  it is possible to obtain indications on what are the possible coordinated and large scale motions of the protein. The first PC encodes a stretching - shortening motion along the major axis of the monomer. The structures characterized by extreme values of this PC are 6LU7 and 5RE4 in our data set. This mode affects also the loops at the active site entry. These movements can be described mainly as a shift of residues Asn 142, Ser 46 and Gln 189. However, this motion is such that there are no significant alterations of the distances between the most protruding residues: for example, the distance between the distal oxygen atoms of Asn 142 and Gln 189 is roughly 8 Å in 6LU7 and 9 Å in 5RE4. As expected, this PC involves a large part of the protein, and particularly residues 30-36, 61, 66, 89, 226-241, 265, 268 and 298, as shown in Figure 3, left panel. Residues close to the catalytic dyad, His 41 and Cys 145, are barely involved in this first



**Figure 3. Principal modes.** The backbone motion of SARS-CoV-2  $M^{pro}$  encoded by the first PC is reported in the left panel, the second PC in the middle panel and the third PC in the right panel. Most mobile regions are in blue, less ones in red. The catalytic dyad in is reported in licorice.

PC. It can be observed that in the 6LU7 structure, which is one the structures outside of the 99% confidence level in the positive side of the first PC, the active site appears deeper relative to that of structures characterized by lower (negative) values of this PC, such as 5RE4. These differences between the extremal structures are mainly due to movements affecting the orientation of the side chain of some residues, particularly Met 49 and Met 165 (see also below). The movement encoded by the second PC can be described as two parallel torsional motions affecting domain III on one hand, and collectively the chymotrypsin-like region (i.e. the domain I and II) on the other. This domain twisting takes place according to an axis perpendicular to the main one of the monomer, which corresponds to the main axis of the *individual* domains. Obviously, this second PC affects the various regions of the protein less extensively than the first one. Residues 109, 131, 182, 189, 245-249 and 282-287 are the most mobile in this second PC, as reported in Figure 3, middle panel. Among these it is worthy of mention Gln 189 which is part of the surface region of the protein that opens to the active site. Also in this case the residues of the catalytic dyad are not very mobile: indeed the His 41 turns out to be virtually immobile in this component. The third mode can be described as a twisting (torsion) centered on the minor axis of the  $\alpha$ -helix containing domain on one hand, and an analogue motion of the chymotrypsin-like domain along a parallel axis (again representing the minor axis of the domain). The  $\alpha$ -helix and the chymotrypsin-like domains slightly rotate relative to the other as a consequence of this motion. This mode, reported in Figure 3, right panel, is even more localized, and particularly affecting the residues 14, 31, 106 and 132-134. Also in this case the catalytic dyad shows a low mobility.

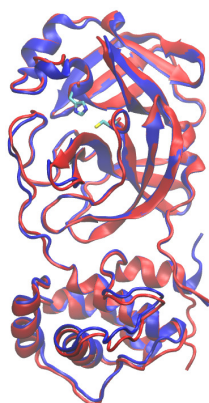
Since the search for specific inhibitors is a priority task, the distribution of the SARS-CoV-2  $M^{pro}$  structures in the landscape outlined by these PCs can be of help in choosing the structures

for docking calculations. This particularly considering the active site plasticity observed in these enzymes (see also below).

### Comparison of SARS-CoV and SARS-CoV-2 main proteases.

As the cuts made by  $M^{pro}$  are essential for the progression of the virus life cycle, this protease is obviously a possible target for COVID-19 therapy. Thence it will be important to evaluate if previously identified inhibitors of SARS-CoV  $M^{pro}$  could be re-proposed as possible drugs in the current pandemic threat. Although it should be stressed that currently no approved therapies based on the inhibition of  $M^{pro}$  are available, a significant number of compounds acting on this enzyme had been brought up to pre-clinical development for the treatment of the previous SARS (and MERS) pandemic [27, 29].

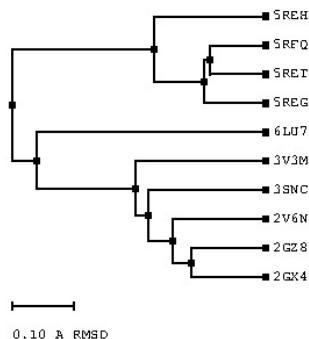
SARS-CoV and SARS-CoV-2 are closely related and their proteases display 96% sequence similarity. Therefore a significant structural similarity of the two proteins is expected. This can be appreciated by looking at the structural superposition between the  $M^{pro}$ s belonging to the two virus strains, as reported in Figure 4. Nearby the active site, there is only a single mutation between the two



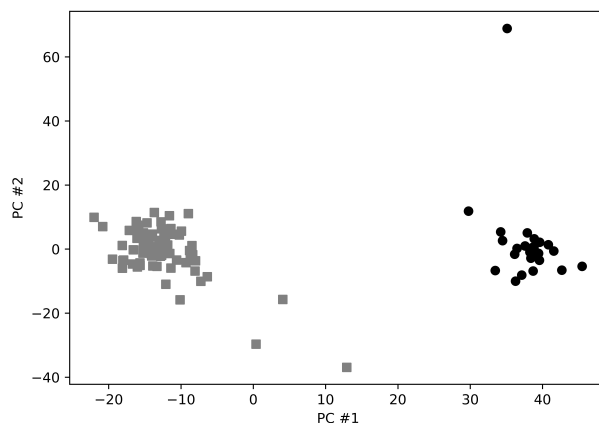
**Figure 4. Superposition of the SARS-CoV main proteases.** The figure reports SARS-CoV (pdb entry 2GX4, in red) and SARS-CoV-2 (pdb entry 5RET, in blue)  $M^{pro}$ s. The catalytic dyad is highlighted in licorice.

proteases (Ser 46 in SARS-CoV-2  $M^{pro}$  is Ala 46 in SARS-CoV). This residue is part of a loop that delimits the region of the active site, but is far from the residues of the catalytic dyad and from the region that binds the substrate more directly, so it is unlikely that it could determine significant changes in the activity or specificity of the two enzymes.

As further demonstration of the similarity of these two enzymes, Figure 5 shows the root-mean-square deviation of atomic positions (RMSD) tree concerning a series of  $M^{pro}$  structures of these strains. The Figure shows only few available PDB entries for clarity, which have been chosen to sample structures across the whole landscape of the first two PCs of both enzymes (the PC #2 *vs* PC #1 landscape for the available SARS-CoV  $M^{pro}$ , has been calculated as reported above for the SARS-Cov-2 enzyme, not shown). As can be appreciated from the above mentioned Figure 5, some structures of SARS-CoV-2 (6LU7 in this case) are more globally similar to structures of SARS-CoV than to those of their own strain. To further assess similarities and differences between the  $M^{pro}$  enzymes, a global PCA was performed on all the available structures belonging to both strains, SARS-CoV and SARS-CoV-2, which is reported in Figure 6. This analysis clearly shows that the two enzymes are linearly separable along a single PC. However, some outliers of the SARS-CoV-2  $M^{pro}$  (for example 6LU7) are closer to the centroid of the other strain than to that of their own group (and this justifies what is observed using RMSD, Figure 5). Anyway, this investigation shows that there are some differences in the global structures, i.e. at the backbone level, between proteases belonging

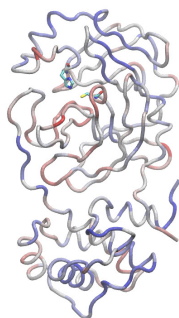


**Figure 5. RMSD tree of some main proteases.** Only few structures have been included in the calculation of the tree. The structures belong to the SARS-CoV  $M^{pro}$  (2GX4, 2GZ8, 2V6N, 3SNC and 3V3M) and to the SARS-CoV-2 enzyme (6LU7, 5REG, 5RET, 5RFQ and 5REH).



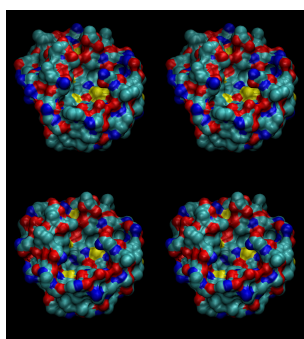
**Figure 6. Principal component analysis of the main proteases of SARS strains.** The figure reports the second *vs* first principal component of the  $M^{pro}$  data set of SARS-CoV (black circles) and SARS-CoV-2 (gray squares).

to the different strains. Figure 7 shows the protein backbone regions more involved in the differences between the two strains. These differences can be described as an elongation of the SARS-CoV enzyme along the major axis of the monomer relative to the SARS-CoV-2 one. As can be appreciated by inspecting the Figure, the most relevant differences are generally observed in residues distant from the enzyme active site. Nevertheless, there are two regions near the catalytic dyad that show a significant change in the backbone configuration between the two strains, namely residues 19-20-21 and 25-26-27. Recently, it has been suggested that there are significant differences in the topology of the two proteins in the region around Asn 142 (this residue is part of a loop that delimits the active site) and a marked flexibility and plasticity of the binding pocket of the SARS-CoV-2  $M^{pro}$  [11,13]. This analysis shows that these differences, as well as the binding pocket plasticity, reduce to side chain displacements, but generally with little involvement of the protein backbone. Figure 8 shows the neighborhood of the catalytic dyad in structures belonging to the two virus strains (2Z3E for SARS-CoV and 5RF4 for SARS-CoV-2), chosen in such a way as to be separated only along the first PC in the structural landscape reported in Figure 6, and such as to be representative of the relative centroids. Indeed this analysis shows that there are differences in depth and accessibility



**Figure 7. Backbone differences in the SARS main proteases.** The figure depicts the region in which the largest differences between SARS-CoV and SARS-CoV-2  $M^{pro}$ s can be observed. These correspond to the most involved regions in the first PC described in Figure 6. Blue regions correspond to the largest backbone displacements, red to the smallest ones. The residues of the catalytic dyad are reported in licorice.

of some residues between the two enzymes, but similar differences can be observed also between structures belonging to the same strain (not shown). Rotational isomerism of some side chains can be observed, particularly Met 165 and Asn 142, but not large displacements of the protein backbone. Although with a note of caution, the structural similarity between the two enzymes suggests that



**Figure 8. Active site of SARS main proteases.** The figure reports the stereo view of the active site surfaces of the  $M^{pro}$ s of SARS-CoV-2 (top image, pdb entry 5RF4). Bottom image refers to the active site surface of the pdb entry 2Z3E of the SARS-CoV.

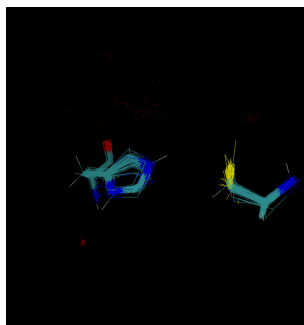
small molecules identified as potential inhibitors of SARS-CoV  $M^{pro}$  may be considered as potential drug candidates for the treatment of COVID-19 [11, 13]. The structural plasticity of this enzyme, however, requires consideration of other possibilities, even different from those already tried.

### The catalytic dyad: a good drug target?

The development of a drug is a long process, hence the need to resort to repurposing in order to quickly exploit drugs whose safety profile is known [11, 13]. Protease inhibitors designed for the treatment of HIV are considered for the SARS-CoV-2 infections, but it should be noted that these inhibitors act on a (very) different type of protease, namely an aspartyl protease characterized by a C2 symmetry of the active site [37]. Consequently these drugs are expected, *a priori*, to be poorly effective against CoV proteases.

Located in the shallow cleft between domains I and II, the active site of  $M^{pro}$  comprises a catalytic dyad consisting of the conserved residues His 41 and Cys 145. The sulfur atom of Cys 145 is located at 3.6 Å from the N- $\epsilon$  of His 41. Usually a water molecule at 3.2 - 3.3 Å from the N- $\delta$  of His 41 is visible

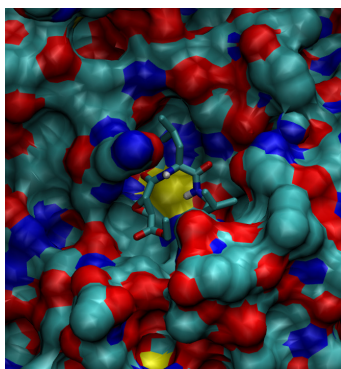
in crystals. This His-Cys-water arrangement is observed in most structures of SARS-CoV-2  $M^{pro}$ , as can be appreciated by inspection of Figure 9. The systematic presence of this water molecule could



**Figure 9. The catalytic dyad.** The figure shows the arrangement of the residues belonging to the SARS-CoV-2  $M^{pro}$  catalytic dyad, His 41 and Cys 145 in the available structures in the PDB. Red dots represent the nearby water molecules.

suggest that a catalytic *triad* is at work in these enzymes. Moreover, unlike the bonding pocket, these catalytic residues have an extremely conserved (and rigid) structural arrangement. A His-Cys dyad is frequently observed in cysteine proteases, including distant and completely unrelated enzymes. This arrangement can be observed in papain-like and 3C-like proteases, i.e. in all cysteine proteases active in the CoVs [18].

The spasmodic search for drugs capable of blocking the replication of CoVs could be facilitated by these similarities of cysteine proteases, and the possibility to exploit a series of inhibitors of these enzymes [49] can be considered. Molecular docking has been performed in order to assess the possibility that some of these substances are at least in principle effective in the SARS-CoV-2 control. These analyses have been performed using as protein target two structures of  $M^{pro}$ , namely 5RET and 6LU7, which were chosen as examples of structures close to the centroid of distribution in the PCA and outside the 99% confidence ellipse respectively. Cathepsin inhibitors, and various



**Figure 10. Aloxistatin binding site** The figure shows the aloxistatin (in licorice) at the active site of the SARS-CoV-2 main protease. The sulfur atom of cysteine 145 is visible near the oxirane ring of the ligand.

sulfur containing drugs (or even generally-recognized-as-safe substances) have been considered. For example, disulfiram [21, 27, 28, 39, 54] or also diallyl thiosulfinate (allicin) [8–10, 40] are *expected* to be broad spectrum inhibitors of these enzymes (see also [53]) and scaffolds for the development of more specific molecules. These molecules are extremely effective in the inhibition of cysteine proteases, including (but not limited to) the cathepsin-like protease histolysain [3, 4], whose active site is similar to that of  $M^{pro}$ , or caspases [34]. Molecular docking of diallyl thiosulfinate show that



this compound adapts perfectly to the active site, in an optimal configuration for the reaction with Cys 145. Binding affinity, as expected, is however too low to consider this substance as a specific inhibitor. Furthermore, it must be considered that this is an extremely reactive molecule, so the use of this substance as the only inhibitor can hardly lead to inhibition of growth of SARS-CoV-2 in real life. A similar argument also applies to disulfiram: it adapts perfectly to the active site of  $M^{pro}$ , in an optimal position for the reaction with Cys 145, but unfortunately the binding affinity, also in this case, is such that this cannot be considered as specific (not shown). In addition to these molecules, many others can theoretically bind the enzyme with high specificity. For example, among those with disulfiram-like effects that have been explored (this effect is often due to drug metabolites and not to the molecule itself), cefoperazone binds to the  $M^{pro}$  in docking experiments, as also shown in [32]. Here very significant results have been obtained regarding the binding of many cephalosporins to the active site of  $M^{pro}$ . Cefoperazone binds with remarkable affinity (-8.8 kcal/mol) to the enzyme (not shown). Unfortunately, unless specific drug delivery systems will be developed to allow high intracellular levels, these substances are unlikely to be effective in real life. Lopinavir binds to the active site with good affinity (up to -7.7 kcal/mol), and can probably act as a competitive inhibitor (not shown). The active site of this protease is however different from that of the HIV enzyme, so the effect is unlikely to be such as to trap the enzyme in some non-reactive conformation.

A molecule that could be promising as a protease inhibitor is aloxistatin (PubChem compound ID 65663). Aloxistatin (E64-d) is an irreversible and membrane permeable cysteine protease inhibitor initially developed for muscular dystrophy, which was brought up to phase 3 in the eighties of the last century [44, 46, 47]. It was administered in healthy adult volunteers without significant alterations in clinical parameters, but the development of this drug has discontinued. (<https://drugs.ncats.io/drug/L5W337AOUR>). It shows also antiviral effects [22]. Our results show that aloxistatin binds specifically to the active site of  $M^{pro}$  (-6.3 kcal/mol), with a distance between the sulfur atom of Cys 145 and a carbon atom of the oxirane ring of 3.7-3.8 Å (see Figure 10) both on 6LU7 and 5RET. Although this is a compound developed as a cathepsin inhibitor, it is interesting to note that its binding to the papain-like protease monomer of SARS-CoV-2 is less specific, and particularly it binds better in a region of the protein which does not coincide with the active site (these last calculations were carried out on the 6W9C structure; not shown).

In conclusion, the structural analysis on  $M^{pro}$  shows that some substances already known are able to specifically bind this enzyme in its various conformations. It is possible that these drugs, alone or in cocktails, may be effective in blocking the reproduction of SARS-CoV-2.

## References

1. K. Anand, G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra  $\alpha$ -helical domain. *EMBO J.*, 21(13):3213–3224, 2002.
2. K. Anand, J. Ziebuhr, P. Wadhvani, J. R. Mesters, and R. Hilgenfeld. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, 300(5626):1763–1767, 2003.
3. S. Ankri and D. Mirelman. Antimicrobial properties of allicin from garlic. *Microbes Infect.*, 1(2):125–129, 1999.
4. S. Ankri, T. Miron, A. Rabinkov, M. Wilchek, and D. Mirelman. Allicin from garlic strongly inhibits cysteine proteinases and cytopathic effects of *Entamoeba histolytica*. *Antimicrob. Agents Chemother.*, 41(10):2286–2288, 1997.
5. A. Bakan, L. M. Meireles, and I. Bahar. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*, 27(11):1575–1577, 2011.

- 
6. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucl. Acids Res.*, 28(1):235–242, 01 2000.
  7. S. Bienert, A. Waterhouse, T. A. de Beer, G. Tauriello, G. Studer, L. Bordoli, and T. Schwede. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.*, 45(D1):D313–D319, 2017.
  8. E. Block. The chemistry of garlic and onions. *Sci. Am.*, 252(3):114–121, 1985.
  9. E. Block. The organosulfur chemistry of the genus *Allium*—implications for the organic chemistry of sulfur. *Angew. Chem.*, 31(9):1135–1178, 1992.
  10. J. Borlinghaus, F. Albrecht, M. C. Gruhlke, I. D. Nwachukwu, and A. J. Slusarenko. Allicin: chemistry and biological properties. *Molecules*, 19(8):12591–12618, 2014.
  11. M. Bzowka, K. Mitusinska, A. Raczynska, A. Samol, J. A. Tuszyński, and A. Gora. Molecular dynamics simulations indicate the covid-19 mpro is not a viable target for small-molecule inhibitors design. *bioRxiv*, 2020.
  12. J. Cui, F. Li, and Z.-L. Shi. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.*, 17(3):181–192, 2019.
  13. A. Fischer, M. Sellner, S. Neranjan, M. A. Lill, and M. Smieško. Inhibitors for novel coronavirus protease identified by virtual screening of 687 million compounds. *ChemRxiv*, 2020.
  14. D. Frishman and P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579, 1995.
  15. A. Gnani, E. De Nitto, S. Scacco, L. Santacroce, and L. L. Palese. A new look at the structures of old sepsis actors by exploratory data analysis tools. *Antibiotics*, 8(4):225, 2019.
  16. A. Gorbalenya, S. Baker, R. Baric, and et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nat. Microbiol.*, pages 536–544, 2020.
  17. N. Halko, P.-G. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.*, 53(2):217–288, 2011.
  18. R. Hilgenfeld. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.*, 281(18):4085–4096, 2014.
  19. W. Humphrey, A. Dalke, K. Schulten, et al. VMD: visual molecular dynamics. *J. Mol. Graph.*, 14(1):33–38, 1996.
  20. J. Jacobs, V. Grum-Tokars, Y. Zhou, M. Turlington, S. A. Saldanha, P. Chase, A. Eggler, E. S. Dawson, Y. M. Baez-Santos, S. Tomar, et al. Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease. *J. Med. Chem.*, 56(2):534–546, 2013.
  21. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature*, 2020.
  22. J. C. Kim, R. A. Spence, P. F. Currier, X. Lu, and M. R. Denison. Coronavirus protein processing and RNA synthesis is inhibited by the cysteine proteinase inhibitor E64d. *Virology*, 208(1):1–8, 1995.

- 
23. M. M. Lai and D. Cavanagh. The molecular biology of coronaviruses. In *Adv. Virus Res.*, volume 48, pages 1–100. Elsevier, 1997.
  24. C.-C. Lee, C.-J. Kuo, M.-F. Hsu, P.-H. Liang, J.-M. Fang, J.-J. Shie, and A. H.-J. Wang. Structural basis of mercury-and zinc-conjugated complexes as SARS-CoV 3C-like protease inhibitors. *FEBS Lett.*, 581(28):5454–5458, 2007.
  25. C.-C. Lee, C.-J. Kuo, T.-P. Ko, M.-F. Hsu, Y.-C. Tsui, S.-C. Chang, S. Yang, S.-J. Chen, H.-C. Chen, M.-C. Hsu, et al. Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. *J. Biol. Chem.*, 284(12):7646–7655, 2009.
  26. T.-W. Lee, M. M. Cherney, C. Huitema, J. Liu, K. E. James, J. C. Powers, L. D. Eltis, and M. N. James. Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide. *J. Mol. Biol.*, 353(5):1137–1151, 2005.
  27. G. Li and E. De Clercq. Therapeutic options for the 2019 novel coronavirus (2019-nCoV), 2020.
  28. M.-H. Lin, D. C. Moses, C.-H. Hsieh, S.-C. Cheng, Y.-H. Chen, C.-Y. Sun, and C.-Y. Chou. Disulfiram can inhibit MERS and SARS coronavirus papain-like proteases via different modes. *Antiviral Re.*, 150:155–163, 2018.
  29. W. Liu, J. S. Morse, T. Lalonde, and S. Xu. Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-ncov. *Chembiochem*, 2020.
  30. I.-L. Lu, N. Mahindroo, P.-H. Liang, Y.-H. Peng, C.-J. Kuo, K.-C. Tsai, H.-P. Hsieh, Y.-S. Chao, and S.-Y. Wu. Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.*, 49(17):5154–5161, 2006.
  31. P. S. Masters. The molecular biology of coronaviruses. In *Adv. Virus Res.*, volume 66, pages 193 – 292. Academic Press, 2006.
  32. L. Mittal, A. Kumari, M. Srivastava, M. Singh, and S. Asthana. Identification of Potential Molecules Against COVID-19 Main Protease Through Structure-Guided Virtual Screening Approach. *ChemRxiv*, 4 2020.
  33. T. Muramatsu, C. Takemoto, Y.-T. Kim, H. Wang, W. Nishii, T. Terada, M. Shirouzu, and S. Yokoyama. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. USA*, 113(46):12997–13002, 2016.
  34. C. S. I. Nobel, M. Kimland, D. W. Nicholson, S. Orrenius, and A. F. Slater. Disulfiram is a potent inhibitor of proteases of the caspase family. *Chem. Res. Toxicol.*, 10(12):1319–1324, 1997.
  35. N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison. Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):33, 2011.
  36. L. L. Palese. Analysis of the conformations of the HIV-1 protease from a large crystallographic data set. *Data Brief*, 15:696–700, 2017.
  37. L. L. Palese. Conformations of the HIV-1 protease: a crystal structure data set analysis. *Biochim. Biophys. Acta*, 1865(11):1416–1422, 2017.
  38. L. L. Palese. A random version of principal component analysis in data clustering. *Comput. Biol. Chem.*, 73:57–64, 2018.
-

- 
39. T. Pillaiyar, S. Meenakshisundaram, and M. Manickam. Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discov. Today*, 2020.
  40. A. Rabinkov, T. Miron, L. Konstantinovski, M. Wilchek, D. Mirelman, and L. Weiner. The mode of action of allicin: trapping of radicals and interaction with thiol containing proteins. *Biochim. Biophys. Acta*, 1379(2):233–244, 1998.
  41. K. Ratia, K. S. Saikatendu, B. D. Santarsiero, N. Barretto, S. C. Baker, R. C. Stevens, and A. D. Mesecar. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. *Proc. Natl. Acad. Sci. USA*, 103(15):5717–5722, 2006.
  42. E. Roberts, J. Eargle, D. Wright, and Z. Luthey-Schulten. MultiSeq: unifying sequence and structure data for evolutionary analysis. *BMC bioinformatics*, 7(1):382, 2006.
  43. R. B. Russell and G. J. Barton. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins*, 14(2):309–323, 1992.
  44. E. Satoyoshi. Therapeutic trials on progressive muscular dystrophy. *Intern. Med.*, 31(7):841–846, 1992.
  45. S. Shityakov and C. Förster. In silico predictive model to determine vector-mediated transport properties for the blood–brain barrier choline transporter. *Advances and applications in bioinformatics and chemistry: AABC*, 7:23, 2014.
  46. M. Tamai, K. Matsumoto, S. Omura, I. Koyama, Y. Ozawa, and K. Hanada. In vitro and in vivo inhibition of cysteine proteinases by EST, a new analog of E-64. *J. Pharmacobiodyn.*, 9(8):672–677, 1986.
  47. M. Tamai, C. Yokoo, M. Murata, K. Oguma, K. Sota, E. Sato, and Y. Kanaoka. Efficient synthetic method for ethyl (+)-(2s, 3s)-3-[(s)-3-methyl-1-(3-methylbutylcarbamoyl) butylcarbamoyl]-2-oxiranecarboxylate (EST), a new inhibitor of cysteine proteinases. *Chem. Pharm. Bull.*, 35(3):1098–1104, 1987.
  48. O. Trott and A. J. Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.
  49. B. Turk. Targeting proteases: successes, failures and future prospects. *Nat. Rev. Drug Discov.*, 5(9):785–799, 2006.
  50. K. H. Verschuere, K. Pumpor, S. Anemüller, S. Chen, J. R. Mesters, and R. Hilgenfeld. A structural view of the inactivation of the SARS coronavirus main proteinase by benzotriazole esters. *Chem. Biol.*, 15(6):597–606, 2008.
  51. C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao. A novel coronavirus outbreak of global health concern. *Lancet*, 395(10223):470–473, 2020.
  52. F. Wang, C. Chen, W. Tan, K. Yang, and H. Yang. Structure of main protease from human coronavirus NL63: insights for wide spectrum anti-coronavirus drug design. *Sci. Rep.*, 6:22677, 2016.
  53. L. Wang, B.-B. Bao, G.-Q. Song, C. Chen, X.-M. Zhang, W. Lu, Z. Wang, Y. Cai, S. Li, S. Fu, et al. Discovery of unsymmetrical aromatic disulfides as novel inhibitors of SARS-CoV main protease: chemical synthesis, biological evaluation, molecular docking and 3d-qsar study. *Eur. J. Med. Chem.*, 137:450–461, 2017.
  54. L. Xu, J. Tong, Y. Wu, S. Zhao, and B.-L. Lin. Targeted oxidation strategy (TOS) for potential inhibition of coronaviruses by disulfiram—a 70-year old anti-alcoholism drug. *ChemRxiv*, 2020.

- 
55. X. Xue, H. Yang, W. Shen, Q. Zhao, J. Li, K. Yang, C. Chen, Y. Jin, M. Bartlam, and Z. Rao. Production of authentic SARS-CoV Mpro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.*, 366(3):965–975, 2007.
  56. S. Yang, S.-J. Chen, M.-F. Hsu, J.-D. Wu, C.-T. K. Tseng, Y.-F. Liu, H.-C. Chen, C.-W. Kuo, C.-S. Wu, L.-W. Chang, et al. Synthesis, crystal structure, structure- activity relationships, and antiviral activity of a potent SARS coronavirus 3CL protease inhibitor. *J. Med. Chem.*, 49(16):4971–4980, 2006.
  57. J. Yin, C. Niu, M. M. Cherney, J. Zhang, C. Huitema, L. D. Eltis, J. C. Vederas, and M. N. James. A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase. *J. Mol. Biol.*, 371(4):1060–1074, 2007.
  58. Q. Zhao, S. Li, F. Xue, Y. Zou, C. Chen, M. Bartlam, and Z. Rao. Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. *J. Virol.*, 82(17):8647–8655, 2008.
  59. L. Zhu, S. George, M. F. Schmidt, S. I. Al-Gharabli, J. Rademann, and R. Hilgenfeld. Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. *Antiviral Res.*, 92(2):204–212, 2011.