Near-Physiological-Temperature Serial Femtosecond X-ray Crystallography Reveals Novel Conformations of SARS-CoV-2 Main Protease Active Site for Improved Drug Repurposing

Serdar Durdagi[1], Çağdaş Dağ[2], Berna Dogan[1], Merve Yigin[2], Timucin Avsar[3], Cengizhan Buyukdag[2], Ismail Erol[1,4], Fatma Betul Ertem[2], Seyma Calis[3,5], Gunseli Yildirim[2], Muge D. Orhan[3], Omur Guven[2], Busecan Aksoydan[1,6], Ebru Destan[2], Kader Sahin[1], Sabri O. Besler[2], Lalehan Oktay[1], Alaleh Shafiei[2], Ilayda Tolu[1], Esra Ayan[2], Busra Yuksel[2], Ayse B. Peksen[2], Oktay Gocenler[2], Ali D. Yucel[2], Ozgur Can[2], Serena Ozabrahamyan[2], Alpsu Olkan[1,7], Ece Erdemoglu[1,8], Fulya Aksit[2], Gokhan Tanisali[2], Oleksandr M. Yefanov[9], Anton Barty[9], Alexandra Tolstikova[9], Gihan K. Ketawala[10,11], Sabine Botha[10,11], E. Han Dao[12], Brandon Hayes[13], Mengning Liang[13], Matthew H. Seaberg[13], Mark S. Hunter[13], Alex Batyuk[13], Valerio Mariani[13], Zhen Su[13,14], Frederic Poitevin[13], Chun Hong Yoon[13], Christopher Kupitz[13], Raymond G. Sierra[13], Edward Snell[15] & Hasan DeMirci[2,12]

[1] Computational Biology and Molecular Simulations Laboratory, Department of Biophysics, School of Medicine, Bahcesehir University, 34734, Istanbul, Turkey
[2] Department of Molecular Biology and Genetics, Koc University, 34450, Istanbul, Turkey
[3] Department of Medical Biology, School of Medicine, Bahcesehir University, 34734, Istanbul, Turkey
[4] Department of Chemistry, Gebze Technical University, 41400, Kocaeli, Turkey
[5]Department of Molecular Biology - Genetics and Biotechnology, Istanbul Technical University, 34469, Istanbul, Turkey
[6] Neuroscience Program, Graduate School of Health Sciences, Bahcesehir University, 34734, Istanbul, Turkey
[7] School of Medicine, Bahcesehir University, 34734, Istanbul, Turkey
[8] Faculty of Medicine, Mersin University, 33070, Mersin, Turkey
[9] Deutsches Elektronen-Synchrotron, Notkestrasse 85, 22607, Hamburg Germany
[10] Department of Physics, Arizona State University, Tempe, AZ, 85287-1504, USA
[11] Biodesign Center for Applied Structural Discovery, Arizona State University, Tempe, AZ, 85287-5001, USA
[12] Stanford PULSE Institute, SLAC National Laboratory, 94025, Menlo Park, CA, USA
[13] Linac Coherent Light Source, SLAC National Accelerator Laboratory, 2575 Sand Hill Rd, Menlo Park, 94025, CA, USA
[14]Department of Applied Physics, Stanford University, Stanford, CA, USA
[15]Hauptman-Woodward Medical Research Institute, 700 Ellicott St, University at Buffalo, Buffalo, NY, USA and Materials Design and Innovation, SUNY at Buffalo, 700 Ellicott St., Buffalo, NY USA

Abstract

The COVID-19 pandemic has resulted in 175 million reported infections and more than 3.8 million deaths as of June 2021 (*covid19.who.int*). Research to identify effective therapies for COVID-19 includes: i) designing a vaccine as future protection; ii) de novo drug discovery; and iii) identifying existing drugs to repurpose them as effective and immediate treatments. To assist in drug repurposing and design, we determined two apo structures of Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2) main protease at ambient-temperature by Serial Femtosecond X-ray crystallography. We employed detailed molecular simulations of selected known main protease inhibitors with the structures and compared binding modes and energies. The combined structural and molecular modeling studies not only reveal the dynamics of small molecules targeting the main protease but also

provide invaluable opportunities for drug repurposing and structure-based drug design strategies against SARS-CoV-2.

One Sentence Summary

Radiation-damage-free high-resolution SARS-CoV-2 main protease SFX structures obtained at near-physiological-temperature offer invaluable information for immediate drug-repurposing studies for the treatment of COVID-19.

INTRODUCTION

In late 2019, after the first patient was diagnosed with pneumonia of unknown etiology reported to the World Health Organization (WHO) from China, millions of cases followed in a short span of four months (WHO). On the 11th of March, 2020; WHO declared the COVID-19 outbreak as a pandemic, which originated from Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2) infection. SARS-CoV-2 has a high spread rate (R_o) value, deeming the pandemic difficult to control (Petersen et al., 2020). Moreover, the lack of effective treatments to control the infection in high comorbidity groups made this pandemic a major threat to global health (Ahn et al., 2020).

The first human coronavirus to cause a variety of human diseases, such as common cold, gastroenteritis, and respiratory tract diseases were identified in the 1960s (Tyrrell and Bynoe, 1965). In 2002, a deadly version of coronavirus, SARS CoV, was identified in China (Heymann and Rodier, 2004). SARS-CoV-2, the most recent member of the coronavirus family to be encountered, is a close relative of SARS-CoV and causes many systemic diseases (Andersen et al., 2020; Dutta and Sengupta, 2020; Braun et al., 2020). COVID-19 patients exhibit (i) high C-reactive protein (CRP) and pro-inflammatory cytokine levels; (ii) macrophage and monocyte infiltration to the lung tissue; (iii) atrophy of spleen and lymph nodes which weakens the immune system; (iv) lymphopenia; and (v) vasculitis (Zhang et al., 2020; McGonagle et al., 2020). The release of a large amount of cytokines results in acute respiratory distress syndrome (ARDS) aggravation and widespread tissue injury leading to multi-organ failure and death. Therefore, mortality in many severe cases of COVID-19 patients has been linked to the presence of the cytokine storm evoked by the virus (Ragab et al., 2020).

The SARS-CoV-2 genome encodes structural proteins including surface/spike glycoprotein (S), envelope (E), membrane (M), and nucleocapsid (N) proteins; and the main reading frames named ORF1a and ORF1b that contain 16 non-structural proteins (NSP) (Gordon et al., 2020; Chen et al.,

2020). Among these, ORF1a/b encodes papain-like protease (PLpro), main protease (Mpro), a chymotrypsin-like cysteine protease, along with polyproteins named polyprotein1a (pp1a) and polyprotein1b (pp1b) (Astell et al., 2005). Encoded polyproteins are then proteolyzed to NSPs by precise Mpro and PLpro cleavages of the internal scissile bonds. NSPs are vital for viral replication, such as RNA-dependent RNA polymerase (RdRp) and Nsp13, which are used for the expression of structural proteins of the virus (Dai et al., 2020; Thiel et al., 2003; Ullrich and Nitsche, 2020; Ziebuhr and Snijder, 2000). SARS-CoV Mpro has no homologous human protease that recognizes the same cleavage site (Pillaiyar et al., 2016). Therefore, drugs that target its active site are predicted to be less toxic and harmful to humans (Zhang et al., 2020). High sequence conservation of Mpro provides minimized mutation-caused drug resistance (Jin et al., 2020). Given its essential role in the viral life cycle (Dai et al., 2020; Thiel et al., 2003; Ullrich and Nitsche, 2020; Ziebuhr et al., 2000), the SARS-CoV-2 Mpro presents a major drug target requiring a detailed structural study.

Structure-based drug repurposing is a rapid method of identifying potential therapies that could be effective against COVID-19 compared to continuous investigative efforts (e.g. identification of new drugs and development of preventive vaccine therapies). The well-studied properties of drugs currently in clinical use mean that such compounds are better understood compared to their counterparts designed *de novo*. A putative drug candidate identified by drug-repurposing studies could make use of existing pharmaceutical supply chains for formulation and distribution, an advantage over developing new therapies (Chen and Li, 2020; Huang et al., 2020; Pushpakom et al., 2018; Jarada et al., 2020). In previous drug repurposing studies, approved drug libraries were typically screened against the active site or an allosteric site of target protein structures which were traditionally obtained by methods with limitations in revealing the enzyme structure, *e.g.* cryogenic temperature or radiation damage (Wang et al., 2020; Kneller et al., 2020). Such investigations generally only involve screening of drugs currently on the market (Zhou et al., 2020; Choudhary et al., 2020; Beck et al., 2020; Wang, 2020). Current structural biology-oriented studies that display a repurposing approach to SARS-CoV-2 research have focused on structure-based drug design, virtual screening and a wide spectrum of inhibitors based on previously published tertiary protein structure alignments (Rathnayake et al., 2020; Chauhan and Kalra, 2020; Chen et al., 2020; Muralidharan et al., 2020; Khan et al., 2020; Kumar et al., 2020; Joshi et al., 2020). Given the high-infectivity and persistence of SARS-CoV-2, any molecular candidates identified through drug-repurposing studies could be a potential treatment in the short term and should thus be investigated using empiric structural data.

In this work, we ascertain some of the previously-unseen conformational dynamics of Mpro via high-throughput serial crystallography at Macromolecular Femtosecond Crystallography (MFX) instrument of the Linac Coherent Light Source (LCLS) which provides ultrafast and ultrabright X-ray pulses allowing data collection without secondary radiation damage at ambient temperature (Sierra et al., 2019; Blaj et al., 2019; Van Driel et al., 2020). Our structural data reveal the active site of SARS-CoV-2 Mpro with ambient-temperature radiation-damage free data and were used for subsequent *in silico* investigation of known Mpro inhibitors with the binding pocket residues.

RESULTS

Ambient-temperature XFEL crystal structures of SARS-CoV-2 Mpro reveal alternate conformations of the drug-binding pocket. We determined two radiation-damage-free SFX crystal structures of SARS-CoV-2 Mpro in two crystal forms, monoclinic at 1.9 Å and orthorhombic at 2.1 Å resolutions (PDB IDs: 7CWB and 7CWC, respectively) (Figure 1A, B) (Supplementary Tables 1 and 2 and Supplementary Figure 1). We used a Mpro structure determined at ambient-temperature using a rotating anode home X-ray source (Kneller et al., 2020) as our initial molecular replacement search model for structure determination (PDB ID: 6WQF). Two high-resolution SFX structures obtained in different space groups were superposed with an overall RMSD of 1.0 Å (Supplementary Figure 2). They reveal novel active site residue conformations and dynamics at the atomic level, revealing several differences compared to the prior ambient-temperature structure of 6WQF (Kneller et al., 2020) (Figure 2A, B).

The unique N-terminal sequence affects the enzyme's catalytic activity (Chang, 2010; Zhang et al., 2020). Besides the native monoclinic form of Mpro at 1.9 Å (Figure 1A), we also determined the structure of the Mpro with additional 4 extra N-terminal amino acids (generated by thrombin specific N-terminal cleavage) at 2.1 Å resolution (Figure 1B). The structure obtained from this modified version of Mpro reveals that minor changes introduced at the N-terminus can affect the three-dimensional structure of the Mpro and promotes the formation of a new orthorhombic crystal form. The biologically relevant dimeric structure of native monomeric Mpro can be generated by adding the symmetry-related chain B (Figure 1A).

Each protomer of SARS-CoV-2 Mpro is formed by three major domains (Jin et al., 2020) (Supplementary Figure 3). Domain I starts from the N-terminus of the protein and includes an antiparallel β-sheet structure. This β-sheet forms a β-barrel fold which ends at residue 100. The domain

II of Mpro resides between residues 101 through 180 and mostly consists of antiparallel β-sheets. The third domain of the Mpro is located between residues 181 through 306 and consists of mostly α helices and has a more globular tertiary structure (Supplementary Figure 3). The intersection between domain I, domain II and loop region of domain III is the key region for enzyme activity (Supplementary Figure 3) which forms the catalytic site and substrate-binding pocket of the enzyme. The biologically active form of SARS-CoV-2 Mpro is a homodimer (Zhang et al., 2020). Previous biochemical studies of SARS-CoV Mpro suggested there is a competition for dimerization surface between domains I and III. In the absence of domain I, Mpro undergoes a new type of dimerization through domain III (Zhong et al., 2008). During our purifications, we repeatedly observed a combination of monomeric and dimeric forms of Mpro on the size exclusion chromatography steps which may be caused by this dynamic compositional and conformational equilibrium.

The two SARS-CoV-2 Mpro SFX crystal structures reveal a non-flexible core active site and the catalytic amino acid Cys145 (Supplementary Figure 4). Temperature factor analysis revealed that the active site is surrounded by mobile regions (Supplementary Figure 4). The presence of these mobile regions was observed in both SFX crystal structures, suggesting intrinsic plasticity rather than structural artifact arising from the crystal lattice contacts (Supplementary Figure 4*;* indicated with red circles). Further, this intrinsic plasticity suggests that inhibitor molecules that interact via noncovalent bonds to the Mpro binding pocket would do so weakly (Gao et al., 2021). Our investigation of PDB structures with available electron densities identified that the majority of the Mpro inhibitors formed covalent bonds with the active site residue Cys145. Additionally, few noncovalent inhibitors were identified among the available structures in September 2020 which have downloadable electron densities and exhibited weak electron densities (Supplementary Figure 5.1 - 5.37).

The cryogenic temperature structures of SARS-CoV-2 Mpro were summarized in an outstanding study and compared them with ambient-temperature 6WQF structure (Kneller et al., 2020). Thus, we compared our monoclinic ambient-temperature SFX structure with the previously published ambient-temperature X-ray structure 6WQF for observing minimization of secondary radiation damage effect on the SARS-CoV-2 Mpro structure by SFX technology. Both structures were largely similar with an RMSD value of 0.404 Å (Figure 2A). We observed significant conformational differences especially in the side chains of Thr24, Ser46, Glu47, Leu50, Asn142, Cys145, Met165, and Gln189 residues (Figure 2B & Supplementary Figure 6). As suggested before, flexibility could conceivably help design by providing alternative binding sites for non-native molecules (Niu et al., 2008). Observed significant conformational differences are caused by intrinsic flexibility around the binding pocket in our SFX

structure. This flexibility provides us to observe the alternative side chain conformations compared to ambient-temperature X-ray structure (PDB ID: 6WQF). The calculated bias-free composite omit map that covers the active region has been shown (Figure 3B & Supplementary Figure 7). This structure offers new insights on the active site of Mpro in addition to 6WQF structure which may be important for the future ligand screening and *in silico* modeling studies. All three domains contribute to the formation of the active site of the protein. The intersection part of the domain I residue His41, domain II residues Cys145 and His164 interact via a coordinated water molecule with Asp187 located at the N-terminal loop region of domain III to form the active site (Figure 2C & Supplementary Figure 8). The N-terminal loop of domain III is suggested to be involved in enzyme activity (Ma et al., 2020). The distance between Cys145 Sγ and His41 Nε2 is 3.7 Å, similar to the ambient-temperature structure 6WQF (Kneller et al., 2020). Oδ1-Oδ2 atoms of Asp187 and NH2-Nε atoms of Arg40 contribute to a salt bridge between these two residues and stabilize the positions of each other. The W5 water molecule (indicated with a red sphere at the figure) in the active site plays crucial roles for catalysis (Kneller et al., 2020). W5 forms triple H-bonds with His41, His164, and Asp187 side chains with the distances of 2.8 Å, 2.9 Å, and 2.8 Å, respectively (Figure 2C & Supplementary Figure 8).

When compared to the room temperature structure of Mpro 6WQF, our structure displays additional active site residue dynamics while it has an overall high similarity (Figure 2D & Figure 4G, H). Canonical chymotrypsin-like proteases contain a catalytic triad composed of Ser(Cys)-His-Asp(Glu) in their catalytic region, however, SARS-CoV-2 Mpro possesses a Cys145 and His41 catalytic dyad which distinguishes the SARS-CoV-2 Mpro from canonical chymotrypsin-like enzymes (Kneller et al., 2020; Gorbelanya and Snijder, 1996). During the catalysis, the thiol group of Cys145 is deprotonated by the imidazole of His41 and the resulting anionic sulfur nucleophilically attacks the carbonyl carbon of the substrate. After this initial attack, an N-terminal peptide product is released by abstracting the proton from the His41, resulting in the His41 to become deprotonated again and a thioester is formed as a result. In the final step, the thioester is hydrolyzed which results in a release of a carboxylic acid and the free enzyme; therefore, restoring the catalytic dyad (Ullrich and Nitsche, 2020; Pillaiyar et al., 2016). Observed catalytic residue conformations of SFX structures are consistent with 6WQF and support the proposed Mpro catalytic mechanism (Figure 2C, D).

The crystal contact of the symmetry-related molecule with the N-terminal region of the Mpro is essential for the formation of the crystal lattice in the monoclinic *C*121 space group (Figure 3A & Supplementary Figure 9). There is an extensive network of hydrogen bonding interactions. The Ser1 amino group and carbonyl O atom engage in two hydrogen bonding interactions with carbonyl O and

backbone amide N of Phe140. The carbonyl group of Gly2 forms a hydrogen bond with the carbonyl Oγ atom of Ser139. The backbone amide group of Phe3 forms a second hydrogen bonding interaction with the carbonyl O atom of Gly138 in the *C*121 space group. The side chain amine and amino group of Arg4 form a hydrogen bond with the carbonyl O atom of Lys137 (Supplementary Figure 9). These interactions are crucial, and elimination of these hydrogen bond network by the addition of 4 N-terminal amino acids switches SARS-CoV-2 Mpro to the secondary crystal conformation in the orthorhombic $P2_12_12_1$ space group (Supplementary Figure 9 & 10). In the new crystal form the active site pocket has a wider binding pocket (Supplementary Figure 11). This wider binding pocket may offer opportunities for obtaining more effective drug soaks for future structural studies of Mpro drug complexes guided by hybrid *in silico* methods.

All-atom molecular dynamics (MD) simulations of apo forms of Mpro reveals that the protomers display asymmetric behavior. To better understand the structural and dynamical properties Mpro, MD simulations were performed using our newly determined SFX crystal structures (Figure 4). As the catalytically active form of Mpro is a homodimer, we have used the dimeric form of crystal structures. For the SFX structure obtained in space group *C*121 with a monomeric asymmetric unit, the relevant dimeric form is generated by symmetry operation. The MD simulations were run for 200 ns to elucidate the dynamic effects on the structures, especially the active site region. The evolutionary changes of atomic coordinates over time were monitored by calculating the RMSD for both chains, i.e. protomers (Supplementary Figures 12 & 13 for 7CWB and 7CWC, respectively). The root-mean-square fluctuations (RMSF) were also computed based on Cα atoms for both protomers separately (Supplementary Figures 14 & 15 for 7CWB and 7CWC, respectively) to determine the flexible regions. In simulations for both 7CWB and 7CWC, the protomers exhibit non-identical behavior as also observed by others in apo form of Mpro (Suárez and Diaz, 2020; Amamuddy et al., 2020) though as expected, higher RMSF values were observed for the loop regions of protomers. Additionally, the protomers of 7CWC display higher fluctuations compared to 7CWB around the loops covering the active site (such as loops containing residues 44-52 and 185-190) which could affect the accessibility of the active site by inhibitors (Supplementary Figure 14 & 15).

Principal component analysis (PCA) correlates the inter-domain motions and its impact on the drug-binding pocket dynamics. The trajectory frames obtained from MD simulations were used to perform PCA to determine the variations of conformers of protein structures, i.e. to observe the slowest motions during MD simulations. As PCA and PCA-based methods are useful to reveal intrinsically accessible movements such as domain motions (Bahar et al., 2010), we have performed PCA for

backbone atoms of dimeric units for the structures belonging to different space groups. We focused on the first three PCs, that show around 40% of the total variance in MD trajectories to determine the regions of protein structures that display the highest variation (Supplementary Figures 16 & 17 for 7CWB and 7CWC, respectively). The first three PCs were projected onto the protein structures to determine the contributions of each residue to specified PCs. The motions of specific regions with blue regions with higher thickness represent more mobile structural parts of the protein along the specified PCs. We also provide supplementary movies to illustrate the movements along the species PCs: In supplementary movies M1-3 the motions along PC1, PC2 and PC3 for 7CWB are displayed, respectively while for 7CWC the movements along first three PCs are displayed in supplementary movies M4-6, respectively. In the movies, different colors represent the different domains of the enzyme such that domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B. In all of the movies presented, protomer-A is shown on the right and protomer-B is shown on the left. It can be seen that both in $C121$ and $P2_12_12_1$ space group structures, again protomers A and B display asymmetric behavior and domain movements along all considered PCs (Figures 4A-C for 7CWB and Figures 4D-F for 7CWC) (Amamuddy et al., 2020). When the motions displayed by two dimeric forms are compared to each other, we observe that domain III is more mobile in 7CWB compared to 7CWC (Supplementary movies M1-M3 for 7CWB and M4-M6 for 7CWC). However, the loop region containing residues 45-53 around the catalytic site is more mobile for both protomers in 7CWC while it is only mobile for the chain B of 7CWB. Interestingly, protomer A of 7CWC is more mobile compared to protomer A of 7CWB which could reflect the differences in dimeric interface due to the additional 4 N-terminal amino acids and missing interactions with chain B N-terminal residues in the dimer interface. However, the loop regions surrounding the binding pockets with residues 166-172 and 185-195 display higher flexibility in chain A of 7CWB while its mobility is somewhat restricted in 7CWC.

We have also performed cross correlation analysis of residues along the PC space to understand the correlation between motions especially for different domains. The correlation between the motions along the first three PCs were plotted as dynamical cross correlation maps displayed in Figures 4G & 4H for 7CWB and 7CWC, respectively. Focusing on the correlation between motions within protomers initially, we see that motions of domains I and III as well as domains I and II are mostly anti-correlated as shown by the purple regions in Figure 4G for 7CWB and Figure 4H for 7CWC. While we observed similar behavior for protomers A and B of 7CWC, the directions of domains I and III movements are different in protomers A and B of 7CWB, specifically for residues 189-192 that is

at the linkage points of domains II and III and residues 45-50 of domain I bordering the catalytic site. While in protomer A, these loops move in the same directions, their movements are mostly anti-correlated in protomer B for the slowest motion along PC1 in 7CWB (Supplementary movies M1, Figure 4G). This could cause differences in the accessibility of catalytic sites in a time-dependent manner in protomers of 7CWB. On the other hand, domains II and III within protomers have mixed correlations with each other in which motions of some residues are along the same directions, others are in opposite directions within both protomers A and B of 7CWB while domain II has limited mobility as being more buried than other domains (Suárez and Diaz, 2020) except for the β-strand segment of residues 166-172 of protomer A. Surprisingly, this segment also has some mobility albeit limited in protomer B of 7CWC instead of protomer A (Supplementary movies M4, Figure 4H). Glu166 of this segment is actually an important residue that plays a role in stabilizing the substrate binding site S1 by interacting with Ser1 of the alternate protomer along with Phe140 (Ghahremanpour et al., 2020 & Jin et al., 2020) and we observed that along PC1, for protomer A of 7CWB and protomer B of 7CWC, the motions of these residues are correlated (Supplementary movies M1 for 7CWB and M4 for 7CWC). This could have implications about information transfer from one protomer to the other though other Mpro structures of SARS-CoV-2 needs to be studied in atomistic details. In addition, we observed that correlations between domains of protomers A and B for 7CWB and 7CWC are dissimilar. For instance, domains I of 7CWB have defined anti-correlated motions as can be seen from the purple areas in Figure 4H while in 7CWC, the movements of domain I are more ambiguous. There are also differences in the behavior of domain I of protomer A and domain II of protomer B in the dimer structures in which for 7CWB the motions of these domains are positively correlated while they are anti-correlated in 7CWC.

Different non-covalent interactions stabilize the dimer interfaces in different space groups. The crystal structures that were obtained in this study, compared with four other Mpro structures crystallized recently. Selected Mpro structures were cryogenic apo form (Zhang et al., 2020) (PDB IDs: 6Y2E & 7C2Y), holoprotein with non-covalent inhibitor, X77 (PDB ID: 6W63) and room temperature apo form structure (Kneller et al., 2020) (PDB ID: 6WQF). Space groups of these structures are $C121$, $P2_12_12_1$, $P2_12_12$, and $I121$ respectively. The completeness of the structures was also evaluated, 6WQF, and 6Y2E crystallized in full-length sequence (1-306), however, 6W63 only lacks Gln306 at both C-terminal ends of its chains. Among compared structures, 7C2Y, which has the same space group as 7CWC has, lacks 24 amino acid residues (chain A, 1-3, 141-142, 281-283, and 298; chain B, 1-2, 45-50, 139-142, and 277-279). In our structures, 7CWB has a full-length sequence,

however, 7CWC lacks its last 6 amino acid residues from C-terminal in chain A, and starts with Phe3 (lacks Ser1 and Gly2), and ends at Ser301, lacks the last 5 residues. When six structures (6W63, 6Y2E, 6WQF, 7C2Y, 7CWB, and 7CWC) were compared based on hydrogen bonding interactions at the dimerization interfaces, more similarities were observed for the latter three (7C2Y, 7CWB, and 7CWC). One of the main differences between our structures and the other four is the hydrogen bond between Ser139 (chain A) and Gln299 (chain B) (Supplementary Figure 18). Although this hydrogen bond is not observed in other structures, the corresponding residues are close to each other, however they are not within hydrogen bonding distance (Supplementary Figure 19).

We also monitored all interface interactions throughout the MD simulations and compared 7CWB and 7CWC. Interestingly, the hydrogen bond between Ser139 (chain A) and Gln299 (chain B) was lost and the interaction was turned into a van der Waals interaction in 7CWB, and conserved as by 64% of the simulation time. In 7CWC structure, Ser139 (chain A) was not in the vicinity of Gln299 (chain B) (Supplementary Figure 20). More interestingly, the same interaction but this time between Gln299 (chain A) and Ser139 (chain B), that is present at the 7CWB crystal structure, was only observed 16% of the simulation time (Supplementary Figure 21). In the 7CWC crystal structure, there was no hydrogen bonding interaction between Gln299 (chain A) and Ser139 (chain B). However, during simulation, this bond was formed and retained in 65% of the simulation time (Supplementary Figure 22).

When static structures were compared based on hydrogen bond formation analysis, 7C2Y, 7CWB, and 7CWC clustered together, having the same hydrogen bonding network at the dimerization interface. The explanation for the hydrogen bond differences observed at the interface is the lack of amino acids at the N-terminal and C-terminal ends of the 7C2Y and 7CWC structures, compared to the 7CWB.

Simulation trajectories of 7CWB and 7CWC were compared based on hydrogen bond occupancies. The major differences were Lys137 (chain A) and Arg4 (chain B), Asn142 (chain A) and Ser301(chain B), and Arg298 (chain A) and Tyr118 (chain B). The first interaction was not observed in 7CWC, however the latter two interactions were not presented in 7CWB. Lys137 (chain A) and Arg4 (chain B) interaction was conserved 82% of the simulation time; and Asn142 (chain A) and Ser301 (chain B); and Arg298 (chain A) and Tyr118 (chain B) interactions retained 67% and 72% of the obtained trajectory frames, respectively (Supplementary Figures 21 & 22). We also observed that during MD simulations, in the case of 7CWC, Gln299 (chain A) and Phe140 (chain B) come closer to each other,

and this van der Waals contact was conserved 64% of the simulation time and not observed in the case of 7CWB (Supplementary Figure 21).

Non-covalent bound inhibitors form unstable complexes with Mpro. In their study Jin *et al* was screened in a library of around 10.000 compounds consisting of approved drugs, compounds in clinical trials and natural products using fluorescence resonance energy transfer assay (Jin et al., 2020). They identified 6 specific hit compounds which range the IC_{50} values from 0.67 to 21.4 μM. In this study, we used top-3 hit compounds (i.e., Ebselen (IC_{50}, 0.67 μM), Tideglusib (IC_{50}, 1.55 μM), and Carmofur (IC_{50}, 1.90 μM)) for the investigation of ligand-target interactions. These compounds were docked to the 7CWB and 7CWC structures and all-atom MD simulations were performed for the top-docking poses. Induced fit docking (IFD) approach was conducted to better account for the flexibility of both of the ligands and the active site of the target structures. These three compounds were also docked to structures with the following PDB IDs: 6W63 and 6Y2E (cryogenic), for comparison. MD simulations were performed using the same MD protocol as performed for apo form structures. Results showed that especially for apo form dimer targets, Ebselen and Carmofur are quite flexible at the binding pocket throughout the simulations and in most of the cases, they do not form a stable complex structure (Supplementary Figure 23). Tideglusib has a more stable structure at the binding pocket of Mpro, however, its binding modes are different at the different targets considered in this study (Figure 5).

Detailed MD simulations were performed for the stably bound SARS-CoV-2 Mpro inhibitor Tideglusib. Tideglusib initially docked to the binding pockets of dimeric Mpro structures 6Y2E and 7CWC using IFD approach. Top-docking poses of these compounds were then used in all-atom MD simulations using the same MD protocols. While Tideglusib was structurally very stable at the binding pocket of the 7CWC during the simulations, it was not so stable at the binding site of the 6Y2E (Figure 5A). Representative trajectory frames (i.e., the frame that has the lowest RMSD to the average structures) were used in the comparison of binding modes (Figure 5B). Results showed that while the binding mode of Tideglusib forms hydrogen bonds and π-π stacking interactions with Glu166, Gln189, and His172, respectively at the 7CWC; its corresponding binding mode at the 6Y2E only forms van der Waals type interactions with hydrophobic moieties. A timeline of protein-ligand contacts was visualized throughout the simulations (Figure 5C). Results showed that Thr25, Leu27, Met49, Glu166, and Gln189 form stable interactions with the ligand at the 7CWC. However, corresponding interactions of the ligand at the active site of 6Y2E were not stable (Supplementary Figure 24). Ligand interactions with protein residues that occur more than 15% of the simulation time are also presented (Figure 5D). Comparison of binding pocket volumes of 7CWC and cryogenic Mpro structure (PDB

ID; 6Y2E) shows the latter has a bigger average binding pocket volume (Supplementary Figure 25). Average binding pocket volumes were 174 and 142 $Å^3$, for the Tideglusib-bound 6Y2E and 7CWC structures, respectively. Corresponding solvent accessible surface area (SASA) values throughout the MD simulations also support this result. The SASA, which is the surface area of a molecule accessible by a water molecule of 7CWC is smaller than 6Y2E (Supplementary Figure 25).

For the comparison of monomer forms of Mpro, we applied IFD protocol to predict the binding mode of Tideglusib at the binding pocket of 6WQF and 7CWB. Carbonyl oxygens of the thiadiazolidine ring of the Tideglusib formed hydrogen bonds between Asn142, Gly143, and Glu166 from their backbone atoms. The naphthalene ring of the ligand formed a π-π stacking interaction with the His41 in the 7CWB structure. A similar binding mode of Tideglusib was observed when 6WQF was used. However, in this binding mode, His41 and Gly143 are found to be important. A backbone hydrogen bond was observed between one of the carbonyl oxygen of the thiadiazolidine ring and Gly143. His41 formed two π-π stacking interactions between the benzyl and the naphthalene rings of the Tideglusib (Supplementary Figure 26 A, B). These two binding modes predicted by IFD were used in 200 ns classical all-atom MD simulations. Each Mpro-Tideglusib system is evaluated based on RMSD changes from the average structure and we obtained representative structures from corresponding trajectories. In the representative structure of 7CWB-Tideglusib complex, we observed van der Waals interactions between surrounding residues, Thr25, His41, Cys44, Thr45, Ser46, Met49, Asn142, Cys145, His163, His164, Met165, Glu166, and Gln189. In the representative structure of 6WQF-Tideglusib complex, in addition to the similar van der Waals interactions, Asn142 and Gln189 formed two hydrogen bonds with the carbonyl oxygens of the thiadiazolidine ring from their backbone atoms (Supplementary Figure 26 C, D). The alignment of the representative frames from 6WQF and 7CWB yielded an RMSD value of 0.97 Å (Supplementary Figure 27).

Translational and rotational motions of the Tideglusib were also examined by the trajectory analyses. Conformational spaces that were explored by the Tideglusib were given in Supplementary Figure 28. We also monitored the binding cavity volume throughout the MD simulations in both complexes and obtained the average values as 154 and 127 $Å^3$ for the Tideglusib-bound 7CWB and 6WQF structures, respectively (Supplementary Figure 29). In the IFD poses, corresponding binding cavity volumes were calculated as 251 and 238 $Å^3$, respectively. Decreased binding cavity volume during the MD simulations may be correlated with the unstable conformations of the Tideglusib in both structures.

Overall, from the MD simulations of Tideglusib with the monomer forms of Mpro, we observed more stable conformations of Tideglusib at the binding pocket of 6WQF compared to the 7CWB.

Discussion

SFX utilizes micro-focused, ultrabright and ultrafast X-ray pulses to probe small crystals in a serial fashion. Structural information is obtained from individual snapshots; capturing Bragg diffraction of single crystals in random orientations (Martin-Garcia et al., 2016). The main advantages of SFX over its counterparts are the capability of working with micron to nanometer sized crystals which does not necessitate lengthy and laborious crystal growth optimization steps and enables working with multiple crystal forms and space groups. It enabled obtaining high resolution structures of SARS-CoV-2 Mpro at physiologically meaningful temperature and confirmed the dynamic regions of the active site. SFX offers great potential and provides critical information for future high-throughput structural drug screening and computational modeling studies with sensitivity to dynamics, resulting in production of detailed structural information.

SARS-CoV-2 Mpro catalyzes the precise cleavage events responsible for activation of viral replication and the expression of structural proteins (Jin et al., 2020). It has been the focus of several structural and biochemical studies, many of which have been performed at cryogenic temperatures, aiming to provide better understanding of the active site dynamics and reveal an inhibitor that affects the enzyme based on structural information. In our study, two crystal forms of Mpro, native and modified, were determined at ambient-temperature with resolutions of 1.9 Å and 2.1 Å, respectively. The two produced crystal forms are optimal for co-crystallization and soaking respectively. Co-crystallization experiments provide efficient interaction in the binding pocket as both drug and protein are stabilized before the formation of crystals. Due to the close crystal lattice contacts, co-crystallization should be the preferred method for larger molecules against the native form of Mpro in the space group *C*121. However, there are numerous successful soaking experiments with small non-covalent fragments (Douangamath et al., 2020). Residues at the N-terminus of Mpro play a critical role in crystal packing. Elimination of the H-bonding network produced orthorhombic crystals in the $P2_12_12_1$ space group yielding a wider binding pocket, increasing the probability of capturing an expanded number of protein-drug complexes by soaking. Obtaining two different crystal forms helped eliminate the artifacts introduced by specific lattice packing restraints of each crystal form and thereby increased the quality of the molecular dynamics analysis. However, the presence of four additional residues at

the N-terminus could hinder soaking or co-crystallization efforts given that ~200 ligand-bound structures of Mpro have been reported (Douangamath et al., 2020).

The high-resolution Mpro SFX structures presented here in two different crystal forms collectively revealed the intrinsic plasticity and dynamics around the enzyme's active site. Due to the anionic nature of Cys145, it seems challenging to design molecules that interact with the active site only through noncovalent bonds as it has a flexible region (Chang, 2010). These findings provide a structural basis for and are consistent with studies claiming that the majority of inhibitors form covalent bonds with the active site of Mpro, through Cys145 (Supplementary Figures 5.4-5.37). Especially, unlike what is suggested by some studies, Ebselen does not form a stable complex structure (Jin et al., 2020; Sies and Parnham, 2020; Menendez et al., 2020; Zmudzinski et al., 2020; Weglarz-Tomezak, 2020), (Supplementary Figure 30). Cys145 residue is a key amino acid for catalytic activity, on the other hand coordinated W5 molecule, regulating the catalytic reaction via triple hydrogen-bonding interactions with the His164, and the Asp187 that stabilizes the positive charge of His41 residue (Kneller et al., 2020). This active site residue conformations of SFX structures are consistent with previous ambient-temperature structure (6WQF) (Figure 2C, D).

N-terminal loop residue Gln189 from domain-III contributes to the stability of the potential inhibitors (Dai et al., 2020; Kneller et al., 2020) along with Asn142 and Ser46, are the active site residues forming the flank of the cavity (Figure 6). In a recent study, Asn142 and Gln189 have been indicated to interact with 11a (N-[(1S)-1-(cyclohexylmethyl)-2-[[(1S)-1-formyl-2-[(3S)-2-oxo-3-pyrrolidinyl]ethyl]amino]-2-oxoethyl]-1H-indole-2-carboxamide) and 11b (N-[(1S)-1-[(3-fluoro phenyl)methyl]-2-[[(1S)-1-formyl-2-[(3S)-2-oxo-3-pyrrolidinyl]ethyl]amino]-2-oxoethyl]-1H-indole-2-carboxamide) inhibitors that have proven *in vitro* effectiveness (Dai et al., 2020). Along with these amino acids, Thr24, which makes van der Waals interaction with the N3 inhibitor, and Ser46 which undergoes conformational changes in the presence of this inhibitor (Jin et al., 2020) (PDB ID: 6LU7), and Leu50 binding to N3, gave different side-chain conformations with SFX, according to 6WQF structure (Kneller et al., 2020) (Figure 2B). There were differences in the crystallization conditions between the SFX and 6WQF studies, the former making use of charge effects and the latter molecular crowding. Kneller et al elegantly presented the major differences between cryogenic and ambient temperature Mpro structures (Kneller, 2020). The SFX study is eliminating potential artifacts to examine crucial amino acids at the atomic level such as decarboxylation of acidic residues is frequently encountered at high X-ray doses, especially in terms of catalytic and inhibitor binding sites (Supplementary Figures 6-8) (Martin-Garcia et al., 2016; Johansson et al., 2017). These two high-

resolution SFX structures in different space groups reveal new active site residue conformations, intra- and inter-domain network and their dynamics at the atomic level, which provides a better understanding on structural allosteric transitions of Mpro structure interacting with the inhibitors (Figure 5 & Supplementary Figures 23-30). Therefore, considering the importance of the required sensitivity in drug design or the use of natural compounds in the studies, these active site residue conformations reveal the critical importance of our study more clearly.

There are many *in silico* docking studies performed based on cryogenic Mpro protein structures of SARS-CoV-2 (Dai et al., 2020; Pillaiyar et al., 2016; Jin et al., 2020; Ton et al., 2020; Durdagi et al., 2020). Although potent antiviral drug candidates are identified and several vaccines on the market, they have not yielded a desirable final cure yet. Having access to the alternative ambient-temperature structures of Mpro and observed conformational changes on active site residues will be a significant boon for the development of therapeutics by providing robust models for computational studies and provide better understanding on ligand and inhibitor binding. At this point, our work has two original aspects. Firstly, we used a comprehensive platform, SFX, which helps to deeply understand the complexity of SARS-CoV-2 and the structural dynamics of the target protein Mpro at near-physiological-temperatures. Secondly, we determined two high-resolution SFX structures of SARS-CoV-2 Mpro in two different space groups due to the new high-throughput data collection setup offered by the MFX instrument of the LCLS-II.

A grave issue with antiviral drug research is the variability of the target proteins as the mutations and modifications may render the found drugs ineffective (Dinesh et al., 2019). For this reason, working with a protein whose biochemical properties are conserved over time and among different strains carries significant importance. Evolutionarily, viruses bypass the host immune system by mimicking the proteins involved in the functioning of the host organism. For instance, cytomegalovirus (HCMV) can mimic a common host protein to hijack normal cell growth machinery or human immunodeficiency virus (HIV) can mimic a high percentage of human T cell receptors (Root-Bernstein, 2017; Robertson, 2003). SARS-CoV-2 virus contains the PLpro enzyme, which is highly similar to the deubiquitinating enzymes (USP7, USP14) in human metabolism, which potentially pose risk of interfering host metabolic pathways (Ratia et al., 2014). In addition, Spike protein, one of the other targeted proteins has a higher mutation rate (Jia et al., 2020) and includes a similar restriction site with epithelial channel protein (Anand et al., 2020). Besides that, inhibitors which target Spike protein provide a more effective solution before the host cell infection by the virus. As there is no Mpro homolog in the human genome, targeting this protease is therefore highly attractive with reduced

cross reactivity and side effects, making it an ideal candidate for drug therapy (Goyal and Goyal, 2020) but covalent inhibitors may show cross reactivity with human enzymes.

Drug repurposing has been the preferred area of research in the insufficiency of time and resources in emergency cases such as a novel pandemic, however, its impact on COVID-19 is debatable (e.g. remdesivir, hydroxychloroquine) (Beigel et al., 2020; Ip et al., 2021). Knowing the beneficial and detrimental effects of targeted drugs, along with well-established precautions will help with time limitations. This procedure has emerged as a fundamental and very strategic approach not only for prospective cohort design but also for many types of clinical trials. Particularly cross-sectional studies as the molecules considered in repurposing studies passed through several stages, have well-defined profiles. They would not require prolonged pre-clinical studies. Hence, they should be potential candidates to consider in case of disease emergencies or outbreaks (Su and Sanger, 2017; Cavalla, 2013; Choo et al., 2019; Aguila and Cua, 2020). Considering all this, adopting the drug repurposing approach and using known inhibitors Ebselen, Tideglusib, and Carmofur to carry out Mpro-based *in silico* molecular docking, MD simulations and post-MD analyses make our hybrid approach more specific for the subsequent future studies. Our novel structural active site dynamics data further provides the nearest physiological template for the future structure-based drug design and development of new Mpro inhibitors.

Acknowledgements

HD acknowledges support from National Science Foundation (NSF) Science and Technology Centers grant NSF-1231306 (Biology with X-ray Lasers, BioXFEL) and The Scientific and Technological Research Council of Turkey (TUBITAK) grant (118C270). HD would like to thank Michelle Young, Ritu Khurana, Lori Anne Love and Tracy Chou for their invaluable support and discussions. Authors would like to thank James Fraser & Galen Correy for their critical reading of our manuscript and their comments. Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The HERA system for in helium experiments at MFX was developed by Bruce Doak and funded by the Max-Planck Institute for Medical Research. Research was supported by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act. The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA

resources). SD and IE acknowledge TRUBA for the computational resources. SD acknowledges support from Bahcesehir University (BAU) Scientific Research Projects (BAP) grant (BAU2020-0101).

Author contributions

HD and SD designed and coordinated the project. HD, CD, MY, CB, FBE, GY, OG, SOB, AS, FA and GT prepared the samples. BH, ML, MHS, MSH, AB, VM, ZS, FP, CHY, CK, RGS performed the sample delivery and on-site data collection. HD, CD, BD, MY, CB, IE, FBE, SC, GY, MDO, OG, ED, SOB, AS, EA, BY, ABP, OG, ADY, OC, SO, EHD performed the remote data collection. SB, OMY, AB, AT, GKK, ZS, FP and CHY executed the data processing and reduction. SD, BD, TA, IE, SC, MDO, BA, KS, LO, IT, AO and EE performed the *in silico* analysis. Structures were refined by HD. Data analyzed by HD, ES, SD, CD, ED, BY, IE, BD, TA. The manuscript was prepared by HD and SD with input from all the coauthors.

References:

Aguila, E. J. T. & Cua, I. H. Y. Repurposed GI Drugs in the Treatment of COVID-19. *Dig. Dis. Sci.* (2020) doi:10.1007/s10620-020-06430-z.

Ahn, D. G. et al. Current status of epidemiology, diagnosis, therapeutics, and vaccines for novel coronavirus disease 2019 (COVID-19). Journal of Microbiology and Biotechnology (2020) doi:10.4014/jmb.2003.03011.

Amamuddy, O. S., Verkhivker, G. M. & Bishop, Ö. T. Impact of emerging mutations on the dynamic properties the SARS-CoV-2 main protease: An in silico investigation. *bioRxiv* (2020) doi:10.1101/2020.05.29.123190.

Anand, P., Puranik, A., Aravamudan, M., Venkatakrishnan, A. J. & Soundararajan, V. SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. Elife (2020) doi:10.7554/eLife.58603.

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine* (2020) doi:10.1038/s41591-020-0820-9.

Astell, C. R., Holt, R. A., Jones, S. J. M. & Marra, M. A. Genome organization and structural aspects of the SARS-related virus. in *Coronaviruses with Special Emphasis on First Insights Concerning SARS* (2005). doi:10.1007/3-7643-7339-3_5.

Bahar, I., Lezon, T.R., Bakan A. & Shrivastava, I. H. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. *Chem. Rev*. (2010), doi:10.1021/cr900095e.

Beck, B. R., Shin, B., Choi, Y., Park, S. & Kang, K. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Comput. Struct. Biotechnol. J.* (2020) doi:10.1016/j.csbj.2020.03.025.

Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *N. Engl. J. Med.* (2020) doi:10.1056/nejmoa2007764.

Blaj, G. *et al.* Performance of ePix10K, a high dynamic range, gain auto-ranging pixel detector for FELs. in *AIP Conference Proceedings* (2019). doi:10.1063/1.5084693.

Braun, F. *et al.* SARS-CoV-2 renal tropism associates with acute kidney injury. *The Lancet* (2020) doi:10.1016/S0140-6736(20)31759-1.

Cavalla, D. Predictive methods in drug repurposing: Gold mine or just a bigger haystack? Drug Discovery Today (2013) doi:10.1016/j.drudis.2012.12.009.

Chang, G. G. Quaternary structure of the SARS coronavirus main protease. in *Molecular Biology of the SARS-Coronavirus* (2010). doi:10.1007/978-3-642-03683-5_8.

Chauhan A., Kalra S., Identification of potent COVID-19 main protease (MPRO) inhibitors from flavonoids. *Preprints* (2020) doi:10.21203/rs.3.rs-34497/v1.

Chen, H., Cheng, F. & Li, J. IDrug: Integration of drug repositioning and drug-target prediction via cross-network embedding. *PLoS Comput. Biol.* (2020) doi:10.1371/journal.pcbi.1008040.

Chen, Y. W., Yiu, C.-P. B. & Wong, K.-Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CLpro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research* (2020) doi:10.12688/f1000research.22457.1.

Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology* (2020) doi:10.1002/jmv.25681.

Choo, O. S. et al. Drugs for hyperlipidaemia may slow down the progression of hearing loss in the elderly: A drug repurposing study. Basic Clin. Pharmacol. Toxicol. (2019) doi:10.1111/bcpt.13150.

Choudhary, S., Malik, Y. S. & Tomar, S. Identification of SARS-CoV-2 Cell Entry Inhibitors by Drug Repurposing Using in silico Structure-Based Virtual Screening Approach. *Front. Immunol.* (2020) doi:10.3389/fimmu.2020.01664.

Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science (80-.).* (2020) doi:10.1126/science.abb4489.

Dinesh, D. C., Tamilarasan, S., Rajaram, K. & Bouřa, E. Antiviral Drug Targets of Single-Stranded RNA Viruses Causing Chronic Human Diseases. Curr. Drug Targets (2019) doi:10.2174/1389450119666190920153247.

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

Durdagi, S., Aksoydan, B., Dogan, B., Sahin, K. & Shahraki, A. Screening of Clinically Approved and Investigation Drugs as Potential Inhibitors of COVID-19 Main Protease: A Virtual Drug Repurposing Study. ChemRxiv (2020).

Dutta, S. & Sengupta, P. SARS-CoV-2 and Male Infertility: Possible Multifaceted Pathology. *Reprod. Sci.* (2021) doi:10.1007/s43032-020-00261-z.

Gao, X. *et al.* Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm. Sin. B* (2021) doi:10.1016/j.apsb.2020.08.014.

Gorbalenya, E., Snijder, E. J. Viral cysteine proteinases. Perspect. *Drug Discov. Des.* (1996), doi:10.1007/BF02174046.

Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* (2020) doi:10.1038/s41586-020-2286-9.

Goyal, B. & Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. ACS Combinatorial Science (2020) doi:10.1021/acscombsci.0c00058.

Ghahremanpour, M. M., *et al*. Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2. *ACS Med. Chem. Lett.* (2020), doi: 10.1021/acsmedchemlett.0c00521.

Heymann, D. L. & Rodier, G. Global Surveillance, National Surveillance, and SARS. *Emerging Infectious Diseases* (2004) doi:10.3201/eid1002.031038.

Huang, F. *et al.* Identification of amitriptyline HCl, flavin adenine dinucleotide, azacitidine and calcitriol as repurposing drugs for influenza A H5N1 virus-induced lung injury. *PLoS Pathog.* (2020) doi:10.1371/journal.ppat.1008341.

Ip, A. et al. Hydroxychloroquine in the treatment of outpatients with mildly symptomatic COVID-19: a multi-center observational study. BMC Infect. Dis. (2021) doi:10.1186/s12879-021-05773-w.

Jarada, T. N., Rokne, J. G. & Alhajj, R. A review of computational drug repositioning: Strategies, approaches, opportunities, challenges, and directions. *Journal of Cheminformatics* (2020) doi:10.1186/s13321-020-00450-7.

Jia, Y. et al. Analysis of the mutation dynamics of SARS-CoV-2 reveals the spread history and emergence of RBD mutant with lower ACE2 binding affinity. bioRxiv (2020) doi:10.1101/2020.04.09.034942.

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

Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. A Bright Future for Serial Femtosecond Crystallography with XFELs. *Trends in Biochemical Sciences* (2017) doi:10.1016/j.tibs.2017.06.007.

Joshi, T. *et al.* In silico screening of natural compounds against COVID-19 by targeting Mpro and ACE2 using molecular docking. *Eur. Rev. Med. Pharmacol. Sci.* (2020) doi:10.26355/eurrev_202004_21036.

Khan, R. J. *et al.* Targeting SARS-CoV-2: a systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2′-O-ribose methyltransferase. *J. Biomol. Struct. Dyn.* (2020) doi:10.1080/07391102.2020.1753577.

Kneller, D. W. *et al.* Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* (2020) doi:10.1038/s41467-020-16954-7.

Kumar, Y., Singh, H. & Patel, C. N. In silico prediction of potential inhibitors for the main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. *J. Infect. Public Health* (2020) doi:10.1016/j.jiph.2020.06.016.

Ma, C. et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. Cell Res. (2020) doi:10.1038/s41422-020-0356-z.

Martin-Garcia, J. M., Conrad, C. E., Coe, J., Roy-Chowdhury, S. & Fromme, P. Serial femtosecond crystallography: A revolution in structural biology. *Arch. Biochem. Biophys.* (2016) doi:10.1016/j.abb.2016.03.036.

McGonagle, D., O'Donnell, J. S., Sharif, K., Emery, P. & Bridgewood, C. Immune mechanisms of pulmonary intravascular coagulopathy in COVID-19 pneumonia. *The Lancet Rheumatology* (2020) doi:10.1016/S2665-9913(20)30121-1.

Menéndez, C. A., Byléhn, F., Perez-Lemus, G. R., Alvarado, W. & de Pablo, J. J. Molecular characterization of ebselen binding activity to SARS-CoV-2 main protease. *Sci. Adv.* (2020) doi:10.1126/sciadv.abd0345.

Muralidharan, N., Sakthivel, R., Velmurugan, D. & Gromiha, M. M. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 protease against COVID-19. *Journal of Biomolecular Structure and Dynamics* (2020) doi:10.1080/07391102.2020.1752802.

Niu, C., Yin, J., Zhang, J., Vederas, J. C. & James, M. N. G. Molecular docking identifies the binding of 3-chloropyridine moieties specifically to the S1 pocket of SARS-CoV Mpro. *Bioorganic Med. Chem.* (2008) doi:10.1016/j.bmc.2007.09.034.

Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet Infectious Diseases* (2020) doi:10.1016/S1473-3099(20)30484-9.

Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S. H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *Journal of Medicinal Chemistry* (2016) doi:10.1021/acs.jmedchem.5b01461.

Pushpakom, S. *et al.* Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery* (2018) doi:10.1038/nrd.2018.168.

Ragab, D., Salah Eldin, H., Taeimah, M., Khattab, R. & Salem, R. The COVID-19 Cytokine Storm; What We Know So Far. *Frontiers in Immunology* (2020) doi:10.3389/fimmu.2020.01446.

Rathnayake, A. D. *et al.* 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV–infected mice. *Sci. Transl. Med.* (2020) doi:10.1126/SCITRANSLMED.ABC5332.

Ratia, K., Kilianski, A., Baez-Santos, Y. M., Baker, S. C. & Mesecar, A. Structural Basis for the Ubiquitin-Linkage Specificity and deISGylating Activity of SARS-CoV Papain-Like Protease. PLoS Pathog. (2014) doi:10.1371/journal.ppat.1004113.

Robertson, D. US FDA approves new class of HIV therapeutics. *Nat. Biotechnol.* (2003), , doi:10.1038/nbt0503-470.

Root-Bernstein, R. Human immunodeficiency virus proteins mimic human T cell receptors inducing cross-reactive antibodies. *Int. J. Mol. Sci.* (2017), doi:10.3390/ijms18102091.

Sierra, R. G. *et al.* The macromolecular femtosecond crystallography instrument at the linac coherent light source. *J. Synchrotron Radiat.* (2019) doi:10.1107/S1600577519001577.

Sies, H., Parnham, M. J. Potential therapeutic use of ebselen for COVID-19 and other respiratory viral infections. *Free Radic. Biol. Med.* (2020), doi:10.1016/j.freeradbiomed.2020.06.032.

Su, E. W. & Sanger, T. M. Systematic drug repositioning through mining adverse event data in ClinicalTrials.gov. PeerJ (2017) doi:10.7717/peerj.3154.

Suárez, D., Díaz, N. SARS-CoV 2 Main Protease: A Molecular Dynamics Study. *J. Chem. Inf. Model.* (2020), doi:10.1021/acs.jcim.0c00575.

Thiel, V. *et al.* Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal of General Virology* (2003) doi:10.1099/vir.0.19424-0.

Ton, A. T., Gentile, F., Hsing, M., Ban, F. & Cherkasov, A. Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Mol. Inform.* (2020) doi:10.1002/minf.202000028.

Tyrrell, D. A. J. & Bynoe, M. L. Cultivation of a Novel Type of Common-cold Virus in Organ Cultures. *Br. Med. J.* (1965) doi:10.1136/bmj.1.5448.1467.

Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorganic and Medicinal Chemistry Letters* (2020) doi:10.1016/j.bmcl.2020.127377.

Van Driel, T. B. et al. The ePix10k 2-megapixel hard X-ray detector at LCLS. J. Synchrotron Radiat. (2020) doi:10.1107/S1600577520004257.

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.* (2020) doi:10.1021/acscatal.0c00110.

Wang, J. Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J. Chem. Inf. Model.* (2020) doi:10.1021/acs.jcim.0c00179.

Węglarz-Tomczak, E., Tomczak, J. M., Talma, M., Brul, S. Ebselen as a highly active inhibitor of PLProCoV2. *bioRxiv* (2020), doi:10.1101/2020.05.17.100768.

World Health Organization, (WHO), "Timeline: WHO's COVID-19 response" (Publication WHO,2020;www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline[the easiest access to this source is via the URL].

Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science (80-.).* (2020) doi:10.1126/science.abb3405.

Zhang, W. *et al.* The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The experience of clinical immunologists from China. *Clinical Immunology* (2020) doi:10.1016/j.clim.2020.108393.

Zhong, N. et al. Without Its N-Finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can Form a Novel Dimer through Its C-Terminal Domain. J. Virol. (2008) doi:10.1128/jvi.02612-07.

Zhou, Y. *et al.* Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* (2020) doi:10.1038/s41421-020-0153-3.

Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology* (2000) doi:10.1099/0022-1317-81-4-853.

Zmudzinski, M. et al. Ebselen derivatives are very potent dual inhibitors of SARS-CoV-2 proteases-PL pro and M pro in in vitro studies. *bioRxiv* (2020) doi.org/10.1101/2020.08.30.273979.

Figures

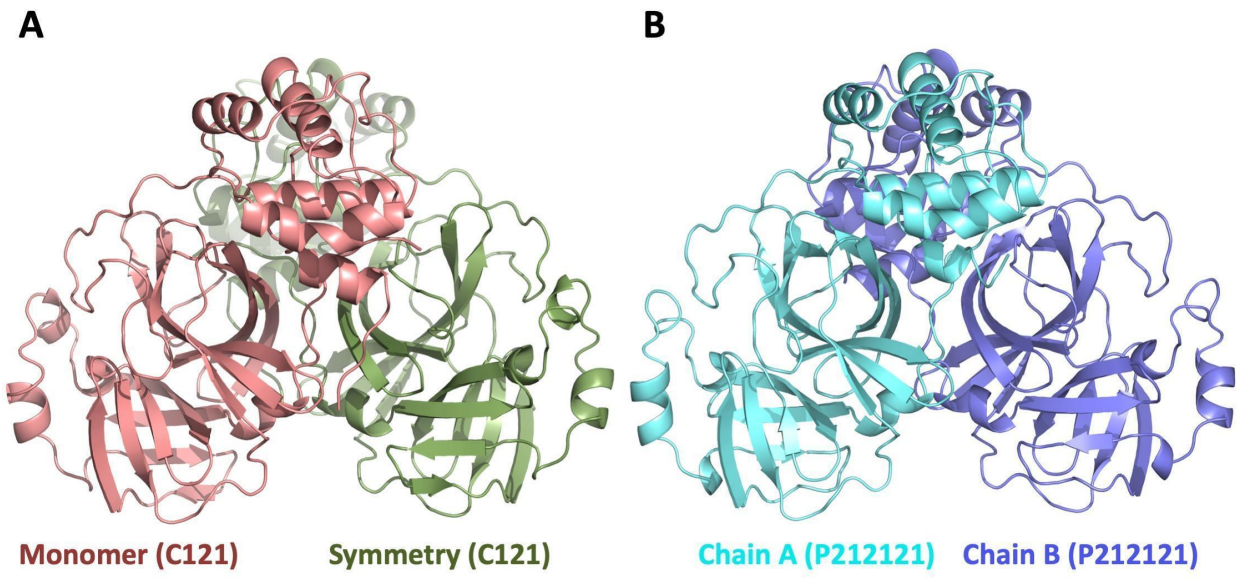

Figure 1 Overview of Mpro crystal structures. a) Cartoon representation of Mpro in space group *C*121. There is one molecule in the asymmetric unit cell colored in dark-salmon (left). Biologically-relevant dimeric form is generated by application of a symmetry operator which is colored in green. **b)** Cartoon representation of the Mpro in space group $P2_12_12_1$ (right). Two chains of the dimer are colored in cyan and slate.

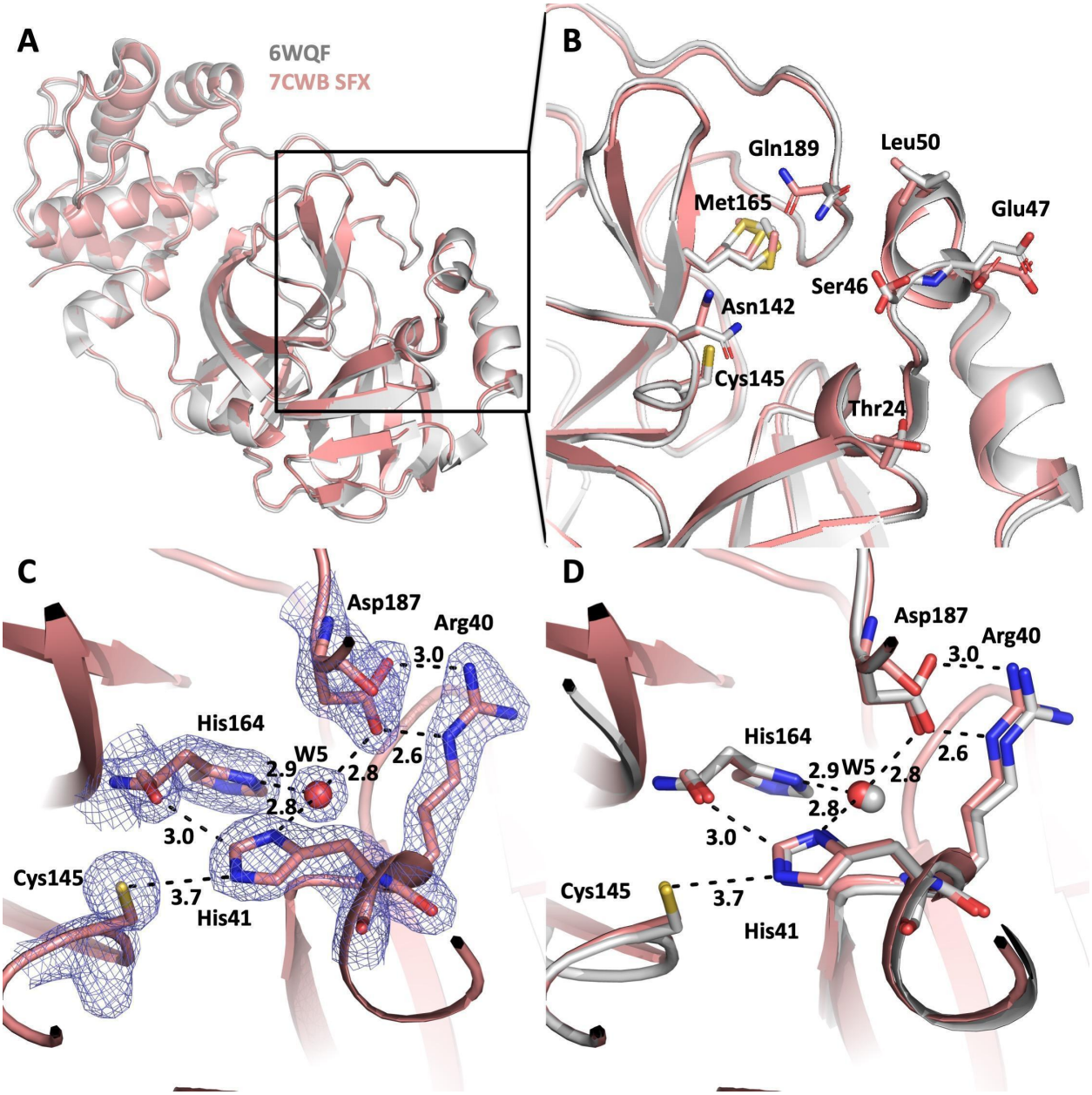

Figure 2 Comparison of SFX structure (PDB ID:7CWB) with ambient temperature structure (PDB ID: 6WQF). a) The two structures align very well with overall RMSD = 0.404 Å SFX structure shown in gray, ambient-temperature structure shown in pink. Same coloring scheme used as in all panels. (All the structural alignments are performed based on the superposition of Cα atoms throughout the manuscript) b) Superposition of the active site pocket reveals the significant conformational states. Residues with altered conformations were labeled and their positions were indicated. c) 2Fo-Fc electron density belonging to the active site residues are contoured in 1 sigma level and colored in slate. H-bonds and other interactions are indicated by dashed lines and distances are given by Angstrom. **d)** Superposition of the 7CWB (SFX) and 6WQF active site reveals very similar states.

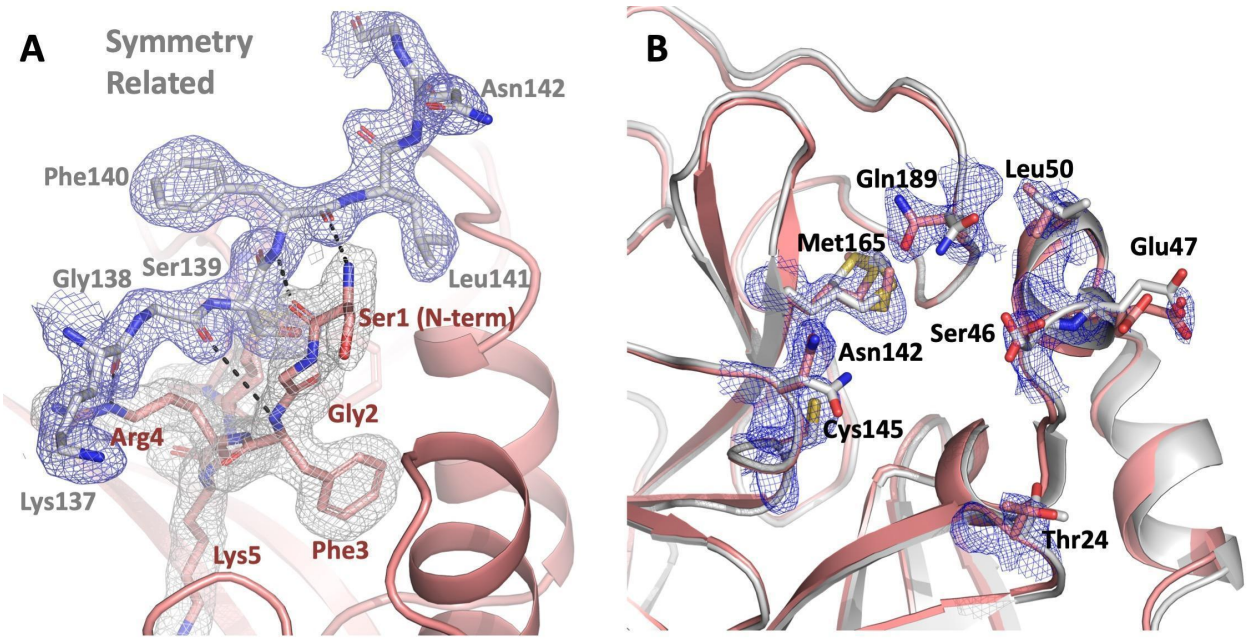

Figure 3 Key crystal contacts of *C*121 crystal form. a) The 2Fo-Fc electron density belongs to the N-terminal region of SARS-CoV-2 main protease contoured at 1 sigma level and colored in gray. Regions of the symmetry related molecule colored in gray, 2Fo-Fc electron density contoured in 1 sigma level and colored in slate. H-bonds are indicated by black dashed lines. Elimination of these H-bonds were essential to obtain the second crystal form in $P2_12_12_1$ space group with more open inhibitor binding pocket for future soaking studies of the putative main protease inhibitors. **b)** Superposition of Cα atoms of our structure (PDB ID: 7CWB) is performed with Mpro structure (PDB ID: 6WQF). Our structure is colored with salmon while chain A of 6WQF is colored in gray. Composite omit map of 7CWB active site is shown in blue color.

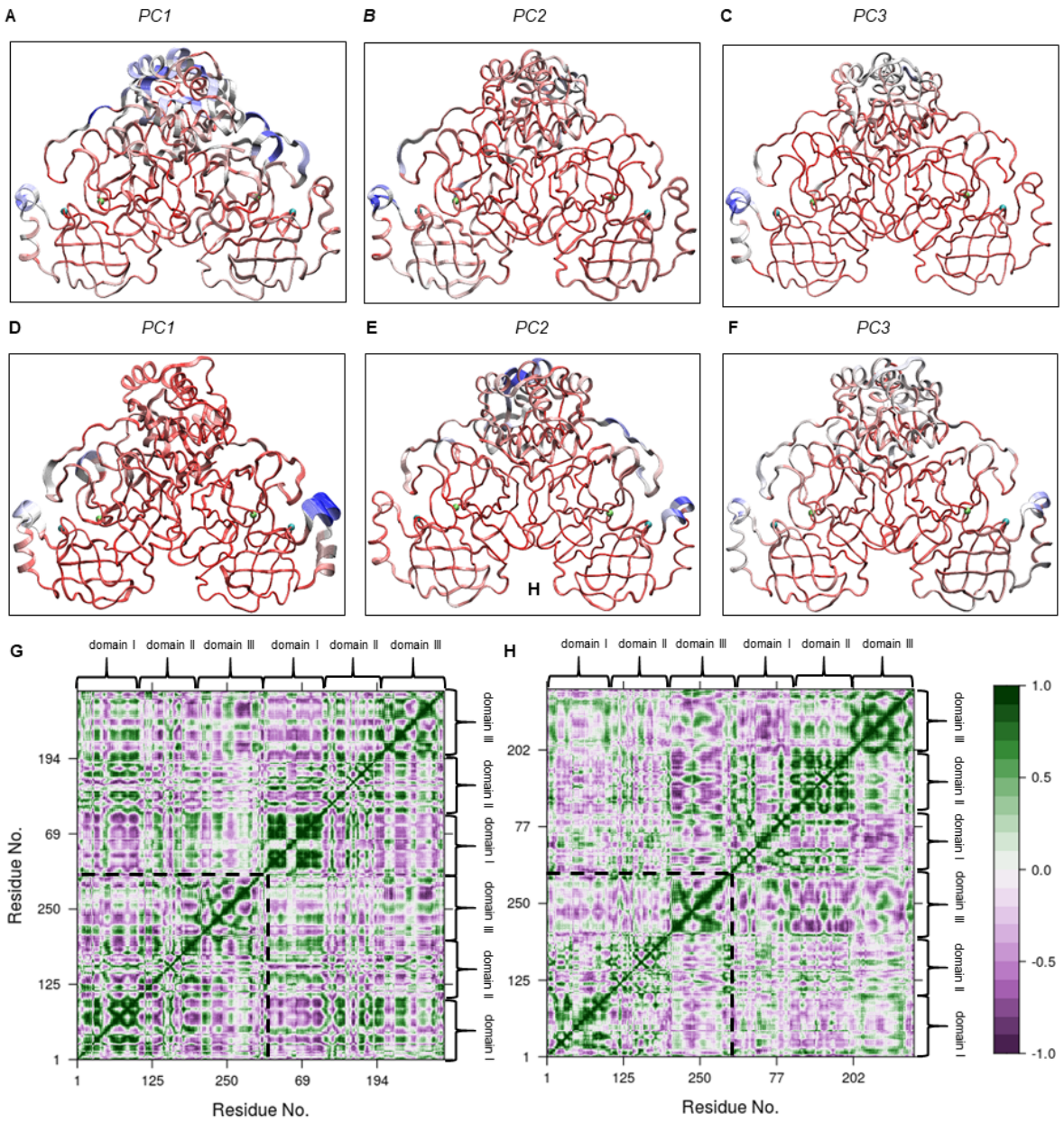

Figure 4 Contributions of residues to the first three PCs. For **a-c)** 7CWB and **d-f)** 7CWC with chain A displayed on the right and chain B on the left side. The aligned trajectory frames were generated to interpolate between the most dissimilar structures in the distribution along specified principal components (PCs). Color scale from red to blue represents low to high displacements along specified PCs with broadening of the tubes depicting the trajectory movements. The Cα atoms of catalytic residues His41 and Cys145 are displayed as spheres with cyan and lime colors, respectively. **g-h)** Dynamic cross correlation matrix generated from the motions observed in PC space with values ranging from -1 (complete anti-correlation) to +1 (complete correlation) **g)** for 7CWB and **h)** for 7CWC. The boundary between chain A and B is denoted by dashed lines.

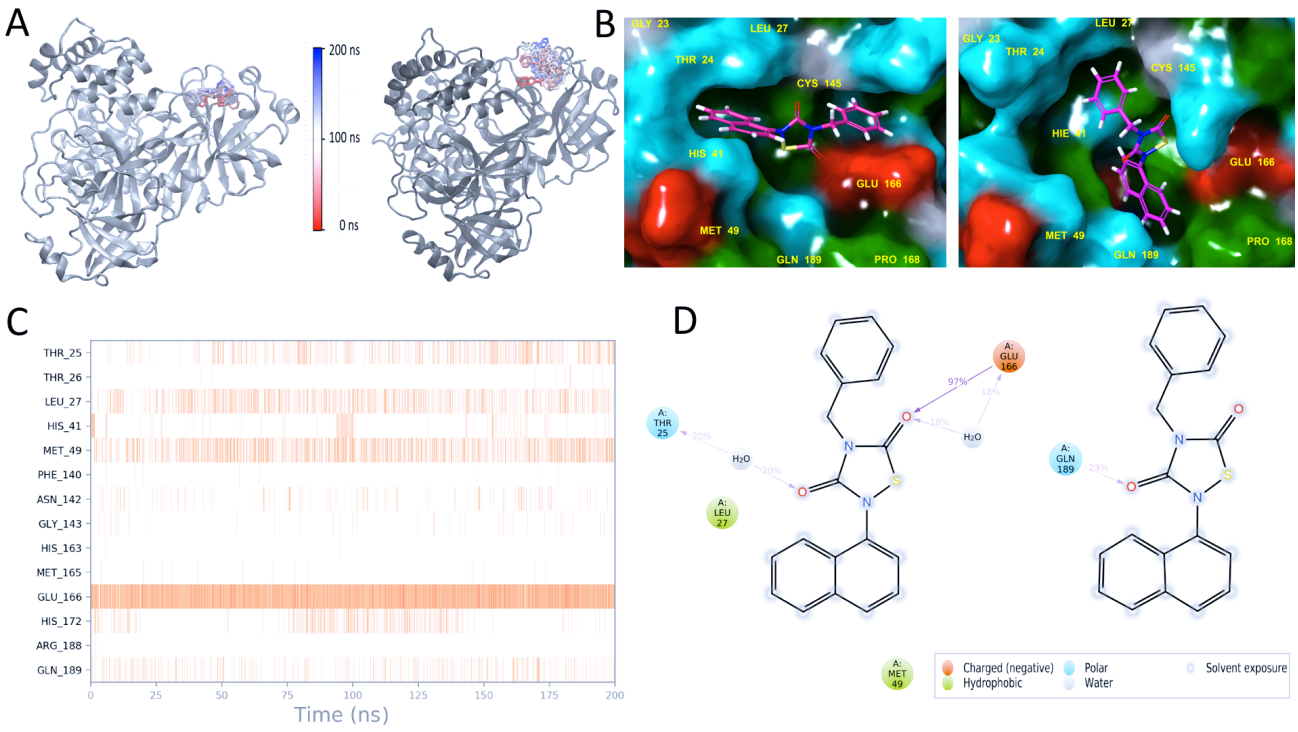

Figure 5 Trajectory analyses of Tideglusib bound 7CWC and 6Y2E PDB coded structures. a) Conformational changes of Tideglusib at the binding pockets of 7CWC (left) and 6Y2E (right) throughout 200 ns MD simulations. (Color scale red-white-blue, red depicts simulation starting time, blue depicts simulation end time, unit nanosecond) **b)** Representative binding poses of Tideglusib at the 7CWC (left) and 6Y2E (right). Residues are colored with amino acid types: red, negative charged; blue, positive-charged; green, hydrophobic; cyan, hydrophilic; **c)** Protein-ligand contacts throughout the MD simulations, 7CWC. The figure shows which residues interact with the ligand in each trajectory frame. Some residues make more than one specific contact with the ligand, which is represented by a darker shade of orange, according to the scale to the right of the plot; **d)** 2D ligand interactions diagrams of Tideglusib, 7CWC (left); 6Y2E (right).

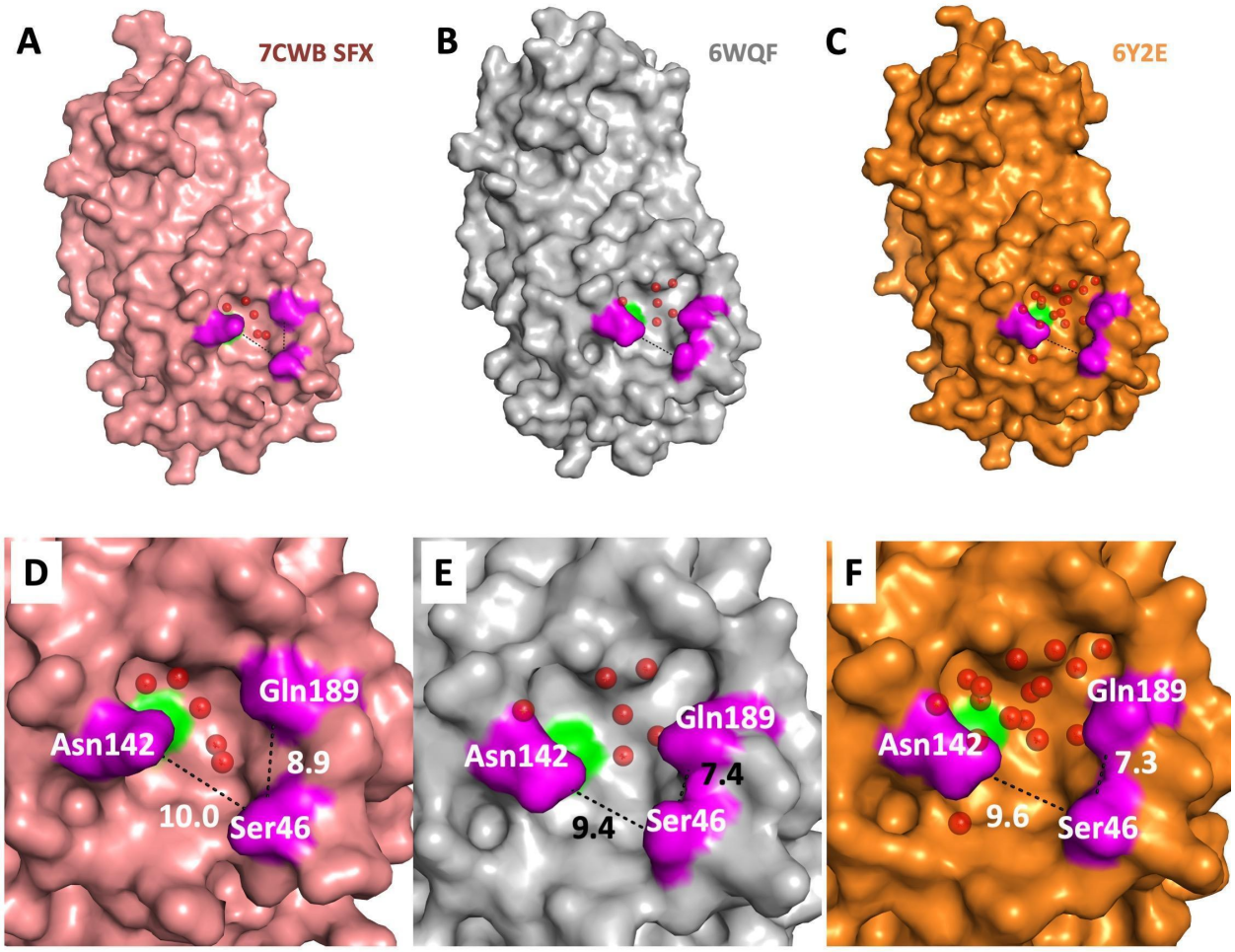

Figure 6 Different monoclinic crystal forms of the SARS-CoV-2 Mpro with catalytic site cavities. **a)** *C*121 SFX, **b)** 6WQF, and **c)** 6Y2E, respectively. Water molecules in the catalytic cavity are shown with red spheres. The catalytic residue (Cys145) is highlighted in green and the flank cavity residues (Asn142, Ser46, Gln189) are highlighted in purple in the catalytic cavity of each crystal form **d)** *C*121 SFX, **e)** 6WQF, and **f)** 6Y2E, respectively. Distance between C (Asn142 Cγ-Ser46 Cβ- Gln189 Cδ) atoms of flank cavity residues is shown as a dashed line. All distances are given in Angstrom (Å).

Supplementary Movie Legends:

Supplementary Movie 1: The motions along the first principal component (PC1) for 7CWB apo structure conformations collected after 200 ns MD simulations. Protomer A is displayed on the right while protomer B is displayed on the left. Domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B.

Supplementary Movie 2: The motions along the second principal component (PC2) for 7CWB apo structure conformations collected after 200 ns MD simulations. Protomer A is displayed on the right while protomer B is displayed on the left. Domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B.

Supplementary Movie 3: The motions along the third principal component (PC3) for 7CWB apo structure conformations collected after 200 ns MD simulations. Protomer A is displayed on the right while protomer B is displayed on the left. Domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B.

Supplementary Movie 4: The motions along the first principal component (PC1) for 7CWC apo structure conformations collected after 200 ns MD simulations. Protomer A is displayed on the right while protomer B is displayed on the left. Domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B.

S**upplementary Movie 5:** The motions along the second principal component (PC2) for 7CWC apo structure conformations collected after 200 ns MD simulations. Protomer A is displayed on the right while protomer B is displayed on the left. Domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B.

Supplementary Movie 6: The motions along the third principal component (PC3) for 7CWC apo structure conformations collected after 200 ns MD simulations. Protomer A is displayed on the right while protomer B is displayed on the left. Domain I is colored in light blue, domain II is colored in cyan and domain III is colored in dark salmon for both protomers A and B.