

# A new inactive conformation of SARS-CoV-2 main protease

Emanuele Fornasier,<sup>a</sup> Maria Ludovica Macchia,<sup>b</sup> Gabriele Giachin,<sup>a</sup> Alice Sosic,<sup>b</sup> Matteo Pavan,<sup>c</sup> Mattia Sturlese,<sup>c</sup> Cristiano Salata,<sup>d</sup> Stefano Moro,<sup>c</sup> Barbara Gatto,<sup>b</sup> Massimo Bellanda<sup>a,e</sup> and Roberto Battistutta<sup>a,e\*</sup>

Received 17 November 2021

Accepted 27 January 2022

Edited by G. Cingolani, Thomas Jefferson University, USA

**Keywords:** SARS-CoV-2; main protease; COVID-19; M<sup>Pro</sup>; crystal structure; inactive conformation.

**PDB reference:** SARS-CoV-2 main protease in a novel conformational state, 7nij

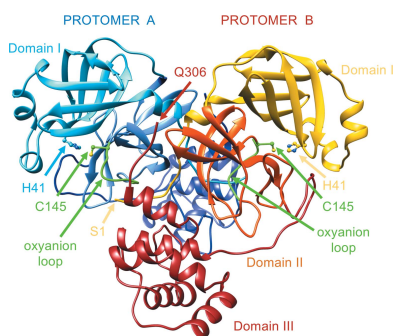
**Supporting information:** this article has supporting information at journals.iucr.org/d

<sup>a</sup>Department of Chemical Sciences, University of Padua, Via F. Marzolo 1, 35131 Padova, Italy, <sup>b</sup>Department of Pharmaceutical and Pharmacological Sciences, University of Padua, Via F. Marzolo 5, 35131 Padova, Italy, <sup>c</sup>Molecular Modeling Section, Department of Pharmaceutical and Pharmacological Sciences, University of Padua, Via F. Marzolo 5, 35131 Padova, Italy, <sup>d</sup>Department of Molecular Medicine, University of Padua, Via Gabelli 63, 35121 Padova, Italy, and <sup>e</sup>Institute of Biomolecular Chemistry of CNR, Padua Unit, Via F. Marzolo 1, 35131 Padova, Italy. \*Correspondence e-mail: roberto.battistutta@unipd.it

The SARS-CoV-2 main protease (M<sup>Pro</sup>) has a pivotal role in mediating viral genome replication and transcription of the coronavirus, making it a promising target for drugs against the COVID-19 pandemic. Here, a crystal structure is presented in which M<sup>Pro</sup> adopts an inactive state that has never been observed before, called new-inactive. It is shown that the oxyanion loop, which is involved in substrate recognition and enzymatic activity, adopts a new catalytically incompetent conformation and that many of the key interactions of the active conformation of the enzyme around the active site are lost. Solvation/desolvation energetic contributions play an important role in the transition from the inactive to the active state, with Phe140 moving from an exposed to a buried environment and Asn142 moving from a buried environment to an exposed environment. In new-inactive M<sup>Pro</sup> a new cavity is present near the S2' subsite, and the N-terminal and C-terminal tails, as well as the dimeric interface, are perturbed, with partial destabilization of the dimeric assembly. This novel conformation is relevant both for comprehension of the mechanism of action of M<sup>Pro</sup> within the catalytic cycle and for the successful structure-based drug design of antiviral drugs.

## 1. Introduction

To face the global COVID-19 pandemic, besides prevention via the use of vaccines, it is also essential to develop targeted therapeutic options for patients infected by the SARS-CoV-2 betacoronavirus. In general, one of the most promising classes of antiviral drug candidates are protease inhibitors, small molecules that are able to inhibit enzymes involved in virus replication within the cell. Very low sequence identity with human proteases and distinct cleavage-site specificities suggest that viral enzymes can be inhibited with very low associated toxic effects ('off-target' effects), if any. Indeed, protease inhibitors have already been efficient in the treatment of viral pathogens such as hepatitis C virus (Pol & Corouge, 2014) and human immunodeficiency virus (HIV; Skwarecki *et al.*, 2021). In coronaviruses, the main protease, M<sup>Pro</sup>, is a cysteine peptidase that is essential for the replication cycle of positive-sense, single-stranded RNA coronaviruses (Xia & Kang, 2011), including SARS-CoV-2. It is also known as 3C-like protease or 3CL<sup>Pro</sup> from the similarity of its active site and its substrate specificity to those of the picornavirus 3C protease (Anand *et al.*, 2002). M<sup>Pro</sup> is involved in the proteolytic processing of the two overlapping polyproteins pp1a and pp1ab, with the formation of individual mature nonstructural



OPEN ACCESS

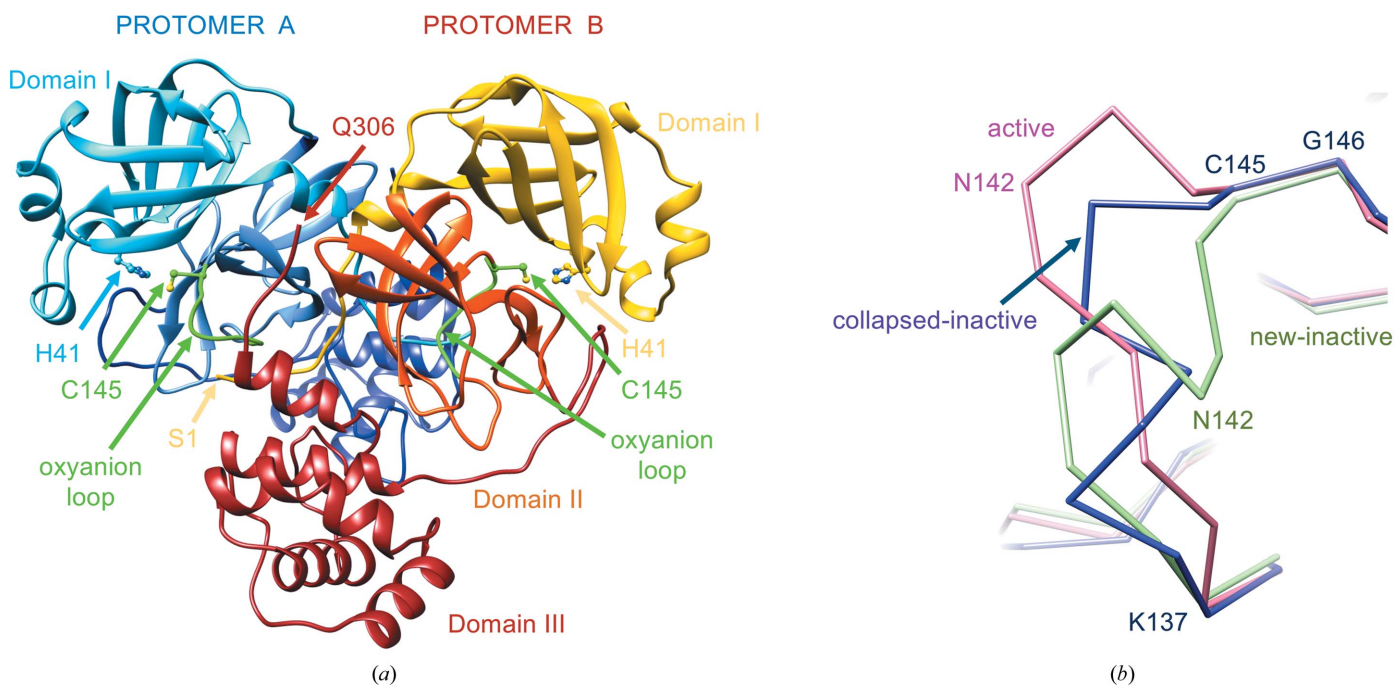
proteins (Snijder *et al.*, 2016), and as such it is a validated antiviral drug target (Dai *et al.*, 2020; Günther *et al.*, 2021; Ullrich & Nitsche, 2020). Currently, there are at least two SARS-CoV-2 M<sup>PRO</sup> inhibitors in phase I clinical trials as candidates with potent antiviral activity: the orally administered PF-07321332 (Pavan *et al.*, 2021) and the intravenously administered PF-00835231 (Ahmad *et al.*, 2021).

SARS-CoV-2 M<sup>PRO</sup> (nsp5), a 306-amino-acid polypeptide of molecular weight 33.8 kDa (Wu *et al.*, 2020), shares 96% sequence identity and a very similar 3D structure with SARS-CoV M<sup>PRO</sup> [0.53 Å r.m.s.d. between PDB entries 6y2e (Zhang *et al.*, 2020) and 2bx4 (Tan *et al.*, 2005)]. Very similar 3D structures have also been found for other coronaviral M<sup>PRO</sup>s such as those from Porcine transmissible gastroenteritis virus (TGEV), which was the first structure of a coronaviral M<sup>PRO</sup> (Anand *et al.*, 2002), Human coronavirus (HCoV) strain 229E (Anand *et al.*, 2003), Infectious bronchitis virus (IBV; Xue *et al.*, 2008) and MERS-CoV (Ho *et al.*, 2015). This structural similarity, which is particularly relevant around the active site, leads to the possibility of the development of pan-coronaviral drugs.

M<sup>PRO</sup> exists in an equilibrium between a monomer and a homodimer (with the two protomers roughly perpendicularly oriented; Fig. 1a), with an apparent  $K_d$  of between 0.8 and 14  $\mu\text{M}$  for the SARS-CoV enzyme, depending on the experimental conditions (Chen *et al.*, 2006). For SARS-CoV-2 M<sup>PRO</sup>, the  $K_d$  has been estimated to be 2.5  $\mu\text{M}$  by analytical ultracentrifugation (Zhang *et al.*, 2020) and 0.14  $\mu\text{M}$  by native mass spectrometry (El-Baba *et al.*, 2020). Unlike 3C protease, only the SARS-CoV M<sup>PRO</sup> dimer shows enzymatic activity (Anand

*et al.*, 2002) and the correct shape of the substrate-binding site, particularly of the S1 subsite; the correct conformation for productive catalytic events is linked to the dimerization process. It has been proposed that the dimerization process has a direct regulatory role of the activity of M<sup>PRO</sup> during the coronaviral replication process (Hsu *et al.*, 2005; Li *et al.*, 2016). Given the high structural similarity, particularly at the dimeric interface, it was reasoned that dimerization of the enzyme is also necessary for the catalytic activity of SARS-CoV-2 M<sup>PRO</sup> (Zhang *et al.*, 2020).

Each M<sup>PRO</sup> protomer is composed of three structural domains (Fig. 1a; Anand *et al.*, 2002). The chymotrypsin-like and 3C protease-like  $\beta$ -barrel domains I (residues 1–99) and II (residues 100–182) directly control the catalytic event. The substrate-binding site is between these two domains and comprises several subsites for substrate binding (from S1 to S6 and from S1' to S3'), corresponding to the P1–P6 and P1'–P3' amino-acid positions of the substrates (according to the convention P6–P5–P4–P3–P2–P1↓P1'–P2'–P3', where ↓ indicates the hydrolyzed peptide bond; Anand *et al.*, 2003). Enzymatic proteolysis by SARS-CoV-2 M<sup>PRO</sup> at the 11 cleavage sites on the viral polyprotein occurs on the C-terminal side of a conserved glutamine in position P1, with the most common consensus sequence being Leu-Gln↓(Ser/Ala), indicating that specificity is determined mostly by the P2, P1 and P1' positions (Ullrich & Nitsche, 2020). Glutamine in position P1 is fully conserved not only for SARS-CoV-2 but also in substrates of SARS-CoV and MERS-CoV. Prime recognition sites at the C-terminus of P1' are not conserved. M<sup>PRO</sup> subsites S4, S2, S1 and S1' have been identified as the most relevant



**Figure 1** SARS-CoV-2 M<sup>PRO</sup> architecture, free form (PDB entry 6y2e). (a) Dimeric assembly of the protease with the main structural features discussed in the text highlighted. Protomer A is in blue-based colors and protomer B is in yellow/red-based colors. The two oxanyonion loops and the two catalytic cysteines 145 are shown in green. (b) Comparison between different oxanyonion-loop conformations of M<sup>PRO</sup>: active in SARS-CoV-2 M<sup>PRO</sup> (PDB entry 6y2e) in pink, collapsed-inactive in SARS-CoV M<sup>PRO</sup> (PDB entry 1uj1 chain B) in magenta and new-inactive in SARS-CoV-2 M<sup>PRO</sup> (this work) in green.

subsites for substrate binding, with regions in the S5, S4 and S2 sites showing considerable conformational flexibility upon binding different chemical groups (Kneller, Galanie *et al.*, 2020). The chymotrypsin-like fold, including domains I and II, is connected by a 16-residue flexible loop to the extra  $\alpha$ -helical domain III (residues 198–306; Fig. 1*a*). Domain III is absent in other RNA virus 3C-like proteases and plays a key role in enzyme dimerization and activity regulation of M<sup>PRO</sup> (Anand *et al.*, 2002; Shi & Song, 2006).

At variance with the classical catalytic triad of chymotrypsin-like proteases, coronaviral M<sup>PRO</sup> has a catalytic dyad, consisting of His41 and Cys145 in SARS-CoV-2 (Fig. 1*a*); a conserved water molecule occupies a position analogous to that of the side chain of the third member of the catalytic triad (for instance, aspartate in chymotrypsin and asparagine in papain) and forms hydrogen bonds to the side chains of His41, His164 and Asp187. It has been proposed that this conserved water is involved in the catalytic event (Anand *et al.*, 2002).

A key role in the proper function of the enzyme is also played by the N-finger (residues 1–7) as the N-terminal tail of one protomer interacts and stabilizes the binding site (S1 subsite) of the other protomer (Verschuere *et al.*, 2008). Indeed, deletion of the N-finger hampers dimerization in solution and abolishes the proteolytic activity. Both the N-finger and the C-terminus are results of the autoproteolytic processing of M<sup>PRO</sup>. Accordingly, in the mature dimeric enzyme both termini of one protomer face the active site of the other.

The important conserved residues Phe140, Leu141, Asn142 and Ser144 (SARS-CoV-2 numbering) are part of a structural element that is essential for a productive catalytic event, the so-called oxyanion loop comprising residues 138–145, which globally lines the binding site for glutamine P1. The central role of the oxyanion loop in the catalytic reaction mechanism of serine proteases and cysteine proteases has been extensively characterized (Frey & Hegeman, 2007). The correct positioning of the oxyanion hole, which is part of the oxyanion loop (formed by the backbone of Gly143, Ser144 and Cys145 in SARS-CoV-2 M<sup>PRO</sup>), is essential for stabilization of the transient tetrahedral acyl (oxyanion) transition state via the hydrogen-bond donor properties of the amides (Anand *et al.*, 2002; Lee *et al.*, 2020; Verschuere *et al.*, 2008). In the known crystal structures of SARS-CoV and SARS-CoV-2 M<sup>PRO</sup>, the oxyanion loop adopts essentially the same ‘active’ conformation; here, we take PDB entry 6y2e as a reference for this conformation (Douangamath *et al.*, 2020; Jin, Du *et al.*, 2020; Jin, Zhao *et al.*, 2020; Zhang *et al.*, 2020). A specific conformation is defined to be active when the amino acids known to participate in the chemical reaction catalyzed by the enzyme are properly positioned and oriented for the reaction to proceed. We also term this conformation catalytically competent.

Variations from the active conformation of the oxyanion loop are found in a few forms of the enzyme, which were consequently considered to be inactive or catalytically incompetent, as in protomer *B* of SARS-CoV M<sup>PRO</sup> (PDB entries 1uj1 and 1uk2; Yang *et al.*, 2003), in the monomeric R298A mutant of SARS-CoV M<sup>PRO</sup> (PDB entry 2qcy; Shi *et al.*,

2008) and in the C172A mutant of 3C<sup>PRO</sup> from the picornavirus hepatitis A virus (Allaire *et al.*, 1994), as well as in IBV 3CL<sup>PRO</sup> (PDB entries 2q6f and 2q6d; Xue *et al.*, 2008). In the inactive monomeric R298A mutant (PDB entry 2qcy), the region of the oxyanion loop, Ser139-Phe140-Leu141, is converted into a short  $3_{10}$ -helix. In PDB entry 1uj1 (SARS-CoV M<sup>PRO</sup> crystallized at pH 6) the oxyanion loop of one of the two protomers exists in a ‘collapsed’ conformation (similar to that found in PDB entry 2qcy), which is considered to be catalytically incompetent, in which the hydrogen bond between Glu166 and His172 that is important for activity is broken (Yang *et al.*, 2003). In the following, we will refer to these two inactive conformations with similar oxyanion-loop conformations as collapsed-inactive (Fig. 1*b*).

In the vast majority of SARS-CoV and SARS-CoV-2 M<sup>PRO</sup> crystal structures, the dimer is crystallographic (Jaskolski *et al.*, 2021); that is, there is only one molecule in the asymmetric unit and therefore the two protomers are perfectly identical. In the very few inactive structures, apart from the artificially induced monomeric forms, the dimer is formed by two different molecules present in the asymmetric unit, one of which is in the inactive state and the other of which is in the active state. Based on molecular-dynamics simulations coupled to activity data in solution, it was suggested that only one protomer at a time is active in the dimer (Chen *et al.*, 2006).

Here, we describe a new inactive structure (called new-inactive) of the main protease of SARS-CoV-2 that is clearly distinct from both the active and the known collapsed-inactive structures, with an oxyanion-loop conformation that is very different from those previously described (Fig. 1*b*). In Section 4, we argue that this conformation has an important functional role as part of the catalytic cycle of coronaviral M<sup>PRO</sup>.

## 2. Materials and methods

### 2.1. Recombinant protein production and purification

The plasmid PGEX-6p-1 encoding SARS-CoV-2 M<sup>PRO</sup> (Zhang *et al.*, 2020) was a generous gift from Professor Rolf Hilgenfeld, University of Lübeck, Lübeck, Germany. Recombinant protein production and purification were adapted from Zhang *et al.* (2020) (where the structure of M<sup>PRO</sup> in the active form was presented; PDB entry 6y2e). The expression plasmid was transformed into *Escherichia coli* strain BL21 (DE3) and then precultured in YT medium at 37°C (100  $\mu\text{g ml}^{-1}$  ampicillin) overnight. The preculture was used to inoculate fresh YT medium supplemented with antibiotic and the cells were grown at 37°C to an OD<sub>600</sub> of 0.6–0.8 before induction with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). After 5 h at 37°C, the cells were harvested by centrifugation (5000g, 4°C, 15 min) and frozen. The pellets were resuspended in buffer *A* (20 mM Tris, 150 mM NaCl pH 7.8) supplemented with lysozyme, DNase I and PMSF for lysis. The lysate was clarified by centrifugation at 12 000g at 4°C for 1 h and loaded onto a HisTrap HP column (GE Healthcare) equilibrated with 98% buffer *A*/2% buffer *B* (20 mM Tris,



150 mM NaCl, 500 mM imidazole pH 7.8). The column was washed with 95% buffer A/5% buffer B, and His-tagged M<sup>Pro</sup> was then eluted with a linear gradient of imidazole from 25 to 500 mM. Pooled fractions containing the target protein were subjected to buffer exchange with buffer A using a HiPrep 26/10 desalting column (GE Healthcare). Next, PreScission protease was added to remove the C-terminal His tag (20 µg of PreScission protease per milligram of target protein) at 12°C overnight. The protein solution was loaded onto a HisTrap HP column connected to a GStap FF column (GE Healthcare) equilibrated in buffer A to remove the GST-tagged PreScission protease, the His tag and the uncleaved protein. M<sup>Pro</sup> was finally purified using a Superdex 75 prep-grade 16/60 SEC column (GE Healthcare) equilibrated with buffer C (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT pH 7.8). Fractions containing the target protein with high purity were pooled, concentrated to 25 mg ml<sup>-1</sup> and flash-frozen in liquid nitrogen for storage in small aliquots at -80°C.

## 2.2. Protein characterization and enzymatic kinetics

The correctness of the M<sup>Pro</sup> DNA sequence was verified by sequencing the expression plasmid. The molecular mass was determined as follows: recombinant SARS-CoV-2 M<sup>Pro</sup>, diluted in 50% acetonitrile with 0.1% formic acid, was analyzed by direct infusion electrospray ionization (ESI) on a Xevo G2-XS QTOF mass spectrometer (Waters). The detected species displayed a mass of 33 796.64 Da, which very closely matches the value of 33 796.81 Da calculated from the theoretical full-length protein sequence (residues 1–306). A representative ESI-MS spectrum is shown in Supplementary Fig. S1. To characterize the enzymatic activity of our recombinant M<sup>Pro</sup>, we adopted a FRET-based assay using the substrate 5-FAM-AVLQ↓SGFRK(DABCYL)K (Proteogenix). The assay was performed by mixing 0.05 µM M<sup>Pro</sup> with various concentrations of substrate (1–128 µM) in a buffer composed of 20 mM Tris, 100 mM NaCl, 1 mM EDTA, 1 mM DTT pH 7.3. Fluorescence intensity (excitation at 485 nm and emission at 535 nm) was monitored at 37°C with a VictorIII microplate reader (Perkin Elmer). A calibration curve was created by measuring multiple concentrations (from 0.001 to 5 µM) of free fluorescein in a final volume of 100 µl reaction buffer. Initial velocities were determined from the linear section of the curve, and the corresponding relative fluorescence units per time unit ( $\Delta\text{RFU s}^{-1}$ ) were converted to the amount of cleaved substrate per time unit ( $\mu\text{M s}^{-1}$ ) by fitting to the calibration curve of free fluorescein. The catalytic efficiency  $k_{\text{cat}}/K_{\text{m}}$  was  $4819 \pm 399 \text{ s}^{-1} \text{ M}^{-1}$ , which is in line with literature data (Ma *et al.*, 2020; Zhang *et al.*, 2020).

## 2.3. Crystallization and data collection

A frozen aliquot of M<sup>Pro</sup> was thawed in ice, diluted in a 1:2 ratio with buffer C (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT pH 7.8) to a final concentration of 12.5 mg ml<sup>-1</sup> and clarified by centrifugation at 16 000g. The inhibitors masitinib, manidipine, bedaquiline and boceprevir were dissolved in 100% DMSO to a concentration of 100 mM. The

protein was crystallized both in the free form and in the presence of inhibitors by co-crystallization. In all cases, final crystal growth was obtained by microseeding starting from small crystals of the free enzyme. The protein in the free form was crystallized using the sitting-drop vapor-diffusion method at 18°C, mixing 1.0 µl M<sup>Pro</sup> solution with 1.0 µl precipitant solution [0.1 M MMT (DL-malic acid, MES and Tris base in a 1:2:2 molar ratio) pH 7.0, 25% PEG 1500] and 0.2 µl seed stock (diluted 1:500, 1:1000 or 1:2000 with precipitant solution) and equilibrating against a 300 µl reservoir of precipitant solution. Crystals appeared overnight and grew for 48 h after the crystallization drops had been prepared. In the case of co-crystallization, M<sup>Pro</sup> was incubated for 16 h at 8°C with a 13-fold molar excess of inhibitor (final DMSO concentration 5%). After incubation with masitinib, manidipine or bedaquiline, a white precipitate appeared and the solutions were clarified by centrifugation at 16 000g; as the protein concentration was essentially unchanged after centrifugation, we concluded that the precipitate is composed of the inhibitors, which are poorly soluble in water. The fact that the protein was later crystallized under the same conditions as described for the free form further confirmed that its concentration was not altered by the centrifugation process. For data collections, crystals were fished from the drops, cryoprotected by a quick dip into 30% PEG 400 (with 5 mM inhibitor in the case of co-crystals) and flash-cooled in liquid nitrogen. The crystals were monoclinic (space group *C2*), isomorphous to the crystals of the free enzyme (PDB entry 6y2e), with one monomer in the asymmetric unit; the dimer is formed by the crystallographic twofold axis.

## 2.4. Structure determination, refinement and analysis

Data were collected on beamlines ID23-2 and ID23-1 at the ESRF. Diffraction data integration and scaling were performed with *XDS* (Kabsch, 2010) and data reduction and analysis were performed with *AIMLESS* (Evans & Murshudov, 2013). Initially, structures were solved by molecular replacement (MR) with *Phaser* (McCoy *et al.*, 2007) in *Phenix* (Liebschner *et al.*, 2019) using PDB entries 6y2e and 5rel (M<sup>Pro</sup> in complex with PCM-0102340; Douangamath *et al.*, 2020) as search models. To limit MR model bias in critical zones (namely residues 139–144, 1–3 and the side chain of His163) we then performed new MR runs using PDB entry 6y2e without residues 139–144 and 1–3, and with an alanine instead of a histidine at position 163, as the search model. Only for co-crystallization experiments with boceprevir was electron density for the ligand clearly visible from the beginning of the refinement (Supplementary Figs. S2 and S3), and the three final structures, modeled from residues 1 to 306 (compared with the ‘new’ structure modeled to residue 301), are virtually identical to those deposited in the PDB (Fu *et al.*, 2020). In all of the other cases, no electron density indicating the presence of the inhibitors masitinib, manidipine or bedaquiline in the active site (or elsewhere) was detectable. For four structures, it was possible to efficiently model residues 139–144, 1–3 and the side chain of His163 in ‘new’ conformations. The final struc-

**Table 1**  
X-ray diffraction data-processing and model-refinement statistics.

Values in parentheses are for the highest resolution shell.

Data collection	
X-ray source	ID23-2, ESRF
Wavelength (Å)	0.873130
Space group	C2
<i>a</i> , <i>b</i> , <i>c</i> (Å)	113.07, 54.71, 44.84
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 101.30, 90.00
Resolution range (Å)	55.44–1.58 (1.61–1.58)
<i>R</i> <sub>merge</sub>	0.070 (1.305)
<i>R</i> <sub>meas</sub>	0.081 (1.505)
<i>R</i> <sub>p.i.m.</sub>	0.040 (0.739)
Total No. of observations	145297 (7276)
No. of unique observations	36653 (1847)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	9.2 (1.0)
CC <sub>1/2</sub> (%)	99.8 (35.7)
Completeness (%)	99.4 (99.5)
Multiplicity	4.0 (3.9)
Wilson <i>B</i> estimate (Å <sup>2</sup> )	23.7
Refinement	
Resolution range (Å)	55.44–1.58
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	17.71/20.31
No. of atoms	
Protein	2350
Water	218
<i>B</i> factors (Å <sup>2</sup> )	
Protein	32.6
Water	43.1
R.m.s.d.	
Bond lengths (Å)	0.008
Bond angles (°)	0.868
Coordinate error (maximum-likelihood-based by <i>Phenix</i> ) (Å)	0.21
Ramachandran statistics	
Favored (%)	97.99
Allowed (%)	2.01
Outliers (%)	0.00
PDB code	7nij
Ensemble refinement	
No. of models	60
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	15.47/20.80

tures were obtained by alternating cycles of manual refinement with *Coot* (Emsley *et al.*, 2010) and automatic refinement with *phenix.refine* (Afonine *et al.*, 2012). At the end, the model was submitted to *phenix.ensemble\_refinement* (Burnley *et al.*, 2012) with default parameters. Data-collection and refinement statistics for the structure obtained by a co-crystallization experiment with masitinib (which was not visible in the final electron density) are reported in Table 1. Secondary-structure analysis was performed with *DSSP* (Kabsch & Sander, 1983; Touw *et al.*, 2015). Local energetic frustration analysis was performed with the *Frustratometer* server (<http://frustratometer.qb.fcen.uba.ar>; Parra *et al.*, 2016). Interface analysis was performed using *PISA* (Krissinel & Henrick, 2007).

## 2.5. Molecular modeling

The majority of the computational work was performed on a Linux desktop workstation (Intel Xeon CPU E5-1620 3.60 GHz) running Ubuntu 16.04 LTS. Molecular-dynamics trajectories were collected on a heterogeneous Nvidia GPU cluster composed of 20 GPUs with models spanning from GTX1080 to RTX2080Ti. For structure preparation, coordi-

nates of the active conformation of SARS-CoV-2 M<sup>Pro</sup> were retrieved from the Protein Data Bank (PDB entry 6y2e). Coordinates for both the active and the new-inactive conformation were processed with the aid of the *Molecular Operating Environment (MOE)* 2019.01 (Chemical Computing Group) structure-preparation tool. Initially, the functional unit of the protease (the dimeric form) was restored by applying a symmetric crystallographic transformation to each asymmetric unit. Residues with alternate conformations were assigned to the highest occupancy alternative. Moreover, missing residues that are present in the primary sequence were added using the *MOE* Loop Modeler tool. The *MOE* Protonate3D tool was used to assign the most probable protonation state to each residue (pH 7.4, *T* = 310 K, i.f. = 0.154). Partial charges were then assigned using the AMBER10 force field and H atoms were energy-minimized until the gradient was below 0.1 kcal mol<sup>-1</sup> Å<sup>-2</sup>. Finally, ions and all co-crystallized molecules except for water were removed before saving the structures. The system setup for the MD simulations was carried out using the *antechamber*, *parmchk* and *tLeap* software implemented in the *AmberTools14* suite (Case *et al.*, 2005). AMBER ff14SB (Maier *et al.*, 2015) was adopted for system parametrization and attribution of partial charges. Protein structures were explicitly solvated in a rectangular prismatic TIP3P (Jorgensen *et al.*, 1983) periodic water box with borders placed at a distance of 15 Å from any protein atom. Na<sup>+</sup> and Cl<sup>-</sup> ions were added to neutralize the system until a salt concentration of 0.154 M was reached. MD simulations were then performed using *ACEMD3* (Harvey *et al.*, 2009), which is based upon an OpenMM 7.4.2 engine (Eastman *et al.*, 2017). Initially, 1000 steps of energy minimization were executed using the conjugate-gradient algorithm. A two-step equilibration procedure was then carried out: the first step consisted of a 1 ns canonical ensemble (NVT) simulation with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic positional constraints applied to each protein atom, while the second step consisted of a 1 ns isothermal–isobaric (NPT) simulation with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic positional constraints applied only to protein C<sup>α</sup> atoms. The production phase consisted of three independent MD replicas for each protein conformation. Each simulation had a duration of 1 μs and was performed using the NVT ensemble at a constant temperature of 310 K with a timestep of 2 fs. For both the equilibration and the production stage, the temperature was maintained constant using a Langevin thermostat. During the second step of the equilibration stage, the pressure was maintained at a fixed value of 1 atm with a Monte Carlo barostat. MD trajectories were aligned using protein C<sup>α</sup> atoms from the first trajectory frame as a reference, wrapped into an image of the system under periodic boundary conditions (PBC), and subsequently saved using a 200 ps interval between each frame and removing any ions and water molecules using *Visual Molecular Dynamics* 1.9.2 (*VMD*; Humphrey *et al.*, 1996). The protein radius of gyration (*R*<sub>g</sub>), the root-mean-square deviation (r.m.s.d.) and the root-mean-square fluctuation (r.m.s.f.) of atomic positions along the trajectory were calculated for protein C<sup>α</sup> atoms exploiting the

MDAnalysis (Gowers *et al.*, 2016; Michaud-Agrawal *et al.*, 2011) Python module. Secondary-structure analysis was carried out with the *STRIDE* package (Frishman & Argos, 1995) as implemented in *VMD* 1.9.2. The collected data were then plotted using the Matplotlib Python library (Hunter, 2007).

Furthermore, two classic MD simulations were performed on the complexes obtained by superposing the coordinates of peptide ligands from PDB entries 2q6g and 7khp on the new-inactive conformation of SARS-CoV-2 M<sup>Pro</sup> using *MOE* 2019.01. For each peptide–ligand complex, a two-stage equilibration protocol followed by a single productive simulation was carried out. The first equilibration step consisted of a 0.1 ns canonical ensemble (NVT) simulation with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic positional constraints applied to each protein atom, while the second equilibration step consisted of a 0.5 ns isothermal–isobaric (NPT) simulation with 5 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic positional constraints applied only to protein C<sup>α</sup> atoms. For both equilibration simulations, the temperature was maintained constant ( $T = 310$  K) using a Langevin thermostat, while during the second equilibration stage the pressure was kept at a constant value of 1 atm using a Monte Carlo barostat. The productive simulation was carried out for 10 ns in the NVT ensemble ( $T = 310$  K).

### 3. Results

#### 3.1. Identification of a new-inactive conformation of M<sup>Pro</sup>

In a campaign to obtain structural insights into SARS-CoV-2 M<sup>Pro</sup>, we analyzed 27 different data sets to determine crystal structures of M<sup>Pro</sup> in complex with different inhibitors, among which were masitinib, manidipine and bedaquiline (Ghahremanpour *et al.*, 2020). As ‘positive’ controls (*i.e.* structures that were already known), we considered ligand-free M<sup>Pro</sup> and M<sup>Pro</sup> in complex with the known  $\alpha$ -ketoamide covalent reversible inhibitor boceprevir, an approved HCV drug that is also able to bind to SARS-CoV-2 M<sup>Pro</sup> (Fu *et al.*, 2020). M<sup>Pro</sup> samples were produced and crystallized in parallel, with very similar experimental procedures, analogous to those of the active enzyme (PDB entry 6y2e; Zhang *et al.*, 2020; see Section 2). Almost all tested crystals were monoclinic (space group  $C2$ , with unit-cell parameters  $a \simeq 113.1$ ,  $b \simeq 54.7$ ,  $c \simeq 44.8$  Å,  $\alpha = 90.0$ ,  $\beta \simeq 101.3$ ,  $\gamma = 90.0^\circ$ ), isomorphous to the crystals of the free active enzyme (PDB entry 6y2e; Zhang *et al.*, 2020) and to most of the deposited M<sup>Pro</sup> structures, signifying the same crystal contacts. After successful molecular replacement and a first round of refinement, in most cases (including the complex with boceprevir) electron density was clearly visible for the entire sequence, indicating a protein matrix with a very similar structure to the search models (PDB entries 6y2e and 5rel; Douangamath *et al.*, 2020). However, there were a significant number of cases, around ten, in which the electron density was of much lower quality or was even absent in particular portions of the protein, namely residues 139–144 of the oxyanion loop, residues 1–3 of the N-finger and the side chain of His163 in the S1 specificity subsite, all of which are

residues that are part of the active site. To cope with the known molecular-replacement bias problem and to correctly rebuild the ambiguous parts, we performed new MR runs using PDB entry 6y2e deprived of residues 139–144 and 1–3, and with an alanine instead of a histidine at position 163 (to remove the His side chain), as a search model. This allowed us to confirm perturbations in the conformation of the selected areas for ten structures, while clear electron density was visible for the remaining cases with the oxyanion loop unambiguously in the active conformation (Supplementary Figs. S2 and S3). In some cases, the electron density was so poor that the tracing of the chain was very problematic, and it was not possible to reliably rebuild the mobile zones entirely (Supplementary Fig. S2*b*). For four structures, it was possible to efficiently model residues 139–144, residues 1–3 and the side chain of His163 in ‘new’ conformations (‘new’ because there are no equivalents in M<sup>Pro</sup> structures deposited in the PDB) that differ from the active conformations and also from the collapsed-inactive conformations, including PDB entry 2qcy, where the oxyanion loop adopts a  $3_{10}$ -helix conformation (Supplementary Fig. S2*c*). In this regard, comprehensive analyses of the available SARS-CoV and SARS-CoV-2 M<sup>Pro</sup> crystal structures have recently appeared in the literature (Behnam, 2021; Brzezinski *et al.*, 2021; Jaskolski *et al.*, 2021; Wlodawer *et al.*, 2020). In no case was a conformation analogous to that presented here described, confirming our assessment of a new-inactive state. The most relevant structures discussed here are reported in Supplementary Table S1.

In summary, we found three different conformational states for the oxyanion loop: active (Supplementary Fig. S2*a*), flexible (*i.e.* with poor electron density; Supplementary Fig. S2*b*) and, strikingly, a new-inactive state (Supplementary Fig. S2*c*). A comparison of the known active and collapsed-inactive conformations with the new-inactive conformation presented here is shown in Fig. 1(*b*).

The new-inactive structures were derived solely from crystals obtained using M<sup>Pro</sup> pre-incubated with the inhibitors masitinib, manidipine or bedaquiline, but in no case was electron density indicating the presence of the inhibitors detected. This is explainable by the medium/high IC<sub>50</sub> (in the range 2.5–19  $\mu$ M; Drayman *et al.*, 2021; Ghahremanpour *et al.*, 2020) and the very low aqueous solubility of the molecules (when inhibitors in 100% DMSO were added to the protein solution, visible white precipitates appeared). It is tempting to speculate that the presence of these inhibitors in solution plays a role in favoring the selection of the new-inactive conformation by the crystallization process. Some structures of crystals from co-crystallization experiments with masitinib or manidipine, again without any evidence for the presence of the ligand in the binding site, show the oxyanion active conformation. This indicates that these molecules, although favoring the new state, are not strict determinants for its formation. In the free form of the enzyme (from crystallization experiments with no ligands), we obtained structures with very clear electron density for the oxyanion loop, as shown in Supplementary Fig. S2(*a*), with low local *B* factors in the refined model, but also structures with a very ‘destabilized’, mobile oxyanion

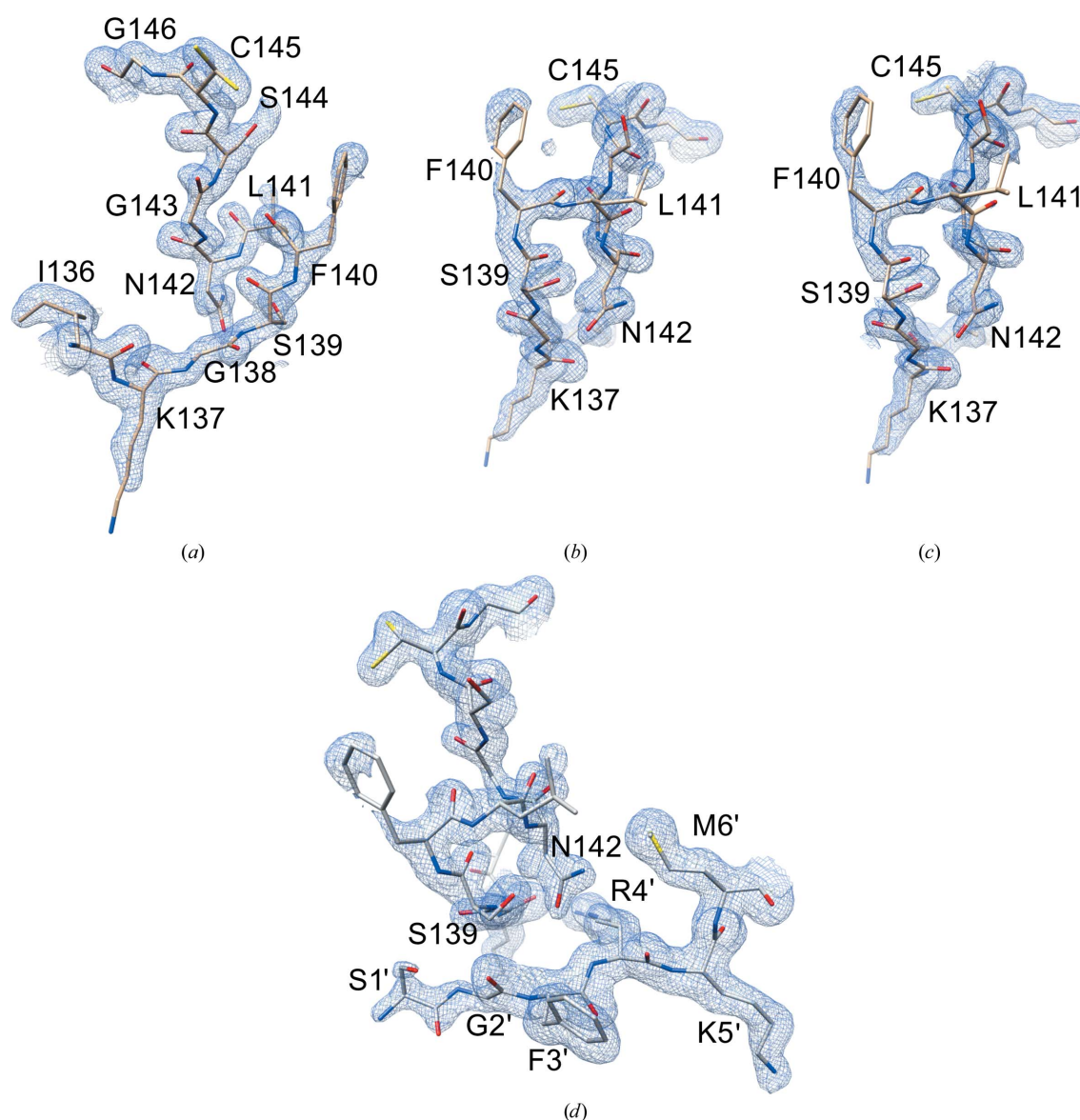
loop, as in Supplementary Fig. S2(b), with much higher  $B$  factors in the final model. This suggests that the high flexibility of the oxyanion loop is an intrinsic property of the free enzyme and is not artificially induced by the presence of ligands in the crystallization experiments.

Here, we describe only one of the structures of  $M^{\text{pro}}$  determined in the new-inactive conformation, which was obtained by co-crystallization experiments with masitinib (no relevant differences exist among the four new-inactive  $M^{\text{pro}}$  structures). Data-collection and final model statistics are reported in Table 1; final electron densities for the most relevant regions discussed in the text are shown in Fig. 2. Unlike in other inactive structures of the enzyme, in which only one protomer adopts the inactive conformation, the

dimeric arrangement of the new structure is due to a crystallographic symmetric axis, and the two subunits are therefore identical and both inactive.

### 3.2. The oxyanion loop adopts a novel inactive conformation

The most striking property of the new structure is the significantly different conformational state of the oxyanion loop (Figs. 1 and 3), which is essential for stabilization of the tetrahedral acyl (oxyanion) transition state during the catalytic cycle. The loop backbone is stabilized by many hydrogen bonds in the new state (Fig. 3a). According to the *DSSP* standardized secondary-structure assignment (Kabsch & Sander, 1983; Touw *et al.*, 2015), in the new oxyanion loop



**Figure 2**

Final electron densities for the most relevant regions of new-inactive  $M^{\text{pro}}$ .  $2F_o - F_c$  maps contoured at the  $1.0\sigma$  level are shown. (a) and (b) show two views of the final electron density for the oxyanion loop in the new conformation. Leu141 and the solvent-exposed Phe140 and Lys137 side chains have incomplete densities indicating various degrees of flexibility. (c) Simulated-annealing omit map (oxyanion-loop residues 138–146 were omitted) viewed as in (b). (d) Electron density in the inter-protomer (intra-dimer) interaction area between the oxyanion loop of one protomer and the N-finger of the other protomer (residues Ser1'–Met6').



there are two consecutive ‘3-turns’ ( $\beta$ -turns) with hydrogen bonds between Leu141 CO and Ser144 NH and between Ser144 CO and Ser147 NH. This region is further stabilized by a ‘4-turn’ ( $\alpha$ -turn) with a hydrogen bond between Ser139 CO and Gly143 NH. *DSSP* does not recognize any  $3_{10}$ -helical segments in the oxyanion loop (as present in the inactive PDB entry 2qcy).

There are other hydrogen bonds involving the backbone that stiffen the oxyanion loop: between Cys145 CO and Asn28 NH, between His163 CO and Gly146 NH and between Ser147 CO and His163 NH (Fig. 3*a*). As a result, the new conformation appears to be quite stable and rigid, as confirmed by the good quality of the local electron density (Fig. 2 and Supplementary Fig. S2*c*).

To analyze the energetics of the local contacts, we performed an energetic frustration analysis (Parra *et al.*, 2016) on the active and new-inactive conformations. The concept of local frustration in protein structure refers to possible residual energetic conflicts in local interactions in folded proteins, using a ‘frustration index’ that measures how favorable a particular contact is relative to the set of all possible contacts in that location (Chen *et al.*, 2020). The ‘principle of minimal frustration’ assumes that proteins find their native state by minimizing the internal energetic conflicts within their polypeptide chain (Bryngelson & Wolynes, 1987). The degree of frustration is therefore dependent on the type of amino acids involved in the interaction. Local violations of this principle have been recognized to be important to exert the proper biological functions, specifically around the active sites of protein enzymes (Freiberger *et al.*, 2019). Analysis of the local configurational frustration of the most interesting contacts around the active site of active and new-inactive  $M^{\text{pro}}$  is shown

in Supplementary Table S2. In both conformations, the catalytic Cys145 is a minimally frustrated ‘hub’ (here we call a position with  $\geq 10$  minimally frustrated interactions a minimally frustrated hub), with a small prevalence of interactions in the active conformation. On the other hand, the difference for Phe140 is striking: eight minimally frustrated interactions are present in active  $M^{\text{pro}}$  (where it is buried in a hydrophobic pocket) as opposed to no interactions in new-inactive  $M^{\text{pro}}$  (where it is solvent-exposed). Differences between the two structures are also evident for other amino acids of the oxyanion loop, namely Leu141, Gly143 and Ser144, indicating their diverse involvement in the local energetic contributions. The oxyanion loop of inactive  $M^{\text{pro}}$  has a larger number of minimally frustrated interactions with Cys117. This residue is a minimally frustrated hub in both conformations; however, given the higher number of minimally frustrated interactions in new-inactive  $M^{\text{pro}}$  (18 versus ten), Cys117 seems to play an important role in the stabilization of the new-inactive conformation. Internal to the oxyanion loop there is also a highly frustrated (unfavorable) interaction involving Leu141, with Ser139 in new-inactive  $M^{\text{pro}}$  and with Ser144 in active  $M^{\text{pro}}$ . This suggests that Leu141 may be important in switching between the two conformations.

### 3.3. Many key interactions of the active enzyme are lost in new-inactive $M^{\text{pro}}$

The correct location of Phe140, Leu141, Asn142, Ser144, Tyr161, His163, Met165, Glu166 and His172 (as seen in the active PDB entry 6y2e, for instance) is an absolute requirement for the reaction catalyzed by  $M^{\text{pro}}$  to properly proceed, with special reference to stabilization of the tetrahedral acyl-

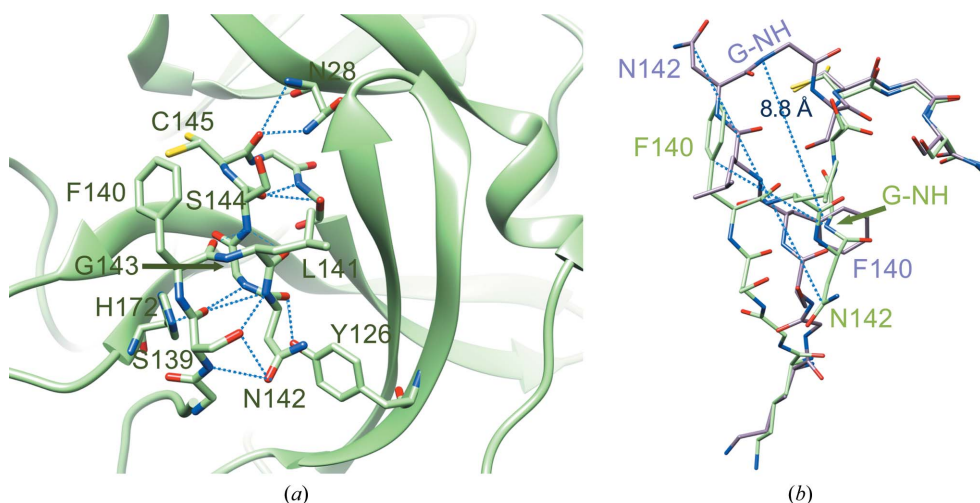


Figure 3

Details of the hydrogen-bond interactions in the oxyanion region of new-inactive  $M^{\text{pro}}$ . (a) The new conformation of the oxyanion loop is stabilized by several backbone hydrogen bonds (blue dashed lines) as described in the main text. The side chain of catalytic Cys145 has a double conformation. (b) Comparison between the new-inactive (green) and active (light magenta; PDB entry 6y2e) oxyanion loops. There are large movements (blue dashed lines) of the side chains of Asn142 and Phe140. In the new-inactive conformation, Asn142 moves from an exposed position with an ASA of  $153.74 \text{ \AA}^2$  to a buried position with an ASA of  $49.00 \text{ \AA}^2$  and Phe140 moves from a buried position with an ASA of  $14.79 \text{ \AA}^2$  to an exposed position with an ASA of  $143.29 \text{ \AA}^2$ . Gly143 NH (G-NH) of the oxyanion hole, which is involved in the stabilization of the tetrahedral intermediate, moves  $8.8 \text{ \AA}$  away.

intermediate (Anand *et al.*, 2002; Lee *et al.*, 2020; Verschueren *et al.*, 2008). Notably, all of these residues are conserved among known coronaviral  $M^{\text{pro}}$ s, underlining their importance. In the new structure of  $M^{\text{pro}}$  most of these residues move away from the ‘active location’: Phe140, Leu141, Asn142 and Ser144 because of displacement of the oxyanion loop (Fig. 3*b*) and His163 and His172 because of rotation of their side chains (Fig. 4).

Specifically, Asn142  $C^\alpha$  and the side chain of Phe140 are remarkably shifted from the active position by  $9.8$  and  $7.5 \text{ \AA}$ , respectively (Fig. 3*b*). Phe140, which is buried in a hydrophobic cleft in active  $M^{\text{pro}}$  with an accessible surface area (ASA) of  $14.79 \text{ \AA}^2$ , is now exposed to the solvent (ASA  $143.29 \text{ \AA}^2$ ), while



Asn142, which is exposed in active  $M^{\text{pro}}$  (ASA 153.74 Å<sup>2</sup>), is now buried (ASA 49.00 Å<sup>2</sup>). The side chain of Asn142 is locked in the new position by hydrogen bonds to the side-chain O<sup>γ</sup> and backbone NH of Ser139. Markedly, the oxyanion hole Gly143 NH, the correct positioning of which is essential for the stabilization of the tetrahedral oxyanion intermediate during catalysis, is moved 8.8 Å away.

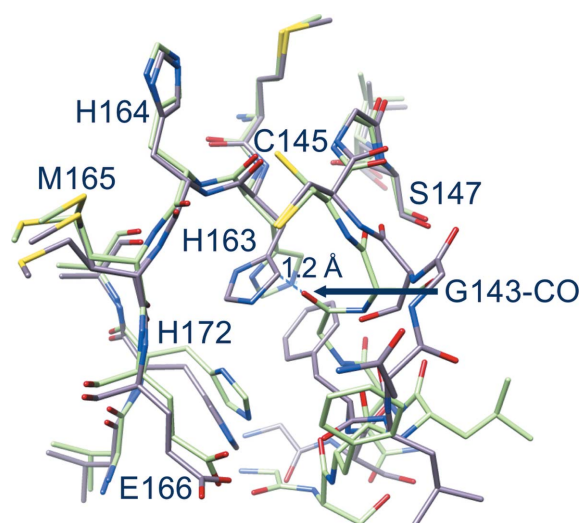
As a consequence, many interactions that are recognized to be important for stabilization of the active conformation are lost, namely hydrogen bonds between Glu166 and His172 and between Tyr161 and His163, as well as the aromatic stacking between His163 and Phe140 (Verma *et al.*, 2020). The rotation of the side chain of His163 (located at the very bottom of the S1 subsite), the hydrogen-bond properties of which seem to be very important in determining both substrate specificity and proper inhibitor binding (Deshmukh *et al.*, 2021), is a noteworthy characteristic of this new conformation of  $M^{\text{pro}}$ . His163 is no longer available for substrate binding as it rotates away to avoid steric clashes with Gly143 CO (Fig. 4). Its position is now 'functionally' occupied by His172, which moves towards the S1 subsite (Fig. 4). The other three important residues, Tyr161, Met165 and Glu166, essentially maintain the same position as adopted in active  $M^{\text{pro}}$ . Despite the large displacement of the oxyanion loop, the position of the catalytic dyad His41 and Cys145 is not significantly altered, especially in the backbone, even though the Cys145 side chain now shows a double conformation (Fig. 5). The conserved water molecule near His41 is still present in the same position, making hydrogen bonds to the side chains of His41, His164 and Asp187 as in active SARS-CoV-2  $M^{\text{pro}}$ .

### 3.4. The N-finger, the C-terminal tail and the dimeric interface are perturbed in new-inactive $M^{\text{pro}}$

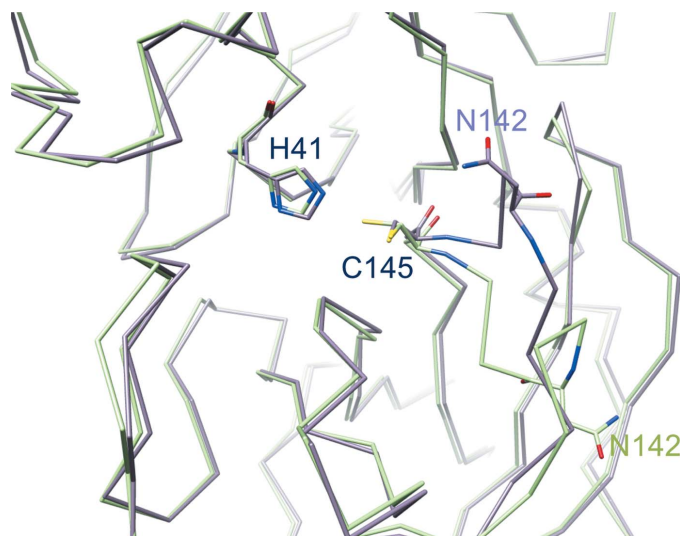
In new-inactive  $M^{\text{pro}}$ , the dimeric interface is altered compared with that of the active conformation. *PISA* analysis of the interface shows that in new-inactive  $M^{\text{pro}}$  the interface area is reduced (from 1661 to 1273 Å<sup>2</sup>), as are the number of hydrogen bonds (from 33 to six) and the number of salt bridges (from 12 to six). However, structural features that are important for stabilization of the dimeric form are essentially conserved, namely (i) the salt bridge between Glu290 of one protomer and Arg4' of the other (Anand *et al.*, 2002), (ii) the hydrophobic aromatic interaction between Tyr126 and Met6' (Wei *et al.*, 2006) and (iii) the interaction of Arg298 with the N-finger and the C-terminus (Shi *et al.*, 2008). This suggests that although new-inactive  $M^{\text{pro}}$  is still able to form dimers, the dimeric state is less stable compared with that of active  $M^{\text{pro}}$ .

At the dimeric interface, relevant changes in both the N- and C-termini are present. In active  $M^{\text{pro}}$ , the N-finger of one protomer interacts and stabilizes the S1 subsite of the other protomer (Verschuere *et al.*, 2008). For instance, in active SARS-CoV-2  $M^{\text{pro}}$  (PSB entry 6y2e) Ser1 of one protomer is hydrogen-bonded both to the carboxylate group of Glu166 and to the main chain of Phe140 of the other protomer. In the new-inactive structure, these interactions are lost as a consequence of the different oxyanion conformation of one

protomer that 'pushes away' residues 1–3 of the N-finger of the other protomer (Fig. 6), with Gly2' CO now at 3.2 Å from Ser139 NH. The rearrangement of the oxyanion loop of one protomer also influences the C-terminal tail of the other protomer, the electron density of which is no longer visible from residue 301 onwards, indicating high flexibility (Figs. 6b and 7). Among the residues of the oxyanion loop, Leu141 shows major changes at the level of the dimeric interface (Fig. 7b), also causing rotation of the side chain of Tyr118 to avoid steric clashes, further supporting its possible central role in switching between the new-inactive and active conformations.



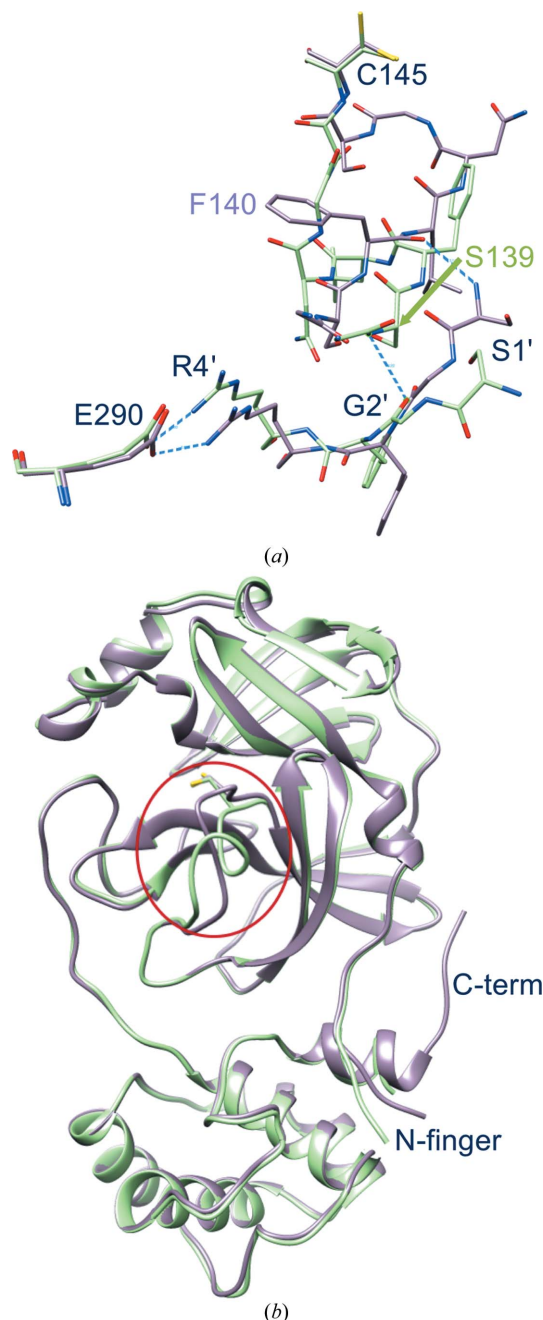
**Figure 4**  
Comparison between new-inactive (green) and active (light magenta)  $M^{\text{pro}}$ . In the new structure the side chain of His163 rotates away to avoid steric clashes with the oxyanion loop: in the active conformation (PDB entry 6y2e) the His163 side chain would be 1.2 Å from the new position of Gly143 CO. Note also the movement of His172.



**Figure 5**  
Catalytic dyad. In new-inactive  $M^{\text{pro}}$  (green) the position of the catalytic dyad His41 and Cys145 is similar to that in the active enzyme (PDB entry 6y2e, light magenta) despite the large shift of residues 138–144. In new-inactive  $M^{\text{pro}}$  Cys145 adopts a double conformation.

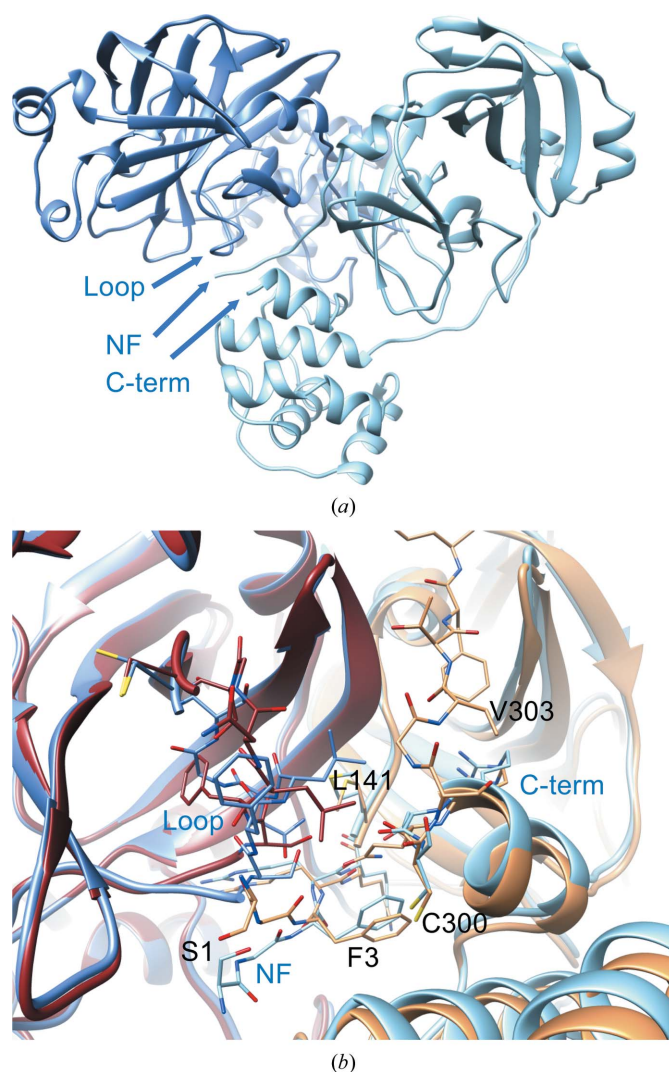
3.5. New-inactive M<sup>PRO</sup> can still bind substrates

Having established that the new structure is catalytically incompetent, we tried to understand whether it is still able to bind natural substrates. Superposition of the new-inactive conformation with either the active conformation in complex with the C-terminal acyl-intermediate (PDB entry 7khp; Lee *et al.*, 2020) or the SARS-CoV M<sup>PRO</sup> active conformation in



**Figure 6** Displacements at the intra-protomer interface. New-inactive M<sup>PRO</sup> is in green and active M<sup>PRO</sup> is in magenta. (a) The new oxanion loop of one protomer pushes away residues 1'–3' of the other protomer; however, the key salt bridge between Arg4' and Glu290, which is important for dimer stabilization, is conserved. (b) Overall superposition of active and new-inactive M<sup>PRO</sup> shows that besides those in the oxanion loop (red ellipsoid), major differences are located in the N-finger and in the C-terminal tail, which is not visible in new-inactive M<sup>PRO</sup>.

complex with its 11-mer substrate complex (PDB entry 2q6g; Xue *et al.*, 2008) does not show evident steric clashes for the substrate. This is also valid for superposition of the new-inactive conformation with two recent complexes between SARS-CoV-2 M<sup>PRO</sup> and two peptide substrates corresponding to the nsp4/5 (Kneller *et al.*, 2021) and nsp8/9 (MacDonald *et al.*, 2021) cleavage sites. Additionally, a short molecular-dynamics refinement of the complexes of the new-inactive conformation of SARS-CoV-2 M<sup>PRO</sup> with either the C-terminal acyl-intermediate or the 11-mer peptide substrate reveal compatible binding modes, with only minor side-chain rearrangements (Fig. 8). The reshaped S1 site of the new-inactive M<sup>PRO</sup> could still host a P1 glutamine, although the rearrangement causes the loss of its interactions with Glu166 O<sup>ε</sup> and Phe140 CO in favor of a single hydrogen bond



**Figure 7** Dimeric architecture of new-inactive M<sup>PRO</sup>. (a) The new conformation of the oxanion loop (labeled 'loop') causes changes in the interface between protomer A (blue) and protomer B (light blue) at the level of the N-finger (labelled 'NF') and the C-terminal tail (labeled 'C-term'). (b) Local differences between the new structure [blue-based colors as in (a)] and the canonical structure (PDB entry 6y2e; brown-based colors, with intact C-terminus): the shift of the Leu141 side chain seems to have major effects in destabilizing the C-terminal tail of the new structure.

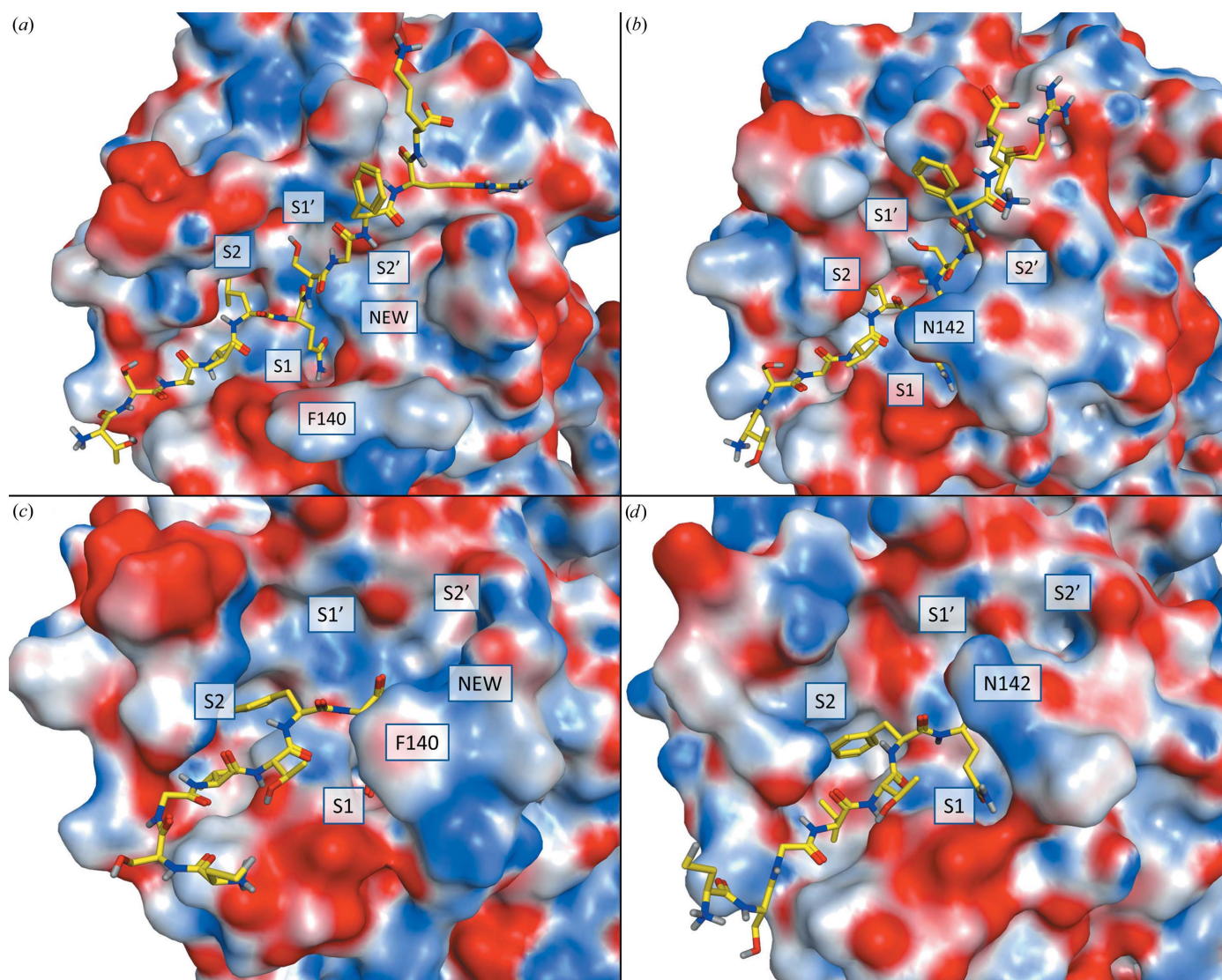


to Gly143 CO (Fig. 9). Aside from the alterations of the S1 subsite, which alter the recognition profile of the P1 glutamine, the other interaction features are retained, namely the hydrogen bonds to Glu166 and Gln189 and the hydrophobic interactions of the P2 phenylalanine within the S2 subpocket. This is a quite remarkable observation because it suggests that the new conformation could be inactive not necessarily because it is incapable of recognizing the substrate, but because the catalytic machinery is not properly organized for an efficient catalytic event, particularly in the oxyanion-hole region, and is unable to stabilize the tetrahedral acyl intermediate. The new conformation of the oxyanion loop generates a new cavity near position S2', as evident from comparison of the new structure with the SARS-CoV-2 acyl-enzyme (PDB entry 7khp; Lee *et al.*, 2020) and the SARS-CoV

11-mer substrate complex (PDB entry 2q6g; Xue *et al.*, 2008) (Fig. 8).

### 3.6. The new-inactive conformation is stable and is in equilibrium with the active conformation in solution

For SARS-CoV M<sup>PRO</sup>, it has been shown that the active-site loops are very dynamic and sensitive to variations in the environmental conditions (Lee *et al.*, 2005; Tan *et al.*, 2005; Xue *et al.*, 2007, 2008; Yang *et al.*, 2003; Zheng *et al.*, 2007). Similarly, the oxyanion loop of SARS-CoV-2 M<sup>PRO</sup> showed conformational flexibility as deduced from room-temperature X-ray crystallography (Kneller, Phillips, Weiss *et al.*, 2020; Kneller, Phillips, O'Neill *et al.*, 2020). To test the stability and to model the dynamics of new-inactive M<sup>PRO</sup>, specifically of the



**Figure 8**  
Reshaping of the S1 and S2' subsites. Molecular-dynamics modeling of the hypothetical interaction of new-inactive M<sup>PRO</sup> with substrates is shown. Top, putative interaction with the 11-mer pseudo-substrate peptide from PDB entry 2q6g: (a) new-inactive M<sup>PRO</sup>, (b) SARS-CoV M<sup>PRO</sup> from PDB entry 2q6g. Bottom, putative interaction with the acyl-intermediate of the M<sup>PRO</sup> C-terminal autoprocessing site: (c) new-inactive M<sup>PRO</sup>, (d) M<sup>PRO</sup> in PDB entry 7khp. As a result of the rearrangement of the oxyanion loop, a new cavity near the S2' site, labeled 'NEW', is formed.



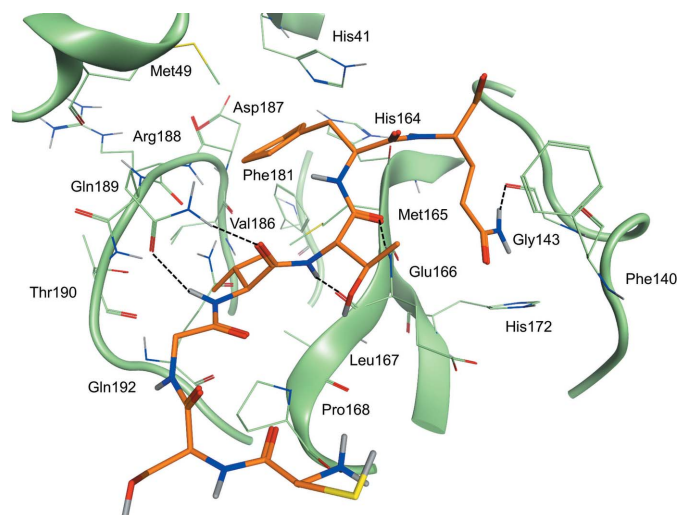
oxyanion loop and regions involved in substrate binding, we performed crystallographic ensemble refinement (Burnley *et al.*, 2012) and MD simulations.

The 60 structures generated by ensemble refinement of new-inactive M<sup>PTO</sup> compatible with the crystallographic restraints confirm the new conformation of the oxyanion loop and reveal that its flexibility is comparable to that of other portions of the substrate-binding region (residues 43–51 in domain I and residues 188–198 in the flexible linker connecting domains II and III; Fig. 10), as also found in the literature. In four out of 60 structures the oxyanion-loop conformation is similar to that in the active form, which is in line with the experimental observation of a residual electron density compatible with the presence of a small fraction of the oxyanion loop and of the side chain of His163 in the active conformation in the crystal state. In this respect, all structures determined here, including new-inactive M<sup>PTO</sup>, were obtained from batches of correctly autoprocessed protein (*i.e.* catalytically active towards itself at the N-terminus) which displayed normal catalytic activity in solution towards substrate peptides.

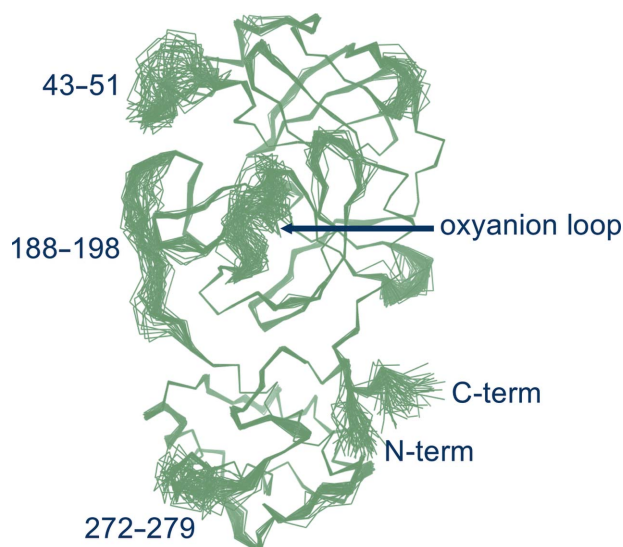
This strongly suggests the presence of a dynamic equilibrium in solution with the coexistence of different conformations, including inactive conformations. In other words, exhibition of the correct catalytic activity on the macroscopic level (with the full ensemble of conformational states available in solution for M<sup>PTO</sup>) does not contrast with the possibility of selection by the crystallization process (in this case probably favored by the presence of certain small molecules) of a subpopulation of a catalytically incompetent form of the enzyme as shown here and for the previous structure with PDB code 1uj1. The conclusion that the dynamic equilibrium in solution includes both the active and the new-inactive conformation is supported by comparing the results of

ensemble refinement of the structure in the free state with very poor electron density for the oxyanion loop (Supplementary Fig. S2*b*). The refined ensemble conformations show a highly dynamic oxyanion loop, with 20% of conformations similar to the active conformation, 23% of conformations similar to the inactive conformation and 57% of conformations in intermediate states.

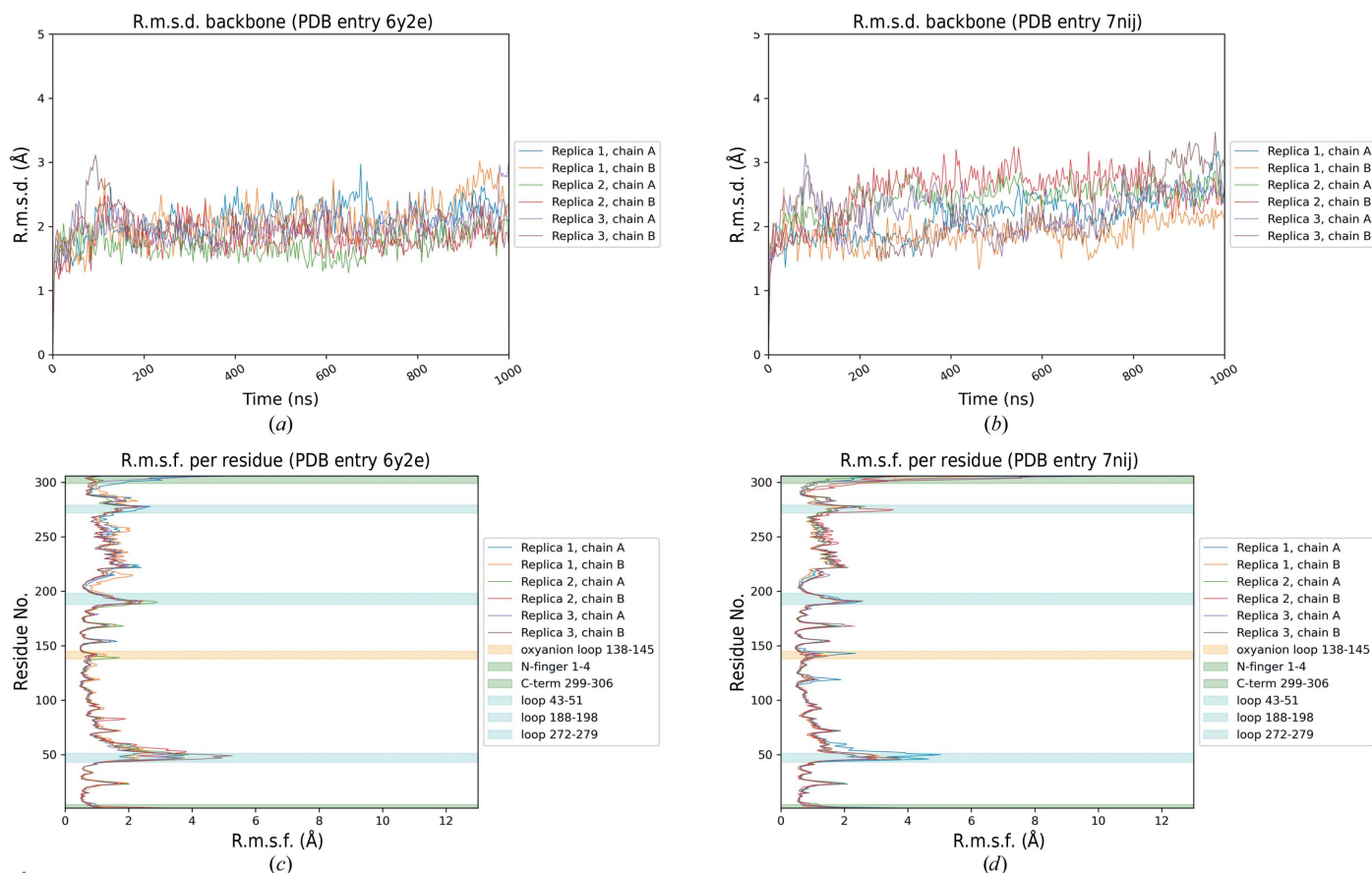
To assess the structural stability of the new-inactive conformation of SARS-CoV-2 M<sup>PTO</sup> and to compare it with the active conformation, three independent 1  $\mu$ s classical molecular-dynamics simulations were performed for both conformations. For the active state, PDB entry 6y2e was taken as a reference. As depicted in Fig. 11, which summarizes the principal geometric analysis performed along the MD trajectories, the two structures show a similar degree of stability. The backbone r.m.s.d. profile for PDB entry 7nij (Fig. 11*b*), representing the new-inactive conformation of M<sup>PTO</sup>, displays moderately higher fluctuations with respect to the active state (Fig. 11*a*). As can be seen in the per-residue r.m.s.f. plots (Figs. 11*c* and 11*d*), this difference can mainly be attributed to major structural fluctuations in the same regions that were marked as flexible by the crystallographic data, namely the three flexible loops 43–51, 188–198 and 272–279 and the C-terminus (299–306), while the rest of the structure is quite stiff, as in the active state. Specifically, the C-terminus in the new-inactive conformation of M<sup>PTO</sup> shows the highest amplitude of movement, as denoted by the high r.m.s.f. values associated with these residues. This result agrees with the absence of electron density for residues 301–306, which indicates high flexibility of this region. Instead, the N-terminus (residues 1–4) shows more limited fluctuations for both M<sup>PTO</sup> conformations, which is in agreement with the presence of well defined electron density in both structures. The overall structural stability of the new-



**Figure 9**  
Details of the putative interaction between new-inactive M<sup>PTO</sup> (green) and the C-terminal acyl-intermediate peptide substrate from PDB entry 7khp (orange). Hydrogen bonds between the substrate and the binding site are depicted as dashed black lines. Aside from the P1 glutamine and its interactions with the P1 pocket, other common interaction features such as hydrogen bonds to Glu166 and Gln189 and hydrophobic interactions of the P2 phenylalanine side chain within the S2 subpocket are retained.



**Figure 10**  
Ensemble refinement. The 60 structures generated by ensemble refinement highlight the mobile regions of new-inactive M<sup>PTO</sup>. The oxyanion loop, which is confirmed in the new conformation, has a flexibility similar to those of residues 43–51 and 188–198 involved in substrate recognition as the S3 and S4 sites.



**Figure 11**

Results of MD simulations. Summary of the key geometric analysis performed along the MD trajectories for both the active (PDB entry 6y2e) and new-inactive (PDB entry 7nij) conformations of SARS-CoV-2 M<sup>Pro</sup>. (a) and (b) highlight the time-dependent variation of the protein root-mean-square deviation (r.m.s.d.) of C $\alpha$  atomic positions for PDB entries 6y2e and 7nij, respectively. (c) and (d) summarize the per-residue mean root-mean-square fluctuation (r.m.s.f.) of atomic positions of protein C $\alpha$  atoms for PDB entries 6y2e and 7nij, respectively. The most relevant regions of the protein are highlighted in the plot for visualization clarity as described in the legend. For both r.m.s.d. and r.m.s.f. analyses, each chain composing the crystallographic dimer is considered separately.

inactive conformation of M<sup>Pro</sup> is also confirmed by the time-dependent evolution of both secondary-structure elements and the protein radius of gyration ( $R_g$ ), with only minor oscillations, similar to those seen in the active conformation (Supplementary Figs. S4, S5 and S6). Despite the slightly higher fluctuations observed in the inactive conformation, no sufficient motions were observed to shed light on a possible transition mechanism between the two conformations. It is not surprising that such rearrangement was not sampled even on a 1  $\mu$ s scale, since such collective motions in proteins usually involve longer timescales (*i.e.* millisecond to microsecond; Orellana, 2019).

#### 4. Discussion

We had the opportunity to capture a new and stable (as seen in MD simulations) inactive state of M<sup>Pro</sup>, called new-inactive, expanding the knowledge of the conformational space accessible to the enzyme. Altogether, the movements in the substrate-binding region and near the catalytic site result in a significant reshaping of the reaction center (Figs. 3, 4 and 8) that has never previously been observed and is much more pronounced than in the previously described collapsed-

inactive M<sup>Pro</sup> conformation. The conformation adopted by residues 139–144 of the oxyanion loop is potentially catalytically incompetent. The backbones of key residues in the oxyanion hole are 8–10 Å away from the catalytically competent position. Fundamental interactions for the proper function of the enzyme are broken or absent, as illustrated in the previous section. Among the residues of the oxyanion loop, Phe140, Leu141 and Asn142 play a major role in the shift between the new-inactive and active conformations. The new state of the oxyanion loop of one protomer pushes the N-finger of the second protomer away from the position adopted in the active enzyme. The last six residues of the C-terminal tail are not visible in the electron-density map and were confirmed to be fully flexible by MD simulations. The novel conformations of the oxyanion loop and of the N- and C-termini result in a weakening of the dimeric architecture, as shown by decreases in the interaction surface area and in the number of inter-protomer interactions. Major variations in the dimeric interface are connected to Leu141 of the oxyanion loop.

This new structure is relevant for the analysis of the M<sup>Pro</sup> catalytic cycle, which was recently investigated using biodynamics theory under non-equilibrium conditions (Selvaggio &

Pearlstein, 2018), using the available crystal structures, which show  $M^{\text{pro}}$  in different conformational states (Wan *et al.*, 2020). This novel approach tries to mimic *in vivo* conditions, which depend on non-equilibrium structure–kinetics relationships. From this analysis a substrate-induced  $M^{\text{pro}}$  activation mechanism was developed, suggesting the existence of a complex substrate-binding activation mechanism in both SARS-CoV and SARS-CoV-2. The proposed catalytic cycle involves transition from the collapsed-inactive conformation of the oxyanion loop, represented by the free form of monomeric  $M^{\text{pro}}$  (PDB entry 2qcy), to the putative substrate-bound form of monomeric  $M^{\text{pro}}$ , represented by one monomer of PDB entry 2q6g (with an active oxyanion loop), and finally to the dimeric fully active state, represented by dimeric  $M^{\text{pro}}$  (PDB entry 6m03; very similar to PDB entry 6y2e). The new-inactive structure presented here shows a new conformational state with an accessible oxyanion loop, adding novel important pieces of information to the structural dynamics of the substrate-induced activation of  $M^{\text{pro}}$  in the context of its catalytic cycle. In the non-equilibrium model, it was hypothesized that transition of the oxyanion loop from the inactive to the active conformation is triggered mainly by solvation/desolvation effects. This also applies to transitions involving our new-inactive structure, where, for activation, Phe140 moves from an exposed position (with no minimally frustrated interactions) to a buried position (with eight minimally frustrated interactions), while Asn142 moves from a buried position to an exposed position. In the context of the conformational dynamics of  $M^{\text{pro}}$ , the intriguing possibility exists that the remodeling of the S2' subsite can be correlated with the large amino-acid variation in position P2' of SARS coronavirus nonstructural protein (nsp) cleavage sites,  $M^{\text{pro}}$  autoprocessing included. Despite being catalytically incompetent, this new state (with a novel cavity in position S2') seems to be able to bind natural substrates of  $M^{\text{pro}}$  (see Figs. 8 and 9). Among the 11 substrates of SARS-CoV-2  $M^{\text{pro}}$ , position P2' is highly variable, hosting nine different amino acids with very different chemical and structural properties: small, such as Gly and Ala, bulky hydrophobic, such as Ile, Val and Leu, positively charged, such as Lys, negatively charged, such as Glu, and polar and hydrogen-bond donor/acceptor, such as Ser and Asn. It is conceivable that the flexibility of the oxyanion-loop conformation is correlated to this variability of the substrates, specifically in position P2', and to the necessity to accommodate the different substrates during the maturation process of the pp1a and pp1ab polypeptides, in the correct succession of proteolytic events. We suggest that this new conformational state is that preferred by the enzyme to efficiently host substrates with bulky hydrophobic residues in position P2', for instance for the processing of nsp7/8 (Ile), nsp12/13 (Val) and nsp14/15 (Leu) cleavage sites. According to the  $M^{\text{pro}}$  reaction scheme proposed by Wan *et al.* (2020), the substrate-binding event triggers the conformational switch of the oxyanion loop, which adopts the necessary conformation for a productive catalytic event. Overall, the following scheme can be proposed: (i) for the initial binding, specific substrates (with bulky residues in position P2') select the new-inactive

conformation among a complex ensemble of different conformations of  $M^{\text{pro}}$  in mutual equilibrium, (ii) the binding event causes conformational changes of the oxyanion loop and, mainly, of the side chains of Glu166, His172 and His163, (iii) the dimeric architecture is stabilized because of rearrangements of the N-finger and the C-terminus and (iv) the resulting activated enzyme is ready to properly hydrolyze the substrate.

The new-inactive structure is also important for the structure-based drug-discovery process that is currently being applied to  $M^{\text{pro}}$  (Deshmukh *et al.*, 2021). The approach of 'repurposing' already known drugs via classical docking methodologies on the 3D structure of the protein target is interesting because, methodologically, it is potentially fast and the safety profiles of the tested compounds are already known. This justifies the large amount of research devoted to repurposing known antiviral drugs against  $M^{\text{pro}}$  (Cannalire *et al.*, 2016). Obviously, the success rate of these campaigns would greatly benefit from the possibility of targeting significantly different, stable, conformations. In this respect, the discovery of the new stable inactive conformation of  $M^{\text{pro}}$  presented here, with the remodeling of the S1 subsite and the formation of the nearby new cavity near subsite S2' (poorly explored until now as known inhibitors usually span the enzyme S1–S4 subsites), offers solid attractive possibilities for the design of completely new classes of antiviral drugs targeting  $M^{\text{pro}}$ . Indeed, a putative binder of the new-inactive form could reduce the population of the active conformation by stabilizing the inactive conformation. Also, a ligand able to bind the novel, readapted site around the catalytic cysteine could sterically hamper the recognition of the substrate. In addition, the possibility of targeting a novel subpocket could increase the affinity by establishing novel contacts and interactions. Most of the more promising  $M^{\text{pro}}$  inhibitors were developed by optimizing starting hits that were further decorated to explore the subpockets located around the catalytic center, following the classic route of fragment maturation in fragment-based lead discovery (Yang & Yang, 2021). One notable example is represented by the optimization of portions of parampanel on S1 and S1' and its engagement of S3–S4, which lead to a fourfold boost in  $IC_{50}$  activity (Zhang *et al.*, 2021).

In conclusion, the new-inactive structure of  $M^{\text{pro}}$  is relevant for better understanding of the function and mechanism of action of this fundamental enzyme for SARS-CoV-2 replication in the cell, with a particular accent on the dynamics within the catalytic cycle of the enzyme, which explores different conformational states including that presented here for the first time. Further, the discovery of this unprecedented inactive conformation of  $M^{\text{pro}}$  provides a unique opportunity for the more successful design of antiviral drugs with improved pharmacological properties using both classical docking-based and innovative non-equilibrium-based approaches.

### Acknowledgements

The authors are grateful to Professor Rolf Hilgenfeld, University of Lübeck, Lübeck, Germany for the free availability



of the plasmid encoding SARS-CoV-2 M<sup>Pro</sup> (Zhang *et al.*, 2020). We thank ESRF beamline ID23-2 local contact Daniele De Sanctis and beamline ID23-1 local contact Romain Talon for help and assistance with data collection. The MMS laboratory is very grateful to Chemical Computing Group, OpenEye and Acellera for the scientific and technical partnership. The MMS laboratory gratefully acknowledges the support of Nvidia Corporation in the donation of the Titan V GPU used for this research. We thank Professor Stefano Mammi for the careful reading of the manuscript and his constructive remarks. Author contributions were as follows. Recombinant protein production and purification were planned and performed by MLM and MB, mass spectrometry by AS and BG, activity measurements by AS, MLM and BG, crystallization experiments and data collection by EF, GG and RB, data processing, structure solution and refinement by RB, MD simulation by MP, MS and SM, overall design of the research by MB, BG, SM, CS and RB, and manuscript writing by RB. The authors declare that they have no conflicts of interest. Open Access Funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement.

### Funding information

This work was supported by funding from the CARIPARO Foundation ('Progetti di ricerca sul Covid-19' No. 55812 to BG and PhD scholarship to EF), from the Department of Chemical Sciences (project P-DiSC #01BIRD2018-UNIPD to MB) and from the University of Padua (Starting Grant STARS@UNIPD – call 2019 to GG).

### References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Ahmad, B., Batool, M., Ain, Q., Kim, M. S. & Choi, S. (2021). *Int. J. Mol. Sci.* **22**, 9124.
- Allaire, M., Chernia, M. M., Malcolm, B. A. & James, M. N. G. (1994). *Nature*, **369**, 72–76.
- Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213–3224.
- Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J. R. & Hilgenfeld, R. (2003). *Science*, **300**, 1763–1767.
- Behnam, M. A. M. (2021). *Biochimie*, **182**, 177–184.
- Bryngelson, J. D. & Wolynes, P. G. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 7524–7528.
- Brzezinski, D., Kowiel, M., Cooper, D. R., Cymborowski, M., Grabowski, M., Wlodawer, A., Dauter, Z., Shabalin, I. G., Gilski, M., Rupp, B., Jaskolski, M. & Minor, W. (2021). *Protein Sci.* **30**, 115–124.
- Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. (2012). *eLife*, **1**, e00311.
- Cannalire, R., Barreca, M. L., Manfroni, G. & Cecchetti, V. (2016). *J. Med. Chem.* **59**, 16–41.
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. & Woods, R. J. (2005). *J. Comput. Chem.* **26**, 1668–1688.
- Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y. & Lai, L. (2006). *J. Biol. Chem.* **281**, 13894–13898.
- Chen, M., Chen, X., Schafer, N. P., Clementi, C., Komives, E. A., Ferreira, D. U. & Wolynes, P. G. (2020). *Nat. Commun.* **11**, 5944.
- Dai, W., Zhang, B., Jiang, X., Su, H., Li, J., Zhao, Y., Xie, X., Jin, Z., Peng, J., Liu, F., Li, C., Li, Y., Bai, F., Wang, H., Cheng, X., Cen, X., Hu, S., Yang, X., Wang, J., Liu, X., Xiao, G., Jiang, H., Rao, Z., Zhang, L., Xu, Y., Yang, H. & Liu, H. (2020). *Science*, **368**, 1331–1335.
- Deshmukh, M. G., Ippolito, J. A., Zhang, C.-H., Stone, E. A., Reilly, R. A., Miller, S. J., Jorgensen, W. L. & Anderson, K. S. (2021). *Structure*, **29**, 823–833.
- Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C. D., Resnick, E., Strain-Damerell, C., Aimon, A., Ábrányi-Balogh, P., Brandão-Neto, J., Carbery, A., Davison, G., Dias, A., Downes, T. D., Dunnett, L., Fairhead, M., Firth, J. D., Jones, S. P., Keeley, A., Keserü, G. M., Klein, H. F., Martin, M. P., Noble, M. E. M., O'Brien, P., Powell, A., Reddi, R. N., Skyner, R., Snee, M., Waring, M. J., Wild, C., London, N., von Delft, F. & Walsh, M. A. (2020). *Nat. Commun.* **11**, 5047.
- Drayman, N., DeMarco, J. K., Jones, K. A., Azizi, S., Froggatt, H. M., Tan, K., Maltseva, N. I., Chen, S., Nicolaescu, V., Dvorkin, S., Furlong, K., Kathayat, R. S., Firpo, M. R., Mastrodomenico, V., Bruce, E. A., Schmidt, M. M., Jedrzejczak, R., Muñoz-Alía, M., Schuster, B., Nair, V., Han, K., O'Brien, A., Tomatsidou, A., Meyer, B., Vignuzzi, M., Missiakas, D., Botten, J. W., Brooke, C. B., Lee, H., Baker, S. C., Mounce, B. C., Heaton, N. S., Severson, W. E., Palmer, K. E., Dickinson, B. C., Joachimiak, A., Randall, G. & Tay, S. (2021). *Science*, **373**, 931–936.
- Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R. & Pande, V. S. (2017). *PLoS Comput. Biol.* **13**, e1005659.
- El-Baba, T. J., Lutomski, C. A., Kantsadi, A. L., Malla, T. R., John, T., Mikhailov, V., Bolla, J. R., Schofield, C. J., Zitzmann, N., Vakonakis, I. & Robinson, C. V. (2020). *Angew. Chem. Int. Ed.* **59**, 23544–23548.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
- Freiberger, M. I., Guzovsky, A. B., Wolynes, P. G., Parra, R. G. & Ferreira, D. U. (2019). *Proc. Natl Acad. Sci. USA*, **116**, 4037–4043.
- Frey, P. A. & Hegeman, A. D. (2007). *Enzymatic Reaction Mechanisms*. Oxford University Press.
- Frishman, D. & Argos, P. (1995). *Proteins*, **23**, 566–579.
- Fu, L., Ye, F., Feng, Y., Yu, F., Wang, Q., Wu, Y., Zhao, C., Sun, H., Huang, B., Niu, P., Song, H., Shi, Y., Li, X., Tan, W., Qi, J. & Gao, G. F. (2020). *Nat. Commun.* **11**, 4417.
- Ghahremanpour, M. M., Tirado-Rives, J., Deshmukh, M., Ippolito, J. A., Zhang, C.-H., Cabeza de Vaca, I., Liosi, M.-E., Anderson, K. S. & Jorgensen, W. L. (2020). *ACS Med. Chem. Lett.* **11**, 2526–2533.
- Gowers, R., Linke, M., Barnoud, J., Reddy, T., Melo, M., Seyler, S., Domański, J., Dotson, D., Buchoux, S., Kenney, I. & Beckstein, O. (2016). *Proceedings of the 15th Python in Science Conference (SCIPY 2016)*, edited by S. Benthall & S. Rostrup, pp. 98–105. [https://conference.scipy.org/proceedings/scipy2016/oliver\\_beckstein.html](https://conference.scipy.org/proceedings/scipy2016/oliver_beckstein.html).
- Günther, S., Reinke, P. Y. A., Fernández-García, Y., Lieske, J., Lane, T. J., Ginn, H. M., Koua, F. H. M., Ehrh, C., Ewert, W., Oberthuer, D., Yefanov, O., Meier, S., Lorenzen, K., Krichel, B., Kopicke, J.-D., Gelisio, L., Brehm, W., Dunkel, I., Seychel, B., Gieseler, H., Norton-Baker, B., Escudero-Pérez, B., Domaracký, M., Saouane, S., Tolstikova, A., White, T. A., Hänle, A., Groessler, M., Fleckenstein, H., Trost, F., Galchenkova, M., Gevorgov, Y., Li, C., Awel, S., Peck, A., Barthelmeß, M., Schlünzen, F., Lourdu Xavier, P., Werner, N., Andaleeb, H., Ullah, N., Falke, S., Srinivasan, V., França, B. A., Schwiner, M., Brognaro, H., Rogers, C., Melo, D., Zaitseva-Doyle, J. J., Knoska, J., Peña-Murillo, G. E., Mashhour, A. R., Hennicke, V., Fischer, P., Hakanpää, J., Meyer, J., Gribbon, P., Ellinger, B., Kuzikov, M., Wolf, M., Beccari, A. R., Bourenkov, G., von Stetten, D., Pompidor, G., Bento, I., Panneerselvam, S., Karpics, I., Schneider, T. R., Garcia-Alai, M. M., Niebling, S., Günther, C.,

- Schmidt, C., Schubert, R., Han, H., Boger, J., Monteiro, D. C. F., Zhang, L., Sun, X., Pletzer-Zelgert, J., Wollenhaupt, J., Feiler, C. G., Weiss, M. S., Schulz, E.-C., Mehrabi, P., Karničar, K., Usenik, A., Loboda, J., Tidow, H., Chari, A., Hilgenfeld, R., Utrecht, C., Cox, R., Zaliani, A., Beck, T., Rarey, M., Günther, S., Turk, D., Hinrichs, W., Chapman, H. N., Pearson, A. R., Betzel, C. & Meents, A. (2021). *Science*, **372**, 642–646.
- Harvey, M. J., Giupponi, G. & Fabritiis, G. D. (2009). *J. Chem. Theory Comput.* **5**, 1632–1639.
- Ho, B.-L., Cheng, S.-C., Shi, L., Wang, T.-Y., Ho, K.-I. & Chou, C.-Y. (2015). *PLoS One*, **10**, e0144865.
- Hsu, M.-F., Kuo, C.-J., Chang, K.-T., Chang, H.-C., Chou, C.-C., Ko, T.-P., Shr, H.-L., Chang, G.-G., Wang, A. H.-J. & Liang, P.-H. (2005). *J. Biol. Chem.* **280**, 31257–31266.
- Humphrey, W., Dalke, A. & Schulten, K. (1996). *J. Mol. Graph.* **14**, 33–38.
- Hunter, J. D. (2007). *Comput. Sci. Eng.* **9**, 90–95.
- Jaskolski, M., Dauter, Z., Shabalin, I. G., Gilski, M., Brzezinski, D., Kowiel, M., Rupp, B. & Wlodawer, A. (2021). *IUCrJ*, **8**, 238–256.
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z. & Yang, H. (2020). *Nature*, **582**, 289–293.
- Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., Zhu, Y., Zhu, C., Hu, T., Du, X., Duan, Y., Yu, J., Yang, X., Yang, X., Yang, K., Liu, X., Guddat, L. W., Xiao, G., Zhang, L., Yang, H. & Rao, Z. (2020). *Nat. Struct. Mol. Biol.* **27**, 529–532.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). *J. Chem. Phys.* **79**, 926–935.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
- Kneller, D. W., Galanie, S., Phillips, G., O'Neill, H. M., Coates, L. & Kovalevsky, A. (2020). *Structure*, **28**, 1313–1320.
- Kneller, D. W., Phillips, G., O'Neill, H. M., Jedrzejczak, R., Stols, L., Langan, P., Joachimiak, A., Coates, L. & Kovalevsky, A. (2020). *Nat. Commun.* **11**, 3202.
- Kneller, D. W., Phillips, G., Weiss, K. L., Pant, S., Zhang, Q., O'Neill, H. M., Coates, L. & Kovalevsky, A. (2020). *J. Biol. Chem.* **295**, 17365–17373.
- Kneller, D. W., Zhang, Q., Coates, L., Louis, J. M. & Kovalevsky, A. (2021). *IUCrJ*, **8**, 973–979.
- Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
- Lee, J., Worrall, L. J., Vuckovic, M., Rosell, F. I., Gentile, F., Ton, A.-T., Caveney, N. A., Ban, F., Cherkasov, A., Paetzel, M. & Strynadka, N. C. J. (2020). *Nat. Commun.* **11**, 5877.
- Lee, T.-W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D. & James, M. N. G. (2005). *J. Mol. Biol.* **353**, 1137–1151.
- Li, C., Teng, X., Qi, Y., Tang, B., Shi, H., Ma, X. & Lai, L. (2016). *Sci. Rep.* **6**, 20918.
- Lieschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Ma, C., Sacco, M. D., Hurst, B., Townsend, J. A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., Marty, M. T., Chen, Y. & Wang, J. (2020). *Cell Res.* **30**, 678–692.
- MacDonald, E. A., Frey, G., Namchuk, M. N., Harrison, S. C., Hinshaw, S. M. & Windsor, I. W. (2021). *ACS Infect. Dis.* **7**, 2591–2595.
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. (2015). *J. Chem. Theory Comput.* **11**, 3696–3713.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. (2011). *J. Comput. Chem.* **32**, 2319–2327.
- Orellana, L. (2019). *Front. Mol. Biosci.* **6**, 117.
- Parra, R. G., Schafer, N. P., Radusky, L. G., Tsai, M.-Y., Guzovsky, A. B., Wolynes, P. G. & Ferreira, D. U. (2016). *Nucleic Acids Res.* **44**, W356–W360.
- Pavan, M., Bolcato, G., Bassani, D., Sturlese, M. & Moro, S. (2021). *J. Enzyme Inhib. Med. Chem.* **36**, 1646–1650.
- Pol, S. & Corouge, M. (2014). *Med. Mal. Infect.* **44**, 449–454.
- Selvaggio, G. & Pearlstein, R. A. (2018). *PLoS One*, **13**, e0202376.
- Shi, J., Sivaraman, J. & Song, J. (2008). *J. Virol.* **82**, 4620–4629.
- Shi, J. & Song, J. (2006). *FEBS J.* **273**, 1035–1045.
- Skwarecki, A. S., Nowak, M. G. & Milewska, M. J. (2021). *ChemMedChem*, **16**, 3106–3135.
- Snijder, E. J., Decroly, E. & Ziebuhr, J. (2016). *Adv. Virus Res.* **96**, 59–126.
- Tan, J., Verschuere, K. H. G., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., Mesters, J. R., Chen, K., Shen, X., Jiang, H. & Hilgenfeld, R. (2005). *J. Mol. Biol.* **354**, 25–40.
- Touw, W. G., Baakman, C., Black, J., te Beek, T. A. H., Krieger, E., Joosten, R. P. & Vriend, G. (2015). *Nucleic Acids Res.* **43**, D364–D368.
- Ullrich, S. & Nitsche, C. (2020). *Bioorg. Med. Chem. Lett.* **30**, 127377.
- Verma, N., Henderson, J. A. & Shen, J. (2020). *J. Am. Chem. Soc.* **142**, 21883–21890.
- Verschuere, K. H. G., Pumpor, K., Anemüller, S., Chen, S., Mesters, J. R. & Hilgenfeld, R. (2008). *Chem. Biol.* **15**, 597–606.
- Wan, H., Aravamuthan, V. & Pearlstein, R. A. (2020). *ACS Pharmacol. Transl. Sci.* **3**, 1111–1143.
- Wei, P., Fan, K., Chen, H., Ma, L., Huang, C., Tan, L., Xi, D., Li, C., Liu, Y., Cao, A. & Lai, L. (2006). *Biochem. Biophys. Res. Commun.* **339**, 865–872.
- Wlodawer, A., Dauter, Z., Shabalin, I. G., Gilski, M., Brzezinski, D., Kowiel, M., Minor, W., Rupp, B. & Jaskolski, M. (2020). *FEBS J.* **287**, 3703–3718.
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C. & Zhang, Y.-Z. (2020). *Nature*, **579**, 265–269.
- Xia, B. & Kang, X. (2011). *Protein Cell*, **2**, 282–290.
- Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., Bartlam, M. & Rao, Z. (2007). *J. Mol. Biol.* **366**, 965–975.
- Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X. C., Liao, M., Bartlam, M. & Rao, Z. (2008). *J. Virol.* **82**, 2515–2527.
- Yang, H. & Yang, J. (2021). *RSC Med. Chem.* **12**, 1026–1036.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & Rao, Z. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 13190–13195.
- Zhang, C.-H., Stone, E. A., Deshmukh, M., Ippolito, J. A., Ghahremanpour, M. M., Tirado-Rives, J., Spasov, K. A., Zhang, S., Takeo, Y., Kudalkar, S. N., Liang, Z., Isaacs, F., Lindenbach, B., Miller, S. J., Anderson, K. S. & Jorgensen, W. L. (2021). *ACS Cent. Sci.* **7**, 467–475.
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K. & Hilgenfeld, R. (2020). *Science*, **368**, 409–412.
- Zheng, K., Ma, G., Zhou, J., Zen, M., Zhao, W., Jiang, Y., Yu, Q. & Feng, J. (2007). *Proteins*, **66**, 467–479.