SARS-CoV-2 M^{pro} responds to oxidation by forming disulfide and NOS/SONOS bonds

Patrick YA Reinke[1,*], Robin Schubert[2,*], Dominik Oberthür[1], Marina Galchenkova[1], Aida Rahmani Mashhour[1], Sebastian Günther[1], Anaïs Chretien[2], Adam Round[2], Brandon Charles Seychell[4], Brenna Norton-Baker[5,13], Chan Kim[2], Christina Schmidt[2], Faisal Koua[2], Alexandra Tolstikova[1], Wiebke Ewert[1], Gisel Esperanza Pena Murillo[1], Grant Mills[2], Henry Kirkwood[2], Hévila Brognaro[12], Huijong Han[2], Jayanath Koliyadu[2], Joachim Schulz[2], Johan Bielecki[2], Julia Lieske[1], Julia Maracke[1], Juraj Knoska[1], Kristina Lorenzen[2], Lea Brings[2], Marcin Sikorski[2], Marco Kloos[2], Mohammad Vakili[1], Patrik Vagovic[1,2], Philipp Middendorf[1], Raphael de Wijn[2], Richard Bean[2], Romain Letrun[2], Seonghyun Han[2,7], Sven Falke[1], Tian Geng[3], Tokushi Sato[2], Vasundara Srinivasan[12], Yoonhee Kim[2], Oleksandr N. Yefanov[1], Luca Gelisio[1], Tobias Beck[4,9], Andrew S. Doré[3], Adrian P Mancuso[2,14], Christian Betzel[12], Saša Bajt[1,9], Lars Redecke[6], Henry Chapman[1, 9, 10], Alke Meents[1], Dušan Turk[8], Winfried Hinrichs[11], Thomas J. Lane[1, 9,#]

* these authors contributed equally to this work
# correspondence: thomas.lane@desy.

1. Center for Free-Electron Laser Science CFEL,
Deutsches Elektronen-Synchrotron DESY,
Notkestr. 85, 22607 Hamburg, Germany

2. European XFEL GmbH,
Holzkoppel 4, 22869 Schenefeld, Germany

3. Sosei Heptares,
Steinmetz Building, Granta Park, Great Abington, Cambridge CB21 6DG, United Kingdom

4. Universität Hamburg,
Department of Chemistry, Institute of Physical Chemistry
Grindelallee 117, 20146 Hamburg, Germany

5. Max Plank Institute for the Structure and Dynamics of Matter,
Luruper Chaussee 149, 22761 Hamburg, Germany

6. Universität zu Lübeck,
Ratzeburger Allee 160, 23562 Lübeck, Germany

7. Gwangju Institute of Science and Technology,
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, Republic of Korea

8. Jožef Stefan Institute,
Jamova cesta 39, 1000 Ljubljana European, Slovenija

9. The Hamburg Centre for Ultrafast Imaging,
Luruper Chaussee 149, 22761 Hamburg, Germany

10. Universität Hamburg,
Department of Physics,
Luruper Chaussee 149, 22761 Hamburg, Germany

11. Universität Greifswald,
Institute of Biochemistry,
Felix-Hausdorff-Str. 4, 17489 Greifswald, Germany

12. Universität Hamburg
Department of Chemistry, Institute of Biochemistry and Molecular Biology,
Laboratory for Structural Biology of Infection and Inflammation,
Build. 22a, c/o DESY, Notkestr. 85, 22607 Hamburg, Germany

13. University of California at Irvine,
Department of Chemistry,
Irvine, CA 92697-2025, USA

14. La Trobe University
Department of Chemistry and Physics, La Trobe Institute for Molecular Science,
Melbourne, Victoria 3086, Australia

Abstract

The main protease (M^{pro}) of SARS-CoV-2 is critical for viral function and is a key drug target. M^{pro} is only active when reduced; turnover ceases upon oxidation but is restored by re-reduction. This suggests the system has evolved to survive periods in an oxidative environment, but the mechanism of this protection has not been confirmed. Here, we report a crystal structure of oxidized M^{pro} showing a disulfide bond between the active site cysteine, C145, and a distal cysteine, C117. Previous work proposed this disulfide provides the mechanism of protection from irreversible oxidation. M^{pro} forms an obligate homodimer, and the C117-C145 structure shows disruption of interactions bridging the dimer interface, implying a correlation between oxidation and dimerization. We confirmed dimer stability is weakened in solution upon oxidation. Finally, we observed the protein's crystallization behavior is linked to its redox state. Oxidized M^{pro} spontaneously forms a new, more loosely packed lattice. Seeding with crystals of this lattice yielded a structure with a novel oxidation pattern incorporating one cysteine-lysine-cysteine (SONOS) and two lysine-cysteine (NOS) bridges. These structures further our understanding of the oxidative regulation of M^{pro} and the crystallization conditions necessary to study this structurally.

INTRODUCTION

During the COVID-19 pandemic, the SARS-CoV-2 main protease (M^{pro}, nsp5 or $3CL^{pro}$) emerged as a key antiviral target and focus of intense study[1],2,3. M^{pro} plays a central role in the SARS-CoV-2 replication cycle, as the viral genome codes for polyproteins that must be cleaved into individual protein units to support viral function. M^{pro} processes at least 11 known sites along polyproteins 1a and 1ab, including its own N- and C-termini[1], and is therefore essential for viral replication. This key role in replication, along with the historical success of viral protease inhibitors, the lack of any similar human protein, and prior work on SARS M^{pro}, has made SARS-CoV-2 M^{pro} the target of several drug discovery programs. These efforts have already yielded an approved molecule, Nirmatrelvir[4]. Given the persistence of the COVID-19 virus and the possible emergence of future pathogenic coronaviruses, it is imperative we develop a deeper understanding of M^{pro} and its role in viral function.

M^{pro}'s activity is regulated by multiple mechanisms, though we have a poor understanding of how these support viral fitness. Most prominently, at sufficiently high concentrations, the enzyme forms a homodimer. Dimerization enhances the catalytic rate, effectively turning M^{pro} from an inactive form into an active one[5]. Structural work suggests this concentration-dependent regulation is not an evolutionary accident. Specifically, M^{pro} adopts a chymotrypsin-like fold, but has a distinct dimerization domain at its C-terminus that many other chymotrypsin-like enzymes lack[6]. Studies of the truncated enzyme lacking this domain, as well as of the domain in isolation, have demonstrated it is both necessary and sufficient for dimer formation[7]. This suggests that this dimerization domain, which is not present in many similar proteases, enables regulation of M^{pro}'s catalytic rate based on the concentration of free enzyme in the cell.

In addition to regulation via dimerization, M^{pro} has been shown to be sensitive to the local redox environment. Including the active site cysteine, the protein sequence contains 12 cysteine residues (~4% in total), an unusually high number[8]. Under mildly reductive conditions all cysteines are reduced, and the protein's catalytic rate is maximized, suggesting this is the active form of the enzyme found in a cellular context[9]. Upon oxidation, a remarkable and growing number of modifications have been reported by both structural and mass spectrometry studies, including glutathionylated C300[10], a peroxy-C145, an N-ethylmaleimide modified C145 and C156[11], a SONOS bridge between C22, C44 and K61[12,13], and a disulfide link between C145 and C117[9].

Given the complexity of viral replication in human hosts, the prevalence or role of these modifications in the natural viral cycle has remained unclear. Oxidative stress in the cell has been shown to regulate the function of other viruses[14], most notably HIV[15–17], and early in the pandemic oxidative stress was hypothesized to play a central role in COVID-19 pathogenesis[18]. It has even been speculated that robustness to oxidative environments might enable corona or other viruses to survive in bat hosts, which are known to exhibit unusual oxidative cellular conditions[10].

The C117-C145 disulfide modification particularly is notable. Funk and colleagues recently performed a systematic study of the behavior of M^{pro} under oxidative conditions and

highlighted this modification as uniquely functional[9]. They produced single point cysteine-to-serine mutants for each cysteine in M^{pro}. Of all these mutants, they found C117S was the only mutant that did not recover activity after being exposed to H_2O_2 and then re-reduced with DTT. This suggests C117 may have a special role in protecting the active site C145 from oxidative damage, but a structural basis for this finding remained lacking.

We determined the structure of M^{pro} with the C117-C145 modification under mildly oxidizing conditions, providing a structural understanding for how this disulfide can protect the enzyme from irreversible oxidation. We found that oxidized protein only crystallized in a more loosely packed, orthorhombic lattice, whereas the reduced protein formed a monoclinic lattice under the same crystallization conditions. Seeding with these orthorhombic crystals enabled us to crystallize M^{pro} exhibiting a previously unobserved set of NOS and SONOS oxidative modifications.

RESULTS

An orthorhombic lattice is flexible enough to produce crystals with oxidative modifications

By delivering streams of microcrystals into the x-ray focus of the SPB/SFX instrument of the European XFEL, we obtained diffraction data yielding two crystal structures of M^{pro}, one active/reduced structure and one inactive/oxidized structure (Table 1). M^{pro} oxidized by air exposure spontaneously crystallized into a different space group and packing as compared to reduced protein, despite being crystallized under the same conditions. Specifically, our reduced crystals formed a monoclinic lattice with C2 symmetry. These crystals contain the native homodimer, with a single protomer in the asymmetric unit and the dimer completed by crystallographic symmetry. Protein subjected to oxidation by exposure to air exhibits a covalent disulfide bond between C117 and C145 and forms crystals in space group $P2_12_12_1$, with the asymmetric unit consisting of the entire homodimer (protomer A-to-B all atom RMSD: 0.96 Å). The orthorhombic lattice exhibits a looser overall packing and higher solvent content (Fig. 1, Table 1). Both datasets were collected at room temperature.

While crystallization conditions for the oxidized and reduced crystals are the same, the lattices obtained differ. This makes a direct comparison of reduced and oxidized structures challenging, as we could not control for differences due to oxidation state *vs.* crystal packing.

Therefore, we attempted to obtain a reduced structure in the orthorhombic lattice seen in our oxidized crystals. By seeding reduced protein with crystals of the oxidized form, we were successfully able to generate crystals of reduced protein in the orthorhombic lattice (Fig. 1). As our XFEL beamtime had concluded by this time, data for these crystals were collected under cryogenic conditions at PETRA III beamline P11 (Table 1). The cryogenic conditions cause a contraction of the lattice and reduction of the solvent content by 4-5% as compared to room temperature collection (Table 1). The molecular structure of the enzyme in the reduced state is similar in both the monoclinic (XFEL/RT) and the orthorhombic (synchrotron/100K) lattices (all atom RMSD: 1.56 Å). Both structures are used here as a basis of comparison to elucidate changes due to oxidation.

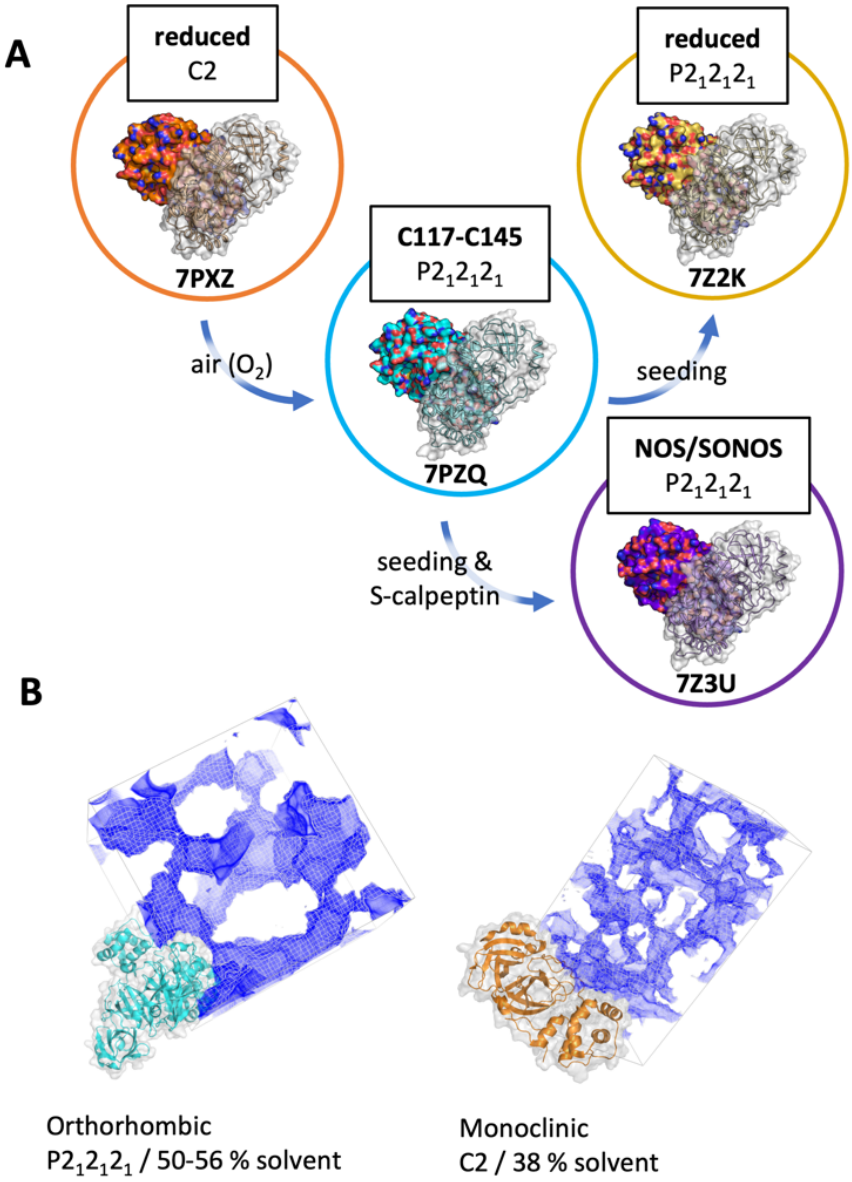


Figure 1. The redox state of M^{pro} is linked to its crystallization behavior. (A) Reduced protein under our crystallization conditions results in monoclinic (C2) protein crystals (PDB 7PXZ), but after oxidation in air, the same protein spontaneously forms crystals with an orthorhombic ($P2_12_12_1$) lattice and exhibit a disulfide link between C117 and C145 (PDB 7PZQ). By producing seed crystals from this oxidized protein, however, we were able to obtain two further structures in the same orthorhombic lattice: first, using reduced protein, an isomorphous structure with reduced C117/C145 (PDB 7Z2K), and second, using the same reduced protein but with the addition of a sulfonated calpeptin ligand, a structure exhibiting NOS and SONOS crosslinks (PDB 7Z3U). (B) Visualization of the orthorhombic and monoclinic lattices, with the solvent content highlighted by *map-channels*[19]. The packing and crystal contact pattern are substantially altered, with the orthorhombic lattice exhibiting significantly larger solvent channels and an overall higher solvent content.

Finally, in conjunction with our ongoing work to develop M^{pro} inhibitors, we employed our oxidized orthorhombic seeds in a co-crystallization experiment with M^{pro} bound to a sulfonated calpeptin derivative[20]. Unexpectedly, the resulting structure exhibits a rich pattern of oxidative modifications. Protomer A contains a SONOS bridge involving C22, C44 and K61, whereas protomer B shows only a NOS bridge involving C22 and K61 at the same site. Both modifications are consistent with previous reports[12,13]. In addition, $2mF_o$-DF_c maps unambiguously show a NOS bridge between K102 and C156 in protomer B, not previously described in the literature, and suggest partial occupancy of the same modification in protomer A.

Table 1. Data collection and refinement statistics

M^{pro}	Reduced Monoclinic*	Oxidized (S-S) Orthorhombic†	Reduced Orthorhombic	Oxidized (NOS/SONOS) Orthorhombic
Data collection	7PXZ	7PZQ	7Z2K	7Z3U‡
Data collection				
Source	EuXFEL	EuXFEL	PETRA-III	PETRA-III
Temperature	297 K	297 K	100 K	100 K
Space group	C2	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions				
a, *b*, *c* (Å)	115.0 54.0 45.0	104.4 104.4 68.7	67.8 101.0 103.9	67.7 99.6 103.261
a, b, g (°)	90 102.0 90	90 90 90	90 90 90	90 90 90
Solvent Content (%)	38.1	56.04	51.34	50.34
Resolution (Å)	31.62-1.75 (1.81-1.75)	24.61-2.25 (2.33-2.25)	67.76-1.65 (1.71-1.65)	49.22-1.72 (1.782-1.72)
R_{sym}[1], R_{split}[2]	0.071 (1.195)[2]	0.169 (3.117)[2]	0.047 (0.576)[1]	0.039 (0.828)[1]
I / sI	11.44 (0.72)	7.41 (0.04)	23.40 (1.26)	12.54 (0.95)
Completeness (%)	99.52 (95.52)	99.88 (99.86)	99.34 (98.46)	99.70 (98.31)
Redundancy	946.0 (24.0)	355.6 (5.4)	6.9 (6.6)	7.5 (7.7)
Refinement				
Resolution (Å)	31.62-1.75 (1.813-1.75)	24.61-2.25 (2.33-2.25)	67.76-1.65 (1.71-1-65)	49.22-1.72 (1.782-1.72)
No. reflections	27225 (2602)	36330 (3592)	85170 (1993)	74542 (7273)
R_{work} / R_{free}	0.1752 / 0.2047	0.1774 / 0.2388	0.1890 / 0.2180	0.1865 / 0.2157
No. atoms	2742	5063	5635	5442
Protein	2498	4819	4966	4911
Ligand/ion	1	12	51	94
Water	729	232	618	437
B-factors				
Protein (Å^2)	39.23	37.58	28.68	40.85
Ligand/ion (Å^2)	28.14	55.23	46.56	53.36
Water (Å^2)	55.27	41.22	35.56	44.02
r.m.s. deviations				
Bond lengths (Å)	0.017	0.008	0.011	0.006
Bond angles (°)	1.90	0.96	1.22	0.92

* Number of crystals merged (serial): 214 954

† Number of crystals merged (serial): 41 771

‡ As reported in [20]

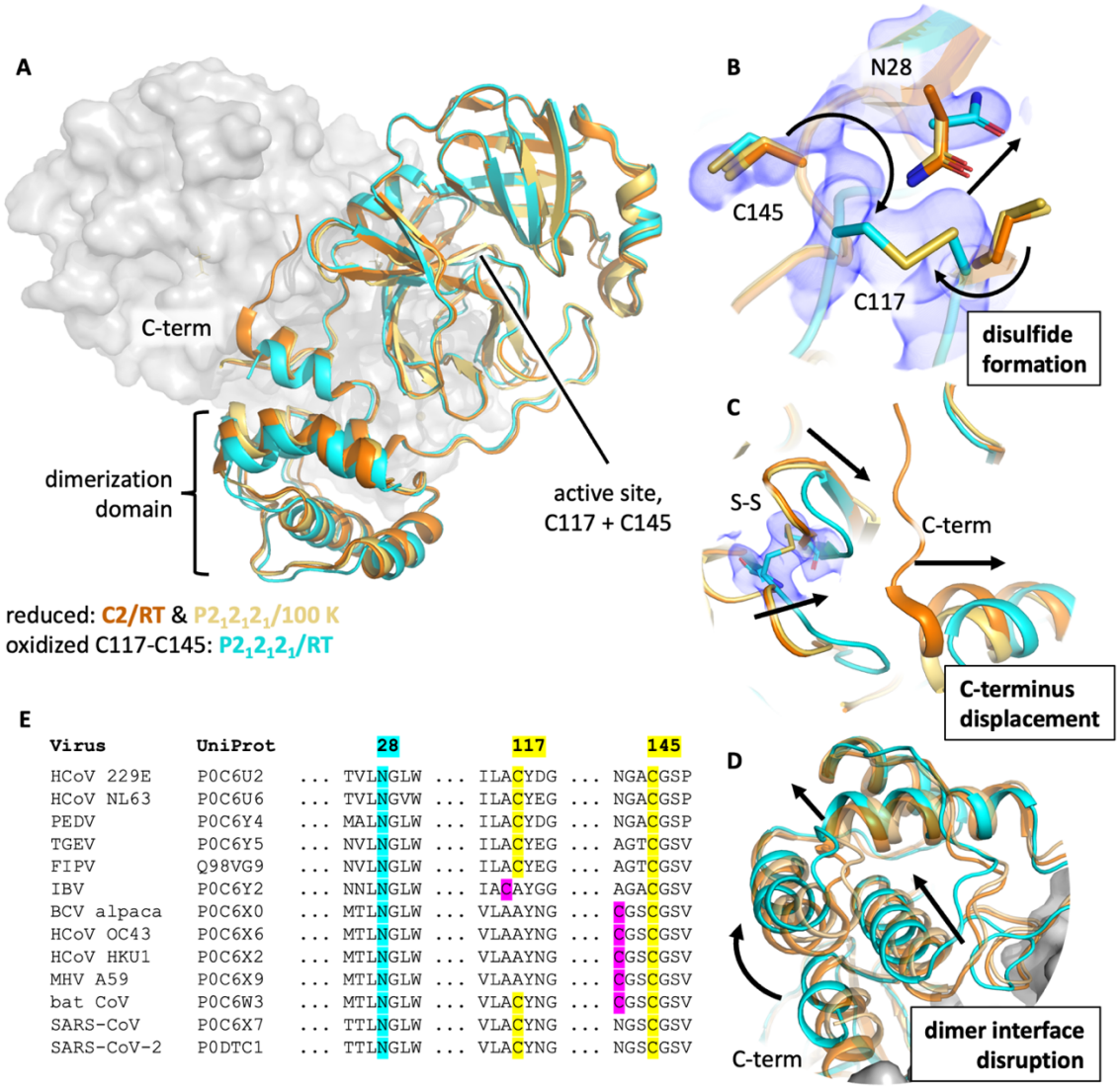


Figure 2. Long-range structural changes correlated with C117-C145 disulfide formation disrupt the dimer interface. (**A**) aligned overlay of reduced (orange: monoclinic/room temperature and yellow: orthorhombic/100 K) and oxidized (cyan: orthorhombic/room temperature) structures, with one monomer of the M^{pro} dimer shown as surface. Oxidation of the active site cysteine, C145, results in (**B**) disulfide bridge formation with C117 and displacement of N28 (density: oxidized $2mF_o$-DF_c at 1 RMSD). Colocalization of C117 and C145 requires (**C**) displacement of C-terminal residues 301-306 from the dimer interface and is correlated with (**D**) a shift of the dimerization domain and disruption of the stabilizing interactions between the two protomers. (**E**) MSA showing N28, C117, and C145 are conserved across related coronaviruses. N28 and C145 are absolutely conserved in the set studied. C117 is partially conserved, but where it is not, another cysteine is present in either position 116 or 142 (magenta) that could conceivably fulfill the same role.

Disulfide formation in M^{pro} precludes catalysis and disrupts the dimer interface

By oxidizing M^{pro} *via* air exposure, we obtained structures with a disulfide bond between C117 and C145. To understand the structural changes that occur upon formation of the C117-C145 disulfide bond (Fig. 2), we determined two reduced reference structures. The first was obtained at room temperature with XFEL radiation, identical to the data collection conditions of our C117-C145 structure, but crystallized in a different space group (C2). The second reference structure is in the same space group as the C117-C145 structure ($P2_12_12_1$), following seeding with crushed oxidized crystals and was obtained at 100 K.

In the reduced form, the catalytically active cysteine C145 sits on a loop in the active site pocket, while C117 forms part of a β-hairpin about 8 Å away (C_α-to-C_α). Oxidative cross-linking of these residues relocates both to a location approximately in the middle of their reduced positions (5.1 Å C_α-to-C_α). This disrupts the β-hairpin motif containing C117 and displaces the conserved N28, which in the reduced structure sits between C145 and C117 but in the oxidized structure undergoes a rotamer shift to make space for the disulfide bridge (Fig. 2). This residue was identified as essential for dimerization and enzymatic activity in SARS-CoV-1 M^{pro}[21]. The rotameric change of N28 was predicted by MD simulations performed by Funk *et al.* and is confirmed by our structures[9]. Disulfide formation partially buries the active site cysteine, which has a solvent exposed area of 24.3 $Å^2$ in the reduced structure but 17.5 $Å^2$ and 13.9 $Å^2$ for the two molecules in the asymmetric unit, A and B respectively, in the oxidized structure.

In protomer A of our oxidized structure, C145 shows residual population at the same position it occupies in the reduced structure. The electron density of this residual population was sufficient to model, resulting in a refined structure with 55% occupancy of the disulfide conformer and 45% population of the reduced conformation. In protomer B, the reduced conformation is insufficiently populated to generate a confident model, and our structure contains a fully occupied disulfide. No evidence of other oxidative modifications was observed in our electron density maps.

The structural rearrangements required to bring C117 and C145 together require a series of long-range structural changes that disrupt the dimer interface (Fig. 2). In the reduced state, the C-termini form part of the dimer interface adjacent to the β-hairpin containing C117, but

upon cross-linking of C117 and C145, C-terminal residues 301-306 become disordered due to the loop rearrangements necessary to bring the two cysteines together.

This ejection of the C-termini from the dimer interface is accompanied by a shift in the entire dimerization domain, which contains contacts that bridge the two protomers. As a result, the entire dimer interface is less well packed in the oxidized structure as compared to its reduced counterpart. The surface area that forms the dimer interface is estimated to be 1301 Å^2 and 1283 Å^2 for protomers A and B, in the orthorhombic reduced structure respectively. This interfacial area decreases to 1198 Å^2 and 1259 Å^2 in the disulfide-containing structure. Notably, the loop formed by residues S284, A285, and L286, which packs tightly with the same residues on the symmetric protomer in the reduced state, is disrupted in the oxidized structure (Fig. 3). In the reduced structure, this loop forms a tight zipper-like packing interface with the opposite protomer, but in the oxidized structure this zipper is out of register and does not form a tight interface (Fig. 3). This disruption of the dimer interface suggests M^{pro}'s dimer affinity is weakened upon oxidation, as recently suggested by Funk *et al.*[9].

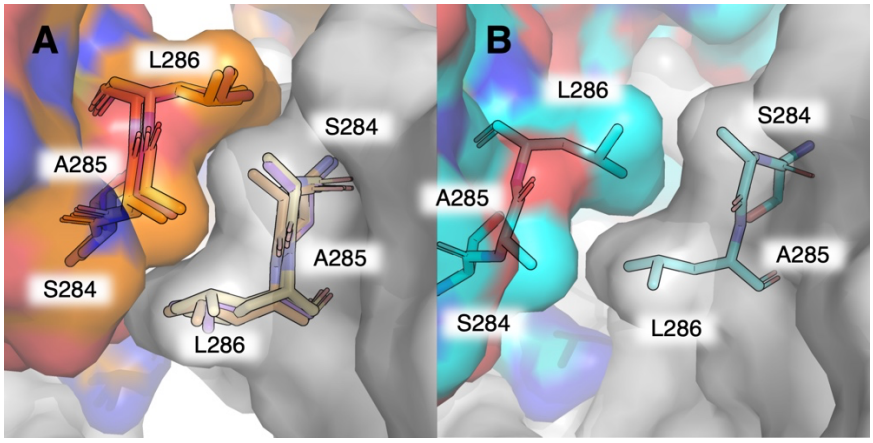
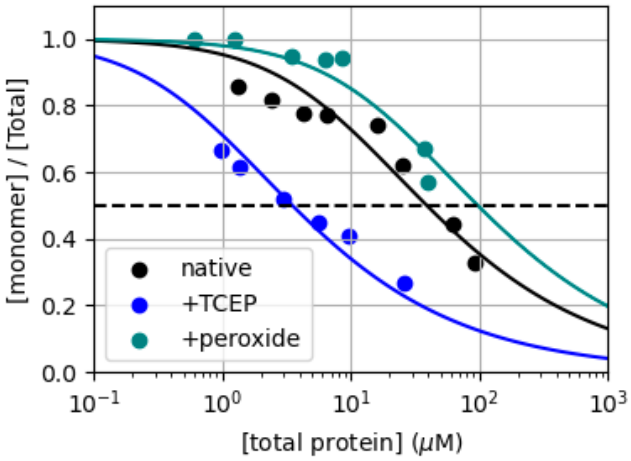


Figure 3. Disruption of the S284/A285/L286 dimerization interface in the disulfide containing structure. Shown is the region where the loop containing S284, A285 and L286 forms a hydrophobic zipper with the same residues on the opposite dimer-forming protomer. All four structures reported are drawn, (**A**) reduced/C2 (orange), reduced/P212121 (yellow), NOS/SONOS (purple) and (**B**) C117-C145 (teal). Only the C117-C145 structure shows a disruption of this dimer interface. For these three residues, the buried surface area decreases from ~110 Å^2 for the structures in (**A**) to 87 and 85 Å^2 for molecules A and B of the C117-C145 structure shown in (**B**), respectively. Structures were aligned by minimizing all heavy atom RMSD prior to visualization. Surfaces are the solvent accessible (Connolly) surface computed with the PyMOL Molecular Graphics System (2.0, Schrödinger LLC)[22].

205

206 **Analytical Size Exclusion Chromatography Confirms Weakened Dimer Affinity**

207 To test if the disruption of the dimer interface observed in our C117-C145 structure translates
208 into a reduction of the dimerization affinity, we performed analytical size exclusion
209 chromatography. In the presence of air, we determined a dimerization dissociation constant
210 (K_D) of 39 µM. The addition of 1 mM TCEP in the running buffer resulted in a higher affinity of
211 3.5 µM, and conversely incubating the protein in 5 mM hydrogen peroxide prior to injection
212 increased the measured K_D to 97 µM, implying that oxidizing conditions decrease the
213 dimerization affinity by an order of magnitude. This agrees with analytical ultracentrifugation
214 performed by Zhang *et al.* (K_D ~2.5 µM, reduced)[2] the SAXS measurements of Silvestrini *et al.*
215 (K_D ~7 µM, reduced)[23] and is qualitatively consistent with analytical ultracentrifugation
216 experiments by Funk *et al.*[9], who determined absolute K_D values that are a factor of 10
217 smaller, but with the same order of magnitude change relative change upon oxidation.
218



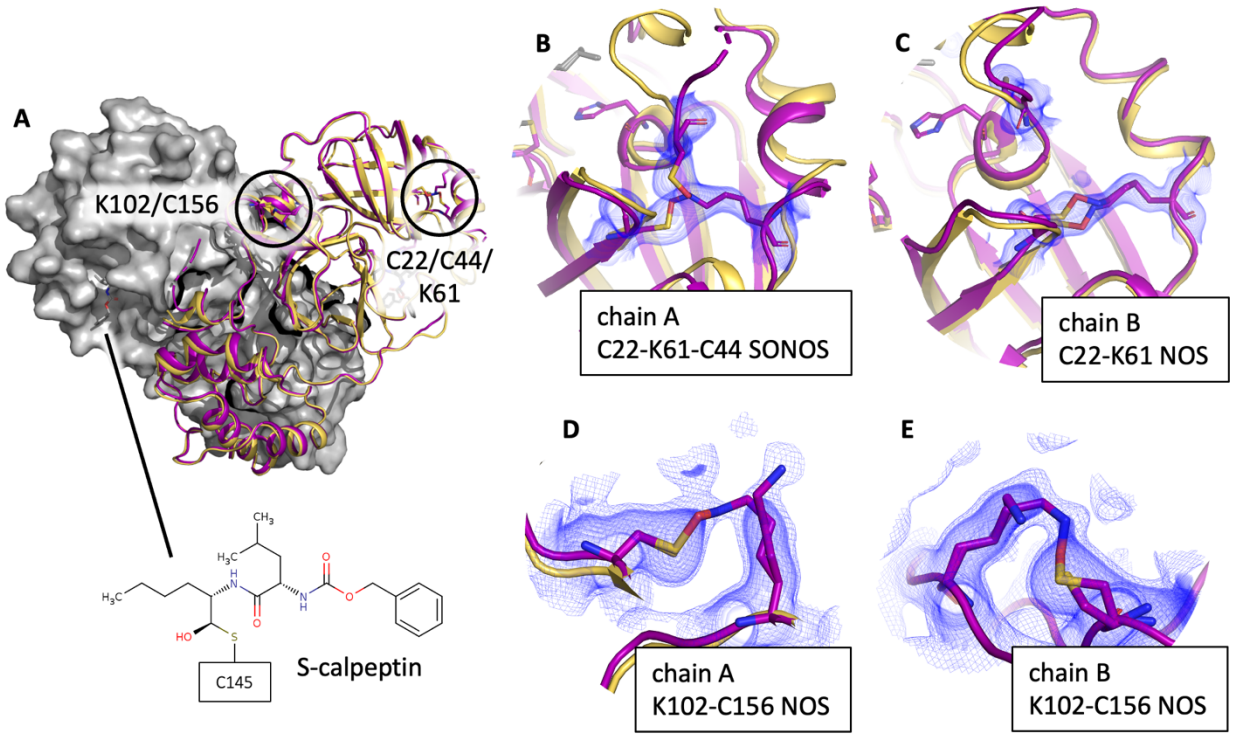
219

220 **Figure 4. Dimer affinity evaluated by analytical size exclusion chromatography.** Size exclusion measurements
221 show that oxidation of M^{pro} by either air (black, K_D 39 +/- 16 µM) or 5 mM peroxide (teal, K_D 97 +/- 43 µM)
222 exhibits significantly weakens the dimer interface as compared to fully reduced protein (blue, K_D of 3.5 +/- 1.1
223 µM, 1 mM TCEP in the running buffer). Errors reported are 95% CIs assuming a Gaussian error model. The x-axis
224 reports the total concentration of single M^{pro} protein chains. The y-axis reports the fraction of monomeric chains.
225

226

227 **Observation of NOS and SONOS modifications upon co-crystallization with a sulfonated**
228 **calpeptin ligand**

During our ongoing structural studies of M^{pro} ligands, we sought to obtain a structure of M^{pro} bound to a ligand of interest, a sulfonated calpeptin derivative that binds covalently to the reactive cysteine in M^{pro}. Co-crystallization attempts of this ligand with reduced, monoclinic seeds failed to yield a high-resolution structure in our hands, instead forming small clusters of crystals that diffracted to low resolution (~5 Å) and could not be indexed. We hypothesized that the looser packing of the orthorhombic lattice provided by our oxidized seeds might better accommodate structural rearrangements caused by ligand binding. Subsequently, we attempted crystallization with our orthorhombic, oxidized seeds and obtained a high-resolution structure clearly showing bound ligand, which we refined against data up to 1.72 Å. Ligand density consistent with full occupancy was present in the active site of both monomers. Unexpectedly, however, the structure shows multiple NOS and SONOS modifications (Fig. 5).



F

Virus	UniProt	22	44	61	102	156
HCoV 229E	P0C6U2	... RVCYGNT ...	HVIASNT ...	YSIMRLH ...	PRHSFRT ...	VFGVNMR ...
HCoV NL63	P0C6U6	... RVCYGST ...	HVIAPST ...	YSTMRLH ...	PKHVFKT ...	VFGVNLR ...
PEDV	P0C6Y4	... RVCYGNM ...	HVIASST ...	LSVLRLH ...	PKYTYRT ...	VYGVNMR ...
TGEV	P0C6Y5	... RVSYGNN ...	HVIASDT ...	MSSVRLH ...	PEHKFKS ...	VYGVNMR ...
FIPV	Q98VG9	... RVAYGNN ...	HVIASDT ...	LSSVRLH ...	PEHKFKS ...	VYGVNMR ...
IBV	P0C6Y2	... SVSYRGN ...	HVLGKF- ...	LNLANNH ...	PKYKFIK ...	LYPVTMR ...
BCV alpaca	P0C6X0	... SVTYGNM ...	HVICSAS ...	LCRVTSS ...	PKYTFGV ...	AFHVTMR ...
HCoV OC43	P0C6X6	... SVTYGNM ...	HVICSAS ...	LCRVTSS ...	PKYTFGV ...	AFHVTMR ...
HCoV HKU1	P0C6X2	... SVTYGSM ...	HVICSSS ...	LCRVTLG ...	PKYTFGN ...	AFHVTMR ...
MHV A59	P0C6X9	... SVTYGNM ...	HVICSSA ...	LCRVTSS ...	PKYSFGV ...	AFHVTLR ...
bat CoV	P0C6W3	... QVTCGSM ...	HIMCPAD ...	LISKTNH ...	PAYTFST ...	VFTVNLR ...
SARS-CoV	P0C6X7	... QVTCGTT ...	HVICTAE ...	LIRKSNH ...	PKYKFVR ...	VYQCAMR ...
SARS-CoV-2	P0DTC1	... QVTCGTT ...	HVICTSE ...	LIRKSNH ...	PKYKFVR ...	VYQCAMR ...

Figure 5. NOS/SONOS crosslinking observed upon co-crystallization with sulfonated calpeptin in the orthorhombic space group. Oxidative NOS and SONOS bridges are seen at four sites in the (**A**) two dimeric protomers that form both the asymmetric and biological unit (all panels, purple: oxidized NOS/SONOS structure in $P2_12_12_1$, yellow: reduced reference in $P2_12_12_1$). (**B**) Protomer A exhibits a SONOS linkage between C22, K61, and C44 that distorts the structure from the reduced form. In contrast, (**C**) protomer B shows only a NOS linkage between C22 and K61 at this site, with the overall structure differing little from the reduced reference. The (**D**) electron density on protomer A between K102 and C156 is ambiguous, consistent with but not conclusively showing a NOS linkage at partial occupancy. In contrast, (**E**) the density at the same site on protomer B clearly shows a NOS bond at partial occupancy. (**F**) None of the residues participating in these linkages are strongly conserved besides C44. Densities shown as blue volumes are the $2mF_o$-DF_c map at 1 RMSD. The same $2mF_o$-DF_c map at 0.5 RMSD is overlaid as a light mesh for the K102-C156 NOS figures, to show partial occupancies more clearly.

The NOS and SONOS modifications exhibit a distinct asymmetry between the monomers that form the dimer in the asymmetric unit. Protomer A shows a SONOS linkage between C22, K61, and C44 (Fig. 5), which has been previously reported[12,13]. This modification results in a shift of the α-helix between E55 and K61 and disrupting the position of a loop between C44 and Y54, which is ordered in the reduced structure but becomes disordered upon SONOS formation. This disorder may be in part because the shifted loop can no longer form a backbone H-bonding contact between M49 and Q189, the latter of which sits on a flexible domain-connecting loop consisting of the residues V186 to G195. This loop is shifted as a result, displaced from its position in the reduced structure.

In contrast, protomer B much more closely resembles the reduced structure. It exhibits a NOS bridge between C22 and K61 (Fig. 5). The effect of this modification is less dramatic, with C22 and K61 separated by 7.2 Å in the reduced structure (C_α-to-C_α) but only 7.4 Å with the NOS bridge present.

Both protomers A and B show evidence for a NOS-bridge between K102 and C156 that connects neighboring β-sheets (Fig. 5). This modification is clear in the density for protomer B, while the density in protomer A is ambiguous but consistent with a NOS bridge at low occupancy. The NOS modification at this site induces essentially no deviation from the reduced structure in the same space group, where K102 and C156 are in close proximity.

DISCUSSION

M^{pro} appears to exhibit an unusually rich set of oxidation modifications, which have been revealed by structural and biochemical methods. While a response to oxidative stress has been implicated in virus biology in general, the possible physiological relevance of each of the observed oxidized states of M^{pro} remains a topic of ongoing investigation.

Our structure of C117-C145 modified M^{pro} provides a mechanistic model for several key observations regarding M^{pro}'s behavior upon change of redox state. Most notably, our structures provide a simple explanation as to why M^{pro}'s dimer affinity decreases by about an order of magnitude upon oxidation[9]. Our structure further confirms a key role of N28, which

rotates to allow space for the C117-C145 disulfide bridge. N28 is highly conserved (Fig. 1), suggesting asparagine at this position is essential for viral fitness[21]. We assume that the small volume and hydrophilic nature of the carboxamide sidechain facilitates this conformational change, enabling M^{pro}'s ability to toggle between reduced and oxidized states.

Our crystals containing the C117-C145 disulfide were studied using XFEL light. We speculate this may have allowed us to observe this modification clearly via "radiation damage free" data collection, as the x-ray exposure (~100 fs) is much more rapid than the nuclear motions required for the two cysteine sidechains to adopt significantly different positions following x-ray induced reduction[24–26].

Funk and colleagues reported that M^{pro} C117S was the only C-to-S mutant that failed to recover activity after exposure to oxidative conditions followed by reduction. Our structures illustrate how, upon oxidation, the catalytic C145 moves from a solvent exposed conformation to a buried, disulfide conformation. Our structure, alongside these previous findings and the conserved nature of cysteines at positions 117 and 145, implicate this modification in a regulatory response to an oxidative environment. We speculate the C117-C145 disulfide provides a protective mechanism against oxidative damage by making harsher, irreversible oxidation to sulfinic and sulfonic acids impossible.[27]

Seeding with crystals containing the C117-C145 disulfide enables kinetic control over the crystallization lattice, allowing us to obtain a novel ligand-bound structure that shows NOS and SONOS bridges. We considered the hypothesis that ligand binding might facilitate these modifications. Yang and colleagues, however recently presented nine M^{pro} structures exhibiting the C22-K61-C44 SONOS bond. Five contain bound inhibitor but four show no ligand of interest[13], demonstrating that ligand binding is not necessary to observe SONOS modifications in M^{pro} crystals. Further, the NOS bond at K102-C156, which has not been reported previously, is far from the active site and seems unlikely to be influenced by the binding of a ligand in that pocket. As no oxidizing agents were added to the crystallization experiment, we attribute NOS/SONOS formation to molecular oxygen introduced by air exposure. The frequency and diversity of NOS/SONOS modifications observed in Mpro suggest these crosslinks may have a functional role in regulating the enzyme's function in

oxidative environments. The possible contribution of these modifications to coronavirus fitness warrants further investigation.

CONCLUSIONS

M^{pro} is a linchpin of coronavirus biology and the premier target for anti-COVID-19 small molecule therapeutics. The enzyme's function has been shown to be regulated via both dimerization and oxidation; further, these regulatory mechanisms are biophysically correlated. While our structures provide mechanistic insight into these properties of M^{pro}, we must now understand how regulation based on oxidative stress or protein concentration impact viral fitness. This will provide deeper insight into viral biology and hopefully open new opportunities to disrupt that biology with life-preserving medicines.

Acknowledgments

We acknowledge T. White for assistance during the XFEL experiment and C. Uetrecht for valuable discussions. TJL was supported by a Helmholtz young investigator award. PYAR was supported by the Helmholtz society through the projects FISCOV, SFragX and the Helmholtz Association Impulse and Networking funds InternLabs-0011 'HIR3X'. We acknowledge financial support obtained from the Cluster of Excellence 'Advanced Imaging of Matter' of the Deutsche Forschungsgemeinschaft (DFG) - EXC 2056 - project ID 390715994 and BMBF via projects 05K19GU4 and 05K20GUB. We acknowledge European XFEL in Schenefeld, Germany, for provision of x-ray free-electron laser beamtime at SPB/SFX and would like to thank the staff for their assistance. Sample reservoirs and the anti-settling device employed in parts of the measurements presented here were designed and fabricated by the Max Planck Institute for Medical Research, Heidelberg, which also provided instruction in its use. This research was supported through computational resources (Maxwell cluster) and experimental facilities (PETRA III beamline P11) operated by Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany, a member of the Helmholtz Association HGF. We acknowledge the P11 staff for their invaluable help.

Methods

Protein production and purification. The protein was overexpressed in *E. coli* and purified for subsequent crystallization using previously published protocols and plasmid constructs[2]. Briefly, cell pellets containing overexpressed protein were lysed in 20 mM TRIS buffer, pH 7.8, supplemented with 150 mM NaCl and 10 mM imidazole using a homogenizer. After removal of insoluble cell matter by ultracentrifugation, a nickel NTA column was used to purify the M^{pro} -histidine-tag fusion protein. Following imidazole elution, the protein buffer was changed to 20 mM TRIS, pH 7.8, 150 mM NaCl, 1 mM TCEP using a PD10 column and the histidine tag was cleaved by 3C protease overnight. Subsequently, the histidine tag and the 3C protease were removed using a nickel NTA column. For the reduced form of M^{pro} a final size exclusion chromatography was performed with an S200 Superdex column using 20 mM TRIS, pH 7.8, 150 mM NaCl, 1 mM TCEP and 1 mM EDTA, while for the oxidized form TCEP was omitted.

Crystallization experiments. M^{pro} microcrystals were grown using seeded batch crystallization in the XBI laboratories of the European XFEL[28]. The initial seed stock was produced by adding M^{pro} protein crystals to a reaction tube containing a glass bead (Beads-for-Seeds, Jena Bioscience) and vortexing periodically for 5 seconds with subsequent incubation at room temperature. For the microcrystal batch crystallization, a volume of 250 µL glass seed beads were used in a 1.5 mL reaction tube. 900 µL precipitant solution (25 % PEG1500, 0.1 M MIB buffer pH 7.5, 5 % DMSO) were mixed with 100 µL seed stock and 100 µL M^{pro} protein solution (35 mg/ml). Crystals were grown in a shaker at 18°C at 900 rpm overnight. Resulting crystals were thin plates with a size ranging from 3-15 µm. Crystal concentration was adjusted by allowing the crystals to settle overnight and removing supernatant accordingly. Final crystal slurry was filtered through a 30 µm mesh gravity filter (Sysmex CellTrics) before injection.

Protein crystals for single crystal rotation experiments were produced as previously reported[29], using orthorhombic seeds and reduced protein at 6.25 mg/mL. For the ligand free and S-Calpeptin containing crystallization experiments, the same reduced protein batch was used. The S-Calpeptin compound was dried in the well prior to crystallization mixture addition, yielding a maximum concentration of 5 mM.

Instrumentation. SFX experiments (7PXZ, 7PZQ) were performed at the SPB/SFX instrument[30] in April 2021 as a part of proposal 2696. The size of the mirror-focused focal spot in the interaction region was estimated to be 4×4 µm^2 FWHM diameter based on optical imaging of single shots using a 20 µm thick Ce:YAG screen. The x-ray pulse energy was in the range of 1.2 - 3.5 mJ at 9.3 keV. Diffraction from the sample was measured using an AGIPD[31] of 1 megapixel located 117.7–118.6 mm downstream of the sample interaction region, with the unused direct beam passing through a central hole in the detector to a beam stop further downstream. The resolution at the edge of the AGIPD was 1.8 Å, and 1.6-Å data were obtained by integrating Bragg reflections into the detector corner. Experiment control was provided by Karabo[32].

We used double-flow focusing nozzles (DFFN) for sample delivery[33,34]. The DFFN had an inner diameter of 75 µm and a liquid jet was established by applying 35 mg/min helium flow, 25 µl/min ethanol flow and 15 - 20 µl/min sample flow. We measured the jet diameter to be about 4.5 µm, with a flow rate of 40 - 45 µl/min under identical conditions to those used for the experiment. This translates into a jet speed of approximately 43 m/s[35]. During injection, sample was at room temperature, approximately 20°C.

Rotation experiments (7Z2K, 7Z3U) were performed at PETRA-III beamline P11, delivering a 100 µm beam of 12 keV x-rays focused by a paired KB mirror system exhibiting 30% transmission[36]. Crystals were mounted robotically on a single-axis goniometer and held at 100 K using a cryojet (Oxford). During data collection, samples were rotated 200 degrees with frames read out from a DECTRIS Eiger detector at a distance of 200 mm every 0.2 degrees, for a total of 1000 images per crystal. Total dose per collection was approximately 1.05 MGy as determined by a calibrated diode measurement of x-ray flux ($0.7 \bullet 10^{12}$ ph/s at 100% transmission).

Data analysis. During SFX experiments, online monitoring of the running experiment was performed with Karabo[32] and OnDA[37]. The AGIPD geometry was refined against lysozyme data taken at the beginning and end of every shift. Preprocessing of images was performed with Cheetah[38] and subsequent crystallographic analysis was done with CrystFEL v0.9.1[39]. MOSFLM was used for preliminary indexing[40], but all reported results used xgandalf[41]. Serial data merging was performed with partialator using the unity model. Data from rotation

experiments with single crystals were processed with XDS[42]. All surface area calculations were performed with PISA[43].

Structure determination. Structures were determined by iterative rounds of model building in Coot[44] and refinement with phenix.refine[45], after molecular replacement using PDB ID 7AR6 as a search model. Disulfide, NOS, and SONOS bonds were generated with phenix using refinement geometry restraints.

Analytical SEC. M^{pro} was prepared in 20 mM Tris (pH 7.8) buffer supplemented with 150 mM NaCl and 1 mM EDTA, and for reduced samples, 1 mM TCEP. Compounds were added to a final concentration of 5 mM and 5% DMSO and incubated for 5 hours. Then, protein solutions were spun down at 16,000 g for 5 minutes and applied to a Cytiva Superdex 75 10/300 increase column using a ÄKTA Pure system from Cytiva. Two peaks are observed in the resulting chromatograms at elution volumes consistent with dimer and monomer species. Relative populations were quantified by fitting a Gaussian to each and integrating the area under the curve.

Code availability. The versions of Cheetah and CrystFEL used in this work are available from the respective websites: https://www.desy.de/~barty/cheetah and https://www.desy.de/~twhite/crystfel.

Data availability. Structural models, structure factor data, and associated metadata are available from the Protein DataBank under PDB IDs 7PXZ, 7PZQ, 7Z2K & 7Z3U. Raw data and processing scripts available upon request.

References

1. Hu, Q. *et al.* The SARS-CoV-2 main protease (Mpro): Structure, function, and emerging therapies for COVID-19. *Med. Comm.* **3:e151**, (2022).
2. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science (1979)* **368**, 409–412 (2020).
3. Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science (1979)* **368**, 1331–1335 (2020).
4. U.S. Food and Drug Administration (FDA). Coronavirus (COVID-19) update: FDA authorizes first oral antiviral for treatment of COVID-19. *Food and Drug Administration* 1 (2021).
5. Ding, L., Zhang, X. X., Wei, P., Fan, K. & Lai, L. The interaction between severe acute respiratory syndrome coronavirus 3C-like proteinase and a dimeric inhibitor by capillary electrophoresis. *Anal Biochem* **343**, (2005).
6. Anand, K. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. *EMBO Journal* **21**, 3213–3224 (2002).
7. Zhong, N. *et al.* C-terminal domain of SARS-CoV main protease can form a 3D domain-swapped dimer. *Protein Science* **18**, (2009).
8. Miseta, A. & Csutora, P. Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Mol Biol Evol* **17**, 1232–1239 (2000).
9. Funk, L.-M. *et al.* Redox regulation of the SARS-CoV-2 main protease provides new opportunities for drug design. *bioRxiv* 2022.04.18.487732 (2022).
10. Davis, D. A. *et al.* Regulation of the Dimerization and Activity of SARS-CoV-2 Main Protease through Reversible Glutathionylation of Cysteine 300. *mBio* **12**, (2021).
11. Kneller, D. W. *et al.* Room-temperature X-ray crystallography reveals the oxidation and reactivity of cysteine residues in SARS-CoV-2 3CL Mpro: Insights into enzyme mechanism and drug design. *IUCrJ* **7**, 1028–1035 (2020).

12. Rabe von Pappenheim, F. *et al.* Widespread occurrence of covalent lysine–cysteine redox switches in proteins. *Nature Chemical Biology 2022 18:4* **18**, 368–375 (2022).
13. Yang, K. S. *et al.* A Novel Y-Shaped, S-O-N-O-S-Bridged Cross-Link between Three Residues C22, C44, and K61 Is Frequently Observed in the SARS-CoV-2 Main Protease. *ACS Chem Biol* (2022) doi:10.1021/acschembio.2c00695.
14. Schwarz, K. B. Oxidative stress during viral infection: A review. *Free Radic Biol Med* **21**, 641–649 (1996).
15. Davis, D. A. *et al.* Regulation of HIV-1 protease activity through cysteine modification. *Biochemistry* **35**, 2482–2488 (1996).
16. Davis, D. A. *et al.* Reversible Oxidative Modification as a Mechanism for Regulating Retroviral Protease Dimerization and Activation. *J Virol* **77**, 3319–3325 (2003).
17. Daniels, S. I. *et al.* The initial step in human immunodeficiency virus type 1 gagpropol processing can be regulated by reversible oxidation. *PLoS One* **5**, (2010).
18. Cecchini, R. & Cecchini, A. L. SARS-CoV-2 infection pathogenesis is related to oxidative stress as a response to aggression. *Med Hypotheses* **143**, (2020).
19. Juers, D. H. & Ruffin, J. MAP-CHANNELS: A computation tool to aid in the visualization and characterization of solvent channels in macromolecular crystals. *J Appl Crystallogr* **47**, (2014).
20. Reinke, P. *et al.* Calpeptin is a potent cathepsin inhibitor and drug candidate for SARS-CoV-2 infections. *Res Sq* (2023) doi:10.21203/rs.3.rs-2450926/v1.
21. Barrila, J., Gabelli, S. B., Bacha, U., Amzel, L. M. & Freire, E. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CLpro. *Biochemistry* **49**, 4308–4317 (2010).
22. Delano, W. L. The PyMOL Molecular Graphics System. *CCP4 Newsletter on protein crystallography* **40**, (2002).
23. Silvestrini, L. *et al.* The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci Rep* **11**, (2021).
24. Caleman, C., Junior, F. J., Grånäs, O. & Martin, A. V. A perspective on molecular structure and bonding-breaking in radiation damage in serial femtosecond crystallography. *Crystals (Basel)* **10**, (2020).
25. Chapman, H. N., Caleman, C. & Timneanu, N. Diffraction before destruction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, (2014).
26. Nass, K. Radiation damage in protein crystallography at X-ray free-electron lasers. *Acta Crystallogr D Struct Biol* **75**, 211–218 (2019).
27. Ulrich, K. & Jakob, U. The role of thiols in antioxidant systems. *Free Radical Biology and Medicine* vol. 140 Preprint at https://doi.org/10.1016/j.freeradbiomed.2019.05.035 (2019).
28. Han, H. *et al.* The XBI BioLab for life science experiments at the European XFEL. *urn:issn:1600-5767* **54**, 7–21 (2021).
29. Günther, S. *et al.* X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science (1979)* **372**, 642–646 (2021).
30. Mancuso, A. P. *et al.* The Single Particles, Clusters and Biomolecules and Serial Femtosecond Crystallography instrument of the European XFEL: initial installation. *urn:issn:1600-5775* **26**, 660–676 (2019).
31. Henrich, B. *et al.* The adaptive gain integrating pixel detector AGIPD a detector for the European XFEL. *Nucl Instrum Methods Phys Res A* **633**, 1–4 (2011).

32. Fangohr, H. *et al.* Data Analysis Support in Karabo at European XFEL. *16th Int. Conference on Accelerator and Large Experimental Control Systems* 245–252 (2018) doi:10.18429/JACOW-ICALEPCS2017-TUCPA01.
33. Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Scientific Reports 2017 7:1* **7**, 1–12 (2017).
34. Knoška, J. *et al.* Ultracompact 3D microfluidics for time-resolved structural biology. *Nat Commun* **11**, (2020).
35. Vakili, M. *et al.* 3D printed devices and infrastructure for liquid sample delivery at the European XFEL. *J Synchrotron Radiat* **29**, (2022).
36. Burkhardt, A. *et al.* Status of the crystallography beamlines at PETRA III. *Eur Phys J Plus* **131**, (2016).
37. Mariani, V. *et al.* OnDA: Online data analysis and feedback for serial X-ray imaging. *J Appl Crystallogr* **49**, 1073–1080 (2016).
38. Barty, A. *et al.* Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J Appl Crystallogr* **47**, 1118–1131 (2014).
39. White, T. A. *et al.* CrystFEL: A software suite for snapshot serial crystallography. *J Appl Crystallogr* **45**, 335–341 (2012).
40. Leslie, A. G. W. & Powell, H. R. Processing diffraction data with mosflm. 41–51 (2007) doi:10.1007/978-1-4020-6316-9_4.
41. Gevorkov, Y. *et al.* XGANDALF - extended gradient descent algorithm for lattice finding. *Acta Crystallogr A Found Adv* **75**, 694–704 (2019).
42. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132 (2010).
43. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* **372**, (2007).
44. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
45. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).