Autoprocessing and oxyanion loop reorganization upon GC373 and nirmatrelvir binding of monomeric SARS-CoV-2 main protease catalytic domain

Nashaat T. Nashed1, Daniel W. Kneller2, Leighton Coates3, Rodolfo Ghirlando4, Annie Aniana1, Andrey Kovalevsky2✉ & John M. Louis1✉

The monomeric catalytic domain (residues 1–199) of SARS-CoV-2 main protease (MPro1-199) fused to 25 amino acids of its flanking nsp4 region mediates its autoprocessing at the nsp4-MPro1-199 junction. We report the catalytic activity and the dissociation constants of MPro1-199 and its analogs with the covalent inhibitors GC373 and nirmatrelvir (NMV), and the estimated monomer-dimer equilibrium constants of these complexes. Mass spectrometry indicates the presence of the accumulated adduct of NMV bound to MProWT and MPro1-199 and not of GC373. A room temperature crystal structure reveals a native-like fold of the catalytic domain with an unwound oxyanion loop (E state). In contrast, the structure of a covalent complex of the catalytic domain-GC373 or NMV shows an oxyanion loop conformation (E* state) resembling the full-length mature dimer. These results suggest that the E-E* equilibrium modulates autoprocessing of the main protease when converting from a monomeric polyprotein precursor to the mature dimer.
The spatio-temporal regulation of virally encoded proteases and ordered processing of viral polyproteins into functional units is indispensable for the assembly of the replication/transcription complex and production of viable progeny virion1–3. In Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) and its closely related SARS-CoV, a single copy of the main protease (MPro, nsp5) encoded within the polyprotein (pp) 1a and 1ab mediates its self-cleavage (autoprocessing) at its termini (nsp4/nsp5 and nsp5/nsp6) and other cleavages between nsp6 and nsp161. Thus, the mature MPro has been at the forefront for drug development and several lead compounds, including nirmatrelvir (NMV) that has recently received emergency use authorization by the U. S. Food and Drug Administration for the treatment of COVID-19, targeting the active site have been described5,8–12. In this regard, a detailed understanding of MPro release from its precursor polyprotein provides an attractive target for structure and mechanism-based design of active site and dimerization inhibitors prior to its maturation.

The fully active mature MPro is a homodimer, with one active site per monomer, and exhibits a monomer–dimer dissociation constant ($K_{dimer}$) in the low micromolar range5,13. The active site consists of a Cys145-His41 catalytic dyad14. Each monomer contains 306 amino acids that make up three domains, I–III. Domains I (residues 8–101) and II (residues 102–184) exhibit a chymotrypsin-like fold, and domain III (residues 201–306), comprises a cluster of five alpha-helices connected to domain II by a long loop (resides 185–200)3,7,32. Notably, domain III is present only in viruses belonging to the order Nidovirales15, which includes coronaviruses.

Similar to a picornavirus 3C-like cysteine protease16,17 which contains the two domains similar to domains I and II of MPro, MPro exhibits substrate specificity for the sequence (Leu/Ile)-Gln1(Set/Ala/Gly), where J indicates the site of cleavage7,18. The mechanism and intermediates produced in the maturation process of MPro and their catalytic activities are not fully understood. In vitro studies of the precursor are complicated to carry out because a single copy of MPro is anchored on either side by membrane spanning regions within nsp4 and nsp6 flanking MPro (nsp5). It is conceivable, however, that an ensemble of folding intermediates between two MPro chains may mediate transient cleavages at its termini during the early steps of the polyprotein processing cascade. Cleavage at the N-terminus of MPro7,19–21 has been proposed to modulate the $K_{dimer}$ and ensuing catalytic activity through conformational rearrangements by forming inter and intra monomer contacts of the free N-terminal residues with domains II and III. Consequently, deletion of the N-terminal residues 1–7 (termed the N-finger) or domain III leads to a major shift in the monomer-dimer equilibrium mainly to the monomer form accompanied by a drastic decrease in mature-like catalytic activity17,22–24. Various mutational analyses of SARS-CoV MPro and structural requirements for its regulation are summarized in references25,26. Despite the monomer form adopting a native-like tertiary fold, as shown for various mutations or deletions in the sequence, monomeric variants of MPro are reported to exhibit very low or no catalytic activity25,26. This has been attributed to an altered active site that occludes binding of Q-P1 of the substrate in the S1 subsite leading to loss of catalytic function25,27–29. Substrate induced dimerization as a pathway to its maturation and catalytic function has been proposed30,31. The interdependency of dimerization to catalytic activity based on numerous mutational studies is summarized in reference26. In this context, the S5 loop residues Q189-A194 and domain III undergo a significant conformational rearrangement upon N-terminal cleavage and dimerization25,27. All of the above studies pertain to the previous SARS-CoV isolate.

Mutations of critical dimer interface residues such as G11, S139, E290 and R298 have been shown to result in significantly increasing the $K_{dimer}$7,27–29,32. In recent studies we demonstrated the modulation of the monomer-dimer equilibrium of a full-length MPro monomeric construct (MProM), bearing 2 substitution mutations (E290A and R298A) in domain III, by a transition state analog inhibitor GC373 (the reactive aldehyde form of GC376)13. The results provide conclusive evidence that the appearance of mature-like catalytic activity was dependent on dimer formation with two equivalent active sites. In this model system, dimerization and inhibitor binding were inseparable events. It also pointed to a conformational active site equilibrium switching between an inactive state (E) and an active state (E*) synchronized with dimerization, particularly that of the N-terminal dimer interface region and domain III reorientation, the active E* state being dominant in the wild-type MPro (MProWT).

In this study, we examined the physical properties, catalytic activities, and room temperature crystal structures of MPro analogs comprising the catalytic domain and the connecting loop up to residue 199. Therefore, these constructs could serve as early mimetics of a folded intermediate of MPro in the polyprotein form with an exclusion of the entire helical domain III. They are monomeric and display catalytic activities and N-terminal autoprocessing exclusive for a monomeric form as compared to that of the dimeric MProWT. The X-ray structure of the monomeric catalytic domain is similar to the dimer but with an unwound oxyanion loop conformation. In contrast, the binding of GC373 and NMV restores the conformation of the oxyanion loop to that of the active dimeric enzyme. Minimal interface regions enabling dimer formation upon inhibitor binding were identified and related to the stability of the enzyme-inhibitor complex. Importantly, these studies also conclusively show that inhibitors designed to bind to mature MPro dimer also bind to its monomer form.

Results and discussion
Characterization and catalytic activity of monomeric MPro1–199. The genomic organization of SARS-CoV-2 and a ribbon representation of the main protease are shown in Fig. 1. For the full-length monomeric double mutant (E290A, R298A) MProM, the observed dissociation constant of the inhibitor GC373 ($K_d = 1/K_a = K_i = 6.2 \mu M$) and monomer–dimer equilibrium constant ($K_{dimer}$) are identical13. To evaluate the binding of reversible covalent inhibitors to the active site of an exclusively monomeric form of MPro without the associated dimerization as observed for MProM, a construct corresponding to residues 1–199 of MPro (MProWP) spanning the catalytic domain and the loop region was expressed and purified (Fig. S1). The MPro1–199 elutes as a monomer at a peak apex concentration of 95 $\mu M$ and an estimated mass of 22.6 kDa as shown by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) analysis (Fig. 2a). Consistent with this result, sedimentation velocity analytical ultracentrifugation (SV-AUC) of MPro1–199 at a loading concentration of 201 $\mu M$ shows a species at 2.24S with an estimated mass of 21.7 kDa (Fig. 2b, blue trace). The distribution accounts for 99.4% of the sedimenting signal and has an integrated value of 0.963 absorbance units. A similar construct MPro1–196, lacking the C-terminal LE-6His residues, was also expressed and purified to provide a choice of two constructs for crystallization with and without the 6His-tag (Fig. S1). Like MPro1–199, MPro1–196 at a loading concentration of 184 $\mu M$ also sediments as a monomer showing a species at 2.18S and an estimated mass of 20.5 kDa (Fig. 2b, red trace).
Fig. 1 Genome organization of SARS-CoV-2: molecular representation and role of the main protease (MPro). a The ~30 kb genome codes for the various proteins in at least 12 open reading frames (ORFs). Two major polyproteins (pp) are encoded in ORFs 1a (nsp1-nsp10) and 1ab (nsp1-nsp16), the processed proteins of which make up the replication/transcription complex. pp1ab is synthesized via a translation frameshifting (denoted FS) mechanism. The two virally encoded proteases PLPro (papain-like, green) and 3C-like main protease (MPro, blue) are responsible for the processing of pp1a and pp1ab. In the precursor form, MPro is anchored on either side with membrane spanning helices within nsp4 and nsp6. MPro is responsible for its own release (termed self-cleavage or autoprocessing) and cleavage of the rest of the sites between nsp4 and nsp16. b Mature homodimeric MPro and regions critical for the modulation of the monomer-dimer (M-D) equilibrium. Subunits of the dimer are colored in blue and white. Regions defining the boundaries of domains are indicated. The loop region connecting the catalytic domain to the helical domain III (red) with D187, T196 and T199 residues shown as sticks. The free N-terminal strand, indispensable for inter- and intra-monomer interface contacts and dimer stability, is also shown in red just for the blue subunit.

Fig. 2 Molecular mass estimation and catalytic activity of monomeric MPro^1-199 and its miniprecurso. a Molecular mass estimation of MPro^1-199 by SEC-MALS. Fractionation was carried out as described in methods. The observed mass is indicated beside the peak. b SV-AUC absorbance c(s) distributions at loading concentrations of 201 and 184 µM of MPro^1-199 and MPro^1-196, respectively. c Linear relationship between the rate of catalyzed hydrolysis vs. the protein concentration of MPro^1-199. d Lineweaver-Burk plot for hydrolysis of substrate by 90.5 µM MPro^1-199. e N-terminal autoprocessing of the miniprecurso ($-25$)MPro^1-199 upon its expression in E. coli. The precursor, product released upon cleavage at the N-terminus of MPro and molecular weight standards (M) are indicated in kDa. f Molecular mass estimation of ($-25$)MPro^1-199(C145A) by SEC-MALS. The observed mass is indicated beside the peak. Fractionation was performed as described in Methods.
Similar to MProWT and MProM, MPro1–199 catalyzes the hydrolysis of a known Förster resonance energy transfer (FRET) peptide substrate4,13,34 corresponding to the nsp4/nsp5 cleavage site in pp1a polyprotein35. Figure S2 shows chromatograms of the hydrolyzed FRET substrate by MProWT and MPro1–199, respectively. The catalytic activity with the FRET substrate is inhibited by the addition of NMV11,12. The rates of hydrolyses of the FRET substrate at a final substrate concentration of 200 µM displays a linear relationship with the concentration of MPro1–199 having an intercept at the origin [multiple correlation coefficient (R = 0.9984)] and calculated kcat/Km is [(1.0 ± 0.03) × 10−6 µM−1 min−1 (Fig. 2c and Table S1)]. The first-order dependency on the protein concentration indicates that the observed catalytic activity is that of a monomeric MPro1–199, and the protein is exclusively in the monomeric form. A plot of 1/v vs. 1/[S] at MPro1–199 concentration of 90.5 µM is linear (R = 0.9987) with a small positive intercept and the calculated kcat/Km is [(1.3 ± 0.02) × 10−6 µM−1 min−1, Fig. 2d]. Thus, MPro1–199 is 5.22 × 105 times more active than MPro1–199 corresponding to 7.7 kcal/mol, representing the contribution of dimer formation to stabilizing the active form of the molecule (Table S1).

To ascertain if MPro1–199 mediates its own cleavage as a polyprotein and further validate its intrinsic catalytic activity, MPro1–199 was expressed as a miniprecursor [ (−25)MPro1–199] appended to 25 amino acids of its N-terminal flanking sequence which correspond to the C-terminal residues of nsp4 (Figs. 1a and S1). Therefore, the construct encompasses the nsp4/nsp5 junction sequence matching that of the FRET peptide substrate. Indeed, (−25)MPro1–199 (27.8 kDa) promotes its own cleavage at the nsp4/nsp5 junction to produce products (22.9 and 4.9 kDa, Fig. 2e) in E. coli although at a slower rate than observed for MProWT (Fig. S3a) and MProM (Fig. S3b) also expressed as a precursor containing a truncated portion of its N-terminal flanking sequence. In agreement with the catalytic activity profile, MProWT >> MProM >> MPro1–199 (Table S1), a similar progressive decrease in the rate of autoprocessing of these constructs is observed (Fig. S3, note that Fig. 2e is reproduced as Fig. S3c for ease of comparison). Consistent with MPro1–199 being a monomer, SEC-MALS analysis of a control construct +25MPro1–199(C145A) bearing an active site mutation, to restrict catalytic activity and enable the isolation and analysis of the intact precursor, showed a single eluting peak with an estimated mass of 24.9 kDa at a peak apex concentration of 86 µM (Fig. 2f). Like (−25)MPro1–199, a miniprecursor construct (−25)MPro1–196 which matches its processed counterpart MPro1–196 exhibits autoprocessing (Fig. S4a) validating its catalytic activity. However, a deletion to exclude the substrate 5 (S5, residues 189–194) loop sequence as in (−25)MPro1–187 construct (Fig. S4b) results in only a very small fraction of the miniprecursor converting to products even after 3.5 h of induction, relative to (−25)MPro1–199 and (−25)MPro1–196 (compare Fig. S4b with S4a and Fig. S3c) indicating that substrate 5 is critical for catalytic activity.

Inhibitor binding to monomeric MPro1–199, MPro1–196 and MPro10–306. A third construct MPro10–306 was included in this analysis in addition to MPro1–199 and MPro1–196. Construct MPro10–306 (Fig. S1) permits evaluation of the contribution of the N-terminal residues 1–9 influencing inhibitor binding in the presence of a full complement of the helical domain (residues 201–306, Fig. 1b). The binding constants of GC373 and NMV to these three constructs were determined by ITC. The KdS and the thermodynamic parameters are listed in Table 1 and compared to MProWT and MProM.12,13

In an aqueous medium, the prodrug GC376 disproportionate to a sulfite ion and aldehyde GC373. GC373 inhibits MPro by reversibly binding and forming a covalent bond between the sulfur of C145 and the carbonyl carbon of GC373 to yield hemithioacetal6,37. Table 1 shows that Kd for GC373 increases from MProWT to MPro10–306. Raw heat deflections and binding isotherms are shown in Figs. 3 and S5. The Kd of GC373 to the catalytic domain alone and to MPro10–306 is 200 to 300-fold weaker compared to MProWT (Kd = 0.15 ± 0.03 µM). The significant decrease in ΔH for binding, which accompanies weaker binding of GC373 to MPro1–199, MPro1–196 and MPro10–306 is offset by an increase in ΔS to give a net decrease in ΔG of about 3 kcal/mol. Notably, the values of ΔG are nearly the same for MPro1–199, MPro1–196 and MPro10–306, indicative of a common binding mode of GC373 and an active site conformational equilibrium that is similar in these constructs as compared to MProWT and MProM. It is noteworthy that Kd decreases with increasing kcal/Km (Table S1).

Similarly, NMV shows a weaker binding to MPro1–199 and MPro1–196 monomers, but with a 2000 to 2700-fold increase in the Kd relative to MProWT dimer (Table 1 and Figs. 3 and S5). The trend in the decrease in ΔH and increase in ΔS is also observed with a net ΔG difference of 4.5 to 4.7 kcal/mol (Table 1). Like GC373, the binding of NMV to MPro1–199 and MPro1–196 is a two-step process. The first step involves the formation of a noncovalent complex (fast step), followed by a slower step leading to the imidate thioester by forming a covalent bond between the sulfur atom of C145 and the nitrile carbon of NMV. Figure 3 shows the isotherms for binding of GC373 and NMV to MPro1–199. While the isotherm for binding of GC373 to MPro1–199 appears to display a single binding step with sharp heat deflections (Fig. 3a, b), the isotherm for NMV binding shows the two-step process with broader heat deflections (Fig. 3c, d), the latter being a dominant portion of the isotherm. The Kd’s derived and listed in Table 1 correspond to the latter process and not the initial binding event. A duplicate titration with 30 injections shows a similar binding isotherm (Fig. S6a) and thermodynamic parameters.

The binding of NMV to MPro1–199 requires the presence of C145 as no thermal response was observed with MPro1–199(C145A) (Fig. S6b). In contrast, the binding of NMV to MProWT is indicative of a fast reaction to the dimeric protein’s active site, with sharp heat deflections, which in part contributes to the higher affinity of NMV as shown before (Fig. S7a)11,12. While no thermal response was observed for the titration of NMV with MPro10–306 under our experimental conditions (Fig. S5d), reversed-phase liquid chromatography with in-line mass spectrometry (RPLC-MS) results indicated very little binding of NMV to the protein (see below). It is noteworthy that unlike NMV, GC373 binding to MPro1–199, MPro1–196 and MPro10–306 exhibit the typical binding with sharp deflections (Figs. 3a and S5a, c) like that observed when titrating with MProWT13.

The results presented above indicate that NMV covalent complex formation requires the catalytically active conformation of the protein. The differences in binding of NMV to MPro1–199, MPro1–196 and MPro10–306 and those of GC373 may be attributed to the difference in the electrophilicity of the carbonyl carbon of GC373 and the nitrile carbon of NMV. Generally, the carbonyl carbon is more electrophilic and susceptible to nucelophilic attack than the nitrile carbon38, as oxygen is more electronegative than nitrogen. Also, the carbonyl group of the aldehyde GC373 is planar, mimicking the carbonyl of a peptide substrate. Thus, the carbonyl oxygen would be in a position to interact with the oxyanion hole of the protein and thereby enhancing the carbonyl carbon electrophilicity and the formation of the hemithioacetal. This interpretation is consistent with the binding of GC373 to MPro10–306 with a similar Kd and thermodynamic parameters as that with MPro1–199 and MPro1–196, although only half the protein is competent to bind.
to the inhibitor ($N = 0.54$, Table 1). In contrast, the nitrile group is linear, and the nitrile nitrogen may not readily interact with the oxyanion hole hydrogen bonds until the formation of the imidate thioester. This is evident from the absence of a thermal response in ITC for MPro10–306 with NMV (Fig. S5d) indicative of very little imidate thioester formation (see below).

**Table 1** Binding affinity of GC373 and NMV to MProWT, MPro1–199, MPro1–196 and MPro10–306 as determined by ITC.

| Compound | Chemical structure | Construct | $N$ | $K_d = K_i$ (µM) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G$ (kcal/mol) |
|----------|--------------------|-----------|-----|-----------------|----------------------|----------------------|----------------------|
| GC373    | ![Chemical structure](image) | MProWT    | 0.99 ± 0.01 | 0.15 ± 0.03 | −6.7 ± 0.1 | 9.1 | −9.4 |
|          |                    | MProM     | 1.07 ± 0.02 | 6.13 ± 0.30 | −6.0 ± 0.2 | 3.9 | −7.2 |
|          |                    | MPro1–199 | 0.9 ± 0.03  | 32 ± 5     | −2.4 ± 0.1 | 12.7 | −6.2 |
|          |                    | MPro1–196 | 0.88 ± 0.09 | 45 ± 20    | −1.48 ± 0.24 | 15 | −6.6 |
|          |                    | MPro10–306| 0.54 ± 0.03 | 44 ± 8     | −1.38 ± 0.1 | 15.3 | −6.6 |
| NMV      | ![Chemical structure](image) | MProWT    | 0.99 ± 0.003| 0.007 ± 0.003| −10.75 ± 0.7 | 15.7 | −11.2 |
|          |                    | MProC45A  | 0.96 ± 0.04 | 2.7 ± 0.9   | −3.89 ± 0.2 | 12.6 | −7.7 |
|          |                    | MPro1–199 | 0.97 ± 0.04 | 19 ± 3     | −3.9 ± 0.1 | 8.8 | −6.5 |
|          |                    | MPro1–196 | 1.05 ± 0.03 | 14 ± 3     | −1.8 ± 0.1 | 16.3 | −6.7 |

ITC experiments were carried out in buffer C at 28 °C. Data were processed and plots were generated with the Origin software provided with the instrument. Titrations of MPro1–199 with NMV shown in Figs. 3c and S6a were fit to a 2 sites model. The mean values obtained for the major isotherm are listed. Thermodynamic parameters derived for MProWT and MProM titrated with GC373, and MProWT titrated with NMV are cited from refs. 12, 13, respectively, solely for comparison with MPro analogs. No thermal response was observed when NMV was titrated with MPro1–199 (Fig. S6b) and MPro10–306 (Fig. S5d), and GC373 with MProC45A as listed in Table S1.

**Fig. 3** Binding isotherms of GC373 and NMV to MPro1–199. Titrations were carried out with MPro1–199 (in the cell) vs. a GC373 and c NMV (in the syringe) in buffer C at 28 °C. Enlarged view of a few deflections are shown for comparison when titrated with GC373 (b) and NMV (d). A slow thermal response is observed for the interaction of NMV with MPro1–199. This slow response is not observed when titrating NMV with MProWT (ref. 12, Fig. S7a) or MProC45A (this work, Fig. S7b). Thermodynamic parameters are listed in Table 1.

Oxyanion loop unwinding in monomeric MPro1–199. To understand the structural implications of our solution measurements of the enzymatic activity and inhibition of MPro1–199, we obtained its room-temperature structure in the inhibitor-free form at 2.25 Å resolution. MPro1–199 crystallizes with two independent molecules present in the asymmetric unit of P2₁2₁2₁.
空间群中缺乏一个晶体学2-轴（图4a）。成熟MProWT主要在具有晶体学2-轴的空间群中结晶，以便在成熟酶中通过这种对称运算形成同源二聚体3,39,40。这表明MPro1 – 199是一个单体在晶体中。为了进一步提供证据，MPro1 – 199在两种独立分子中形成单体，并不形成原本的同源二聚体。如图4b所示，两个独立分子的MPro1 – 199没有形成同源二聚体。在不同的位置和氧基环残基139-142形成一个短的螺旋。e 共同位置的MProWT与MPro1 – 199显示结构重排的结果，从原结构到截短结构的橙色箭头。氢键对的键合线为点状虚线。所有超共线位使用基于α残基在MPro1 – 199中模型的最小平方拟合。
from its position in MProWT, making interactions with the helical domain of the native dimer is not formed by the juxtaposition of the two monomers in the S4 β-hairpin loop (residues 165–199) in the loop preceding the 310 helix, residues 117–126 following the S2 helix.

The orientation of substrate binding subsite S1 is preserved in MPro1 capable of binding to the monomeric MProWT structure with its N-terminus. In the absence of the stabilizing effect of the N-terminus, the N-terminus of the second protomer that usually caps S1 as part of the N-terminus of the second protomer that usually caps S1 as part of the mature homodimer is no longer present in the truncated enzyme. In the absence of the stabilizing effect of the N-terminus, the side chain of Glu166 rotates from its position observed in MProWT, where it hydrogen bonds the N-terminus of the second protomer toward the oxyanion loop (Fig. S8). In the new structure, hydrogen bonds with the main chain amide NH of Gly143 that was part of the oxyanion hole in MProWT. Of note, the S2 helix (residues 46–51) moves ~2 Å away from the catalytic site (Cys145/His41 dyad) and subsite S1. This movement results in the active site opening by ~1 Å in MPro1 relative to MProWT measured by the increase in the distance between Ca atoms of Ser46 in the S2 helix and Pro168 in the S4 β-hairpin loop (residues 165–170), a characteristic feature usually associated with binding of a ligand to the enzyme.

Conversely, the carbonyl of the carbamate group makes identical 2.8 Å hydrogen bonds with the main chain amide NH of Glu166 in both structures, whereas carbamate’s NH recruits the side chain of Glu189 in MPro1, GC373 forming a direct 2.9 Å hydrogen bond. In MProWT, the carbonyl is instead rotated away from the inhibitor facing the bulk solvent, and a water molecule is inserted between the carbamate’s NH and Glu189 side chain, replacing a direct contact with a water-mediated interaction (Fig. 5c). Similar to GC373, NMV interacts with the carbonyl of the His163 imidazole in GC373 and the carbonyl of the carbamate group makes identical 2.8 Å hydrogen bonds with the main chain amide NH of Glu166 in both structures, whereas carbamate’s NH recruits the side chain of Glu189 in MPro1-NMV complex (Fig. 5d).

We also superimposed the inhibitor complexes MPro1, GC373 and MPro1-NMV onto the inhibitor-free structure of MPro1 to visualize the differences in the active site organization and conformation. As is evident from Fig. 5e, when MPro1 is represented with a surface, the inhibitor groups P1 and P2 clash with the residues in subsites S1 and S2, providing further evidence that the active site shape in the inhibitor-free structure does not match the inhibitor structures. Thus, GC373 and NMV would either bind to the monomeric enzyme by induced fit, reshaping the active site into the active conformation, or bind through the conformational selection, when some proportion of MPro1 molecules contains active sites with the appropriate conformation for the inhibitor binding.

Our structures indicate that inhibitor bound conformation of the oxyanion loop of MPro1 is nearly identical to that of MProWT. Thus, it is the transitioning of the oxyanion loop from the unwound to the native state that likely accounts for the weaker binding and measured differences in the thermodynamic properties of the inhibitor complexes.
parameters observed for MPro1-199 and MPro1-196, relative to MProWT. Presumably, the \( \Delta H \) is related to the magnitude of GC373 and NMV reactivity, and \( \Delta S \) to the associated conformational changes including those of the terminal residues and the exclusion of the hydration water from the active site.

Inhibitor binding promotes dimerization of MPro1-199 and MPro1-196. As stated above, the binding of GC373 to a predominantly monomeric MPro [MProWT] is accompanied by an increase in dimer formation and catalytic activity. To ascertain the influence of binding of GC373 and NMV on dimer formation in the absence of the helical domain or the N-terminal residues 1-9, a series of SV-AUC analyses were carried out at concentrations of ~50 and ~200 µM of MPro1-199, MPro1-196 and MPro10-306 and two-fold molar excess (hereafter also referred to as 2x) of inhibitors GC373 and NMV, the higher concentrations reflect nearly the same protein/inhibitor ratios upon completion of the ITC experiments (~200 µM).

At a loading concentration of ~50 µM MPro1-199 and in the absence of GC373, a single species at 2.28S accounts for the total signal with an estimated mass of 21.6 kDa (Fig. 6a). This single peak slightly shifts to 2.31S and becomes slightly broader in the presence of 2x GC373 indicative of a small second population in fast equilibrium with the major monomeric species. A clear separation of two peaks corresponding to 2.34 (21 kDa) and 3.1 S (32 kDa) is observed with ~200 µM MPro1-199 with 2x GC373, the faster sedimenting peak accounting for about 14% of the signal (Fig. 6b). The same trend is observed in the presence of NMV mixed with MPro1-199, at ~50 µM the 2.38S peak gets even broader, and at ~200 µM, the 2.31S (19.7 kDa) and 3.14S (31.2 kDa) species account for 32% and 68% of the integrated signals (Fig. 6a, b). Unlike MPro1-199, no dimer was observed at ~200 µM MPro1-196 concentration with 2x GC373, presumably due to the shortened loop region by three amino acids. The observed S values were 2.23 and 2.26 at ~50 and ~200 µM MPro1-196, respectively, in the presence of GC373 and no significant broadening of the peak is observed (Fig. 6c, d). In the presence of
of NMV and ~50 µM MPro1-196, a single peak corresponding to a monomer of 2.24S is observed with no significant broadening of the peak as seen with ~50 µM MPro1-199 mixed with NMV (Fig. 6c). At ~200 µM MPro1-196 mixed with 2x NMV, two species with equal distribution (50:50) corresponding to 2.25 and 2.86S were evident (Fig. 6d). It is of significance that at ~200 µM MPro10-306 mixed with 2x the concentration of either inhibitor, only sharp peaks of the monomer and no dimer peak was observed (Fig. 6e). The observed S values were 2.65 with an estimated mass of 33.8 kDa in the presence of 2x GC373 and 2.63 with an estimated mass of 34.6 with 2x NMV.

Interestingly, our ITC data indicates that GC373 is a better binder to monomeric MPro10-306, relative to NMV. It could be that the oxyanion loop equilibrium does not shift as readily to the inhibitor bound state with NMV in the absence of residues 5-9, this being yet another equilibrium process (or conformational selection) coupled to the oxyanion loop equilibrium.

Apparent dimer dissociation constants ($K_{d}_{dimer\_app}$) for ~190 µM MPro1-199 and ~186 µM MPro1-196 in the presence of NMV were determined by single concentration Lamm equation modeling of the absorbance and interference sedimentation velocity data as described in the Experimental section. The best-fit dimer dissociation constants are listed in Fig. 6f (see Fig. S9). In addition, a side-by-side comparison of the estimated $K_d$ and $K_{dimer\_app}$ values by ITC and SV-AUC, respectively, and the fold difference $K_{dimer\_app}/K_d$ are listed in Table S1. It is evident from this data that $K_{dimer\_app}$ for MPro1-199 and its analogs in the presence of inhibitor is significantly higher than the corresponding $K_d$, contrary to that observed for MProM previously, pointing to a distinct separation of inhibitor binding from dimer formation. As pointed out earlier, GC373 binding to the predominantly monomeric MProM leads to dimer formation concomitant with restoring the oxyanion loop conformation to the active state. Contrastingly, the results presented in this study indicate that inhibitor binding to monomeric MPro1-196 leads to establishing the active conformation of the oxyanion hole and the thermodynamic stability of the resulting complex promotes dimer formation.

Identifying an adduct of NMV and not that of GC373 in solution. The $K_d$ for the binding of NMV to MProC145A is 2.7 µM as determined by ITC (Table 1 and Fig. S7). Thus, the nitrile warhead reacting with C145 increases the binding affinity by ~400 times, with a measured $K_d$ of 7 nM for NMV with MProWT. Since imidate thioesters are known compounds and
significantly more stable than hemiacetals and hemithioacetals, proteins mixed with either GC373 or NMV and equilibrated for 2 h were analyzed by RPLC-MS. Typically, incubated mixtures of protein (50 and 200 µM) and 2× inhibitor, similar to those used in Fig. 6 for SV-AUC, were diluted to a concentration of 20 µM in 5% aqueous acetic acid and 10 µl of the sample was subjected to RPLC-MS. Fractionation by RPLC results in a dilution of the sample based on their retention volumes by 500 to 600-fold for MPro1–199 and MPro1–196 and ~700 fold for MPro WT and MPro10–306. The mass spectra show a nearly equal distribution of MProWT and the MProWT-NMV adduct (Fig. 7a). In 50 µM mixtures, representing a concentration ~3-fold above the observed dissociation constants (Table 1), protein-NMV adducts were observed for MPro1–199, MPro1–196 and MPro10–306 (Figs. 7c and S10a, c). In a control experiment, only the molecular ion for MPro1–199(C145A) was observed and not of the protein-adduct (Fig. 7b). This is consistent with the very weak or lack of binding observed with ITC (Fig. S6b) and no faster sedimenting species other than a monomer was observed (Fig. S6c). At higher concentrations (~200 µM) of protein-NMV mixtures, MPro1–199, MPro1–196 and MPro10–306 show an increased population of the adduct (Figs. 7d and S10b, d). The protein-adducts are stable for days when maintained at 20 µM in acetic acid at ambient temperature as shown in Figs. 7d and S10b. Even though ITC and SV-AUC data suggest either very weak or lack of binding of NMV and its associated dimerization with MPro10–306, as seen for MPro1–199 and MPro1–196, mass spectra indicate that NMV binds to the active site of MPro10–306 as it does with MPro WT. This indicates the presence of a small population of MPro10–306 having the appropriate conformation to allow NMV binding that leads to the formation of the imidate thioester adduct. In contrast to the observed adducts with NMV, mass spectra of mixtures containing GC373 display only the molecular ions of the unmodified protein as verified both with MPro1–199 and MPro1–196 incubated at high concentrations and subjected to RPLC-MS.
The adduct is an imidate thioester (Fig. 7e) observed with MProWT and MPro1–196 as shown in crystal structures reported previously3,12. Also, GC373 is known to form a covalent hemithioacetal adduct to the catalytic C145 of MProWT which we have not observed in RPLC-MS analysis. There are two possible chemical pathways for the decomposition of the adduct in aqueous solution. As suggested by Owen et al.11, the formation of the imidate thioester is reversible, but the reverse reaction is slow, leading to its accumulation in solution. Alternatively, imidate thioesters are reactive esters susceptible to hydrolytic reactions leading to the formation of amides and/or thioesters41 which may be subject to further hydrolysis to the corresponding carboxylic acids. Proteases such as α-chymotrypsin, subtilisin and carboxy-peptidase. A have been reported to catalyze the hydrolyses of esters and amide substrates in which the ester or amide carbonyl oxygen is replaced with a sulfur, nevertheless at much slower rates than the corresponding amides and esters42,43 and thus, it is possible for MProWT to catalyze the hydrolysis of the imidate thioester adduct (Fig. 7e) to produce one or more of the three possible products; thioester, amide, and carboxylic acid (Fig. 7f). It should be noted that thioesters and amides are substrates that can undergo catalytic hydrolysis by the enzyme to produce the carboxylic acid as an end product. Our results and the conclusion presented above are consistent with the results obtained for the inhibition of cysteine proteases such as papain by nitrile analogs of their substrate which establish the formation of imidate thioester adducts via an enzymatic process44–46. Imidate thioester of papain was reported to undergo hydrolysis to carboxylic acid and free enzyme in a very slow process45.

Conformational stabilization of the oxyanion loop upon inhibitor binding promotes dimer formation. The results listed in Tables 1 and S1 indicate that inhibitors GC373 and NMV bind to monomeric MPro1–199 (and MPro1–196) with $K_d$'s in the range of 32–45 µM and 14–19 µM, respectively. In contrast, the apparent $K_{d_{\text{dimer}}}$ for GC373 and NMV are more than 175 µM. The observed catalytic activity of MPro1–199 indicates the presence of a very small fraction being in the catalytically competent conformation (E*) resembling the active site of the MProWT dimer. The scheme in Fig. 8 is a proposed mechanism to account for the observed results with NMV. Both MPro1–199 and MPro1–196 are mostly in the inactive conformation (E). Binding of GC373 or NMV to MPro1–199 and MPro1–196 leads to a rearrangement of the oxyanion loop equilibrium (E–E*) to that of the active form (E*) of the enzyme confirmed by our X-ray structures. Such an E → E* rearrangement and accompanied conformational changes facilitate dimer formation. The $K_{d_{\text{dimer}}}^{\text{app}}$ for dimer formation of MPro1–199/GC373 complex is 38-fold greater than the dissociation constant ($K_d = 32 \pm 5$ µM) of GC373 to the monomeric MPro1–199. Similarly, $K_{d_{\text{dimer}}}^{\text{app}}$ for dimer formation of MPro1–199/NMV complex is 9-fold greater than the dissociation constant ($K_d = 19 \pm 3$ µM) of NMV to the monomeric MPro1–199 (summarized in Table S1). This difference is 36-fold for MPro1–199. Lack of discernable dimerization of MPro1–196 in the presence of inhibitors up to ~200 µM suggests that the N-terminal residues 5–9, for which density is accounted for in the inhibitor bound crystal structures of MPro1–199 and MPro1–196, play a critical role in stabilizing the catalytically active conformations of E* and/or E*I of the protein, the latter being more predisposed to dimer formation.

Fig. 8 Inhibitor-binding induced conformational change of the oxyanion loop and dimerization of MPro1–199 and its analogs. E, E* and I denote active site conformation of a monomer (E, inactive state) which is in equilibrium with an active state (E*) resembling the active dimer, and inhibitor, respectively. Inhibitor bound active states before (E*I) and after (E–I) covalent bond formation. $K_{d_{\text{dimer}}}^{\text{app}}$ is slow for NMV because of adduct formation in solution and not for GC373 for which no adduct could be observed by RPLC-MS.

Conclusions

The structural studies presented above indicate that the catalytic domain encompassing the region 1–199 of MPro adopts a native-like fold for residues 7–188 with an unwound oxyanion loop conformation (E), which defines the catalytically inactive state of monomeric MPro. The observed first-order dependency of the rate of hydrolysis by the catalytic domain on the protein concentration indicates that the monomeric catalytic domain exists in at least two conformers. A major inactive conformer (E) which is in dynamic equilibrium with an enzymatically active minor conformer (E*) having an active site oxyanion hole similar in conformation to that of the MProWT dimer. Covalent inhibitors GC373 and NMV bind to the monomeric catalytic domain, as shown from our crystal structures and ITC data. Structural studies of the catalytic domain in complex with GC373 and NMV show the native E*I conformation as well as the appearance of electron density for the S5 loop up to residue 196. The E*-NMV complex is predisposed to dimerize at a lower protein and inhibitor concentration because of its enhanced conformational stability, relative to the E*-GC373 complex. While NMV binds to MPro1–199 displaying a well-organized structure for the N-terminal residues 5–9 as well as facilitate dimer formation, NMV binds very weakly to MPro1–306 with no detectable dimer formation. This result confirms that in addition to residues 5–9, interactions of domain III with residues 1–4 further enhance dimer stability.

The observed autoprocessing at the N-terminus of the mini-precursor (~25) MPro1–199 to produce the catalytic domain MPro1–199 indicates that both, the catalytic domain, and its mini-precursor are enzymatically active exhibiting an E* equilibration. Deletion of the S5 loop (residues 189–194) drastically impairs the N-terminal autoprocessing. Based on the ability of substrates and inhibitors to promote an equilibrium shift to the E* form3,13,36,31,47, we propose that autoprocessing of MPro from its precursor polypeptide may also be governed by the binding of the N-terminal cleavage site sequence (nsp4/nsp5). This binding stabilizes the conformation of the oxyanion loop in the E* conformation leading to the liberation of its own free N-terminus. Thus, the conformational stability of the mature dimer is a collective effect of several interactions such as the interface formed between the N-terminal residues 1–9 with the reoriented domain III as well as the capping of the S1 subsite through the interaction of S1 residue of one subunit with F140/E166 of the second subunit when present in an active E* loop conformation. The E state of monomeric MPro or its polypeptide precursor present a strategic target for inhibitor design.
of MPro followed by GP-6His to facilitate its purification following removal of GP-6His via human rhinovirus 3C (HRV-3C) protease cleavage. Construct (−25)MPro1–199 comprises residues of MPro fused to 25 amino acids of the native flanking C-terminal residues of nps4. This construct contains a 6His at both ends as shown in Fig. S1. A similar construct (−25)MPro1–199 bearing an active site C145A mutation but with only the 6His at the C-terminus was also synthesized and cloned. Construct (−25)MPro1–196 is similar to (−25)MPro1–199 with a 6His-tag only at its N-terminus. Construct (−25)MPro1–187 spans residues 1–187 of MPro fused to 25 amino acids of the native flanking C-terminal residues of nps4. Amino acid sequence and designations of all MPro constructs used in this study are listed in Fig. S1.

Expression and purification. Plasmids were transformed into BL21-DE3 cells (Agilent) and induced for expression at 0.7–0.8 optical density with 1 mM isopropyl β-D-1-thiogalactopyranoside typically for 3 h. Proteins were purified from the cell lysates by nickel-affinity chromatography (Ni-NTA, step 1). The event was subject to isocratic fractionation on Superose-12 column (step 2, Cytiva Life Sciences) and HRV-3C protease cleavage (step 3, purchased from Sigma-Aldrich) at 28 °C. For details, see references3,13,54. The substrate was dissolved in Tris-HCl (pH 8), 150 mM NaCl, 20 mM imidazole and 2 mM 2-mercaptoethanol. (Malvern Instruments Inc., Westborough, MA). Data were processed using the Origin software provided with the instrument. For competitive inhibitors that bind at only one site, the dissociation constant (Kd = 1/κi) is equivalent to the inhibition constant measured by enzyme kinetics (Ki).

Protein crystallization and room-temperature X-ray crystallography. MPro1–199, MPro1–196 and MProWT protein samples were concentrated to 5–8 mg/mL. Stocks of inhibitors were prepared at 50 mM NMV in 100% dimethyl sulfoxide (DMSO) and 10 mM GC373 in buffer C for crystallization purposes and stored at −30 °C. For co-crystallization, MPro1–196 was mixed with GC373 or NMV, and MProWT with GC373, at 1:3 molar ratio and allowed to incubate at room temperature for a minimum of 60 min before setting up crystal trays. Crystals were grown by sitting drop vapor diffusion method with 18–21% PEG3350, 0.1 M Bis-Tris pH 6.5 or pH 7.0 (1 ml) as the precipitant solution. Crystallization drops of 20 µl at 1:1 ratio were seed struck using the crystals of the native MPro in complex with noncovalent ligand Mcule-9487700074 as described.12,13. Crystals of MPro1–199 appeared after several weeks and grew to the final size in about 2 months at 10°C. Co-crystals of MPro1–199 in complex with GC376 and NMV grew in several days of incubation at 14 °C. Crystals were mounted in MiTeGen (Itahca, NY) room-temperature capillary setups for data collection.

All room temperature X-ray crystallographic data were collected with a Rigaku HighFlux HomeLab instrument equipped with a MicroMax-007 HF X-ray generator, Osmic VarioMax optics, and a DÉCTRIS Eiger R 4 M hybrid photon counting detector. X-ray diffraction data integration were performed using the scitools software suite (Rigaku Inc., The Woodlands, TX) then reduced and scaled using Aimless from the CCP4 suite.35. Structures were solved by molecular replacement using Phaser,36 PDB code 2QCY37 was used as a search model to solve the inhibitor-free structure MPro1–199 by first truncating the sequence in the model up to residue 6 at the N-terminus. Crystal structures of MPro1–199, GC373 and MPro1–199-NMV complexes were solved similarly using PDB code 6XQI.39. A full-length model from PDB 6XQU was used to solve the structure of MProWT-GC373. Each model was iteratively refined with Phenix.refine from the PHENIX suite and COOT.40. Geometry validation was performed with Molprobity.41 All ligands were energy minimized by using geometry optimized by quantum mechanical calculations in Gaussian 16 at B3LYP/6-31g(d) level of theory.42 Final data collection and refinement statistics can be found in Table 2.

Reversed-phase liquid chromatography-mass spectrometry (RPLC-MS). Proteins and their inhibitor complexes were subjected to mass spectrometry, using the Thermos DIONEX Ultimate 3000 HPLC system and Thermo MSQ Plus single quadrupole mass spectrometer. Typically, 10 µl of sample at 20 µM concentration, diluted from a reaction mixture in 5% acetic acid, was loaded onto an Acclaim PepMap 300 C4 column (1.0 × 15 mm, Thermo Fisher Scientific) at 40 °C, with 0.2 ml/min flow rates in 2% acetonitrile/95% water/0.01% TFA. The column was washed for 10 min and the bound protein was fractionated using a linear solvent gradient from 2% acetonitrile/95% water/0.01% TFA to 60% acetonitrile/40% water/0.01% TFA over 25 min. Chromatome software provided with the instrument and MagTran (Amgen) were used to analyze the data and estimate mass.

Statistics and reproducibility. Expressed proteins were verified both by DNA sequencing and mass spectrometry. The reproducibility of enzyme kinetics was tested at least 2–3 times with freshly prepared enzyme and stock solutions of the enzyme and inhibitor. Once this was determined to provide consistent reaction rates within an error limit of 5%, the final experiment for the data displayed in the manuscript was carried out in duplicate and 4 reads per well for each time point. The mean of the data points was used for fitting. The same stock solutions of enzyme and inhibitor were used for SV-AUC and ITC analyses to determine the dimer dissociation constant (Kd, native) and the binding constant of the inhibitor to the enzyme (Kd, mutant). Kd, native and Kd, mutant and molecular mass were determined with multiple protein constructs (Fig. S1) and concentrations. Each ITC experiment was carried out with a minimum of 20 injections. The apparent dimer dissociation constant (Kd) was estimated from the nuclear magnetic resonance modeling of the absorbance and interference data. X-ray diffraction data and refinement statistics are shown. Gel images are best representative for each of the construct analyzed.

54 MPro1–199, MPro1–196 and MProWT protein samples were concentrated to 5–8 mg/mL. Stocks of inhibitors were prepared at 50 mM NMV in 100% dimethyl sulfoxide (DMSO) and 10 mM GC373 in buffer C for crystallization purposes and stored at −30 °C. For co-crystallization, MPro1–196 was mixed with GC373 or NMV, and MProWT with GC373, at 1:3 molar ratio and allowed to incubate at room temperature for a minimum of 60 min before setting up crystal trays. Crystals were grown by sitting drop vapor diffusion method with 18–21% PEG3350, 0.1 M Bis-Tris pH 6.5 or pH 7.0 (1 ml) as the precipitant solution. Crystallization drops of 20 µl at 1:1 ratio were seed struck using the crystals of the native MPro in complex with noncovalent ligand Mcule-948770004 as described.12,13. Crystals of MPro1–199 appeared after several weeks and grew to the final size in about 2 months at 10°C. Co-crystals of MPro1–199 in complex with GC376 and NMV grew in several days of incubation at 14 °C. Crystals were mounted in MiTeGen (Itahca, NY) room-temperature capillary setups for data collection.

All room temperature X-ray crystallographic data were collected with a Rigaku HighFlux HomeLab instrument equipped with a MicroMax-007 HF X-ray generator, Osmic VarioMax optics, and a DÉCTRIS Eiger R 4 M hybrid photon counting detector. X-ray diffraction data integration were performed using the scitools software suite (Rigaku Inc., The Woodlands, TX) then reduced and scaled using Aimless from the CCP4 suite.35. Structures were solved by molecular replacement using Phaser,36 PDB code 2QCY37 was used as a search model to solve the inhibitor-free structure MPro1–199 by first truncating the sequence in the model up to residue 6 at the N-terminus. Crystal structures of MPro1–199, GC373 and MPro1–199-NMV complexes were solved similarly using PDB code 6XQI.39. A full-length model from PDB 6XQU was used to solve the structure of MProWT-GC373. Each model was iteratively refined with Phenix.refine from the PHENIX suite and COOT.40. Geometry validation was performed with Molprobity.41 All ligands were energy minimized by using geometry optimized by quantum mechanical calculations in Gaussian 16 at B3LYP/6-31g(d) level of theory.42 Final data collection and refinement statistics can be found in Table 2.

Reversed-phase liquid chromatography-mass spectrometry (RPLC-MS). Proteins and their inhibitor complexes were subjected to mass spectrometry, using the Thermos DIONEX Ultimate 3000 HPLC system and Thermo MSQ Plus single quadrupole mass spectrometer. Typically, 10 µl of sample at 20 µM concentration, diluted from a reaction mixture in 5% acetic acid, was loaded onto an Acclaim PepMap 300 C4 column (1.0 × 15 mm, Thermo Fisher Scientific) at 40 °C, with 0.2 ml/min flow rates in 2% acetonitrile/95% water/0.01% TFA. The column was washed for 10 min and the bound protein was fractionated using a linear solvent gradient from 2% acetonitrile/95% water/0.01% TFA to 60% acetonitrile/40% water/0.01% TFA over 25 min. Chromatome software provided with the instrument and MagTran (Amgen) were used to analyze the data and estimate mass.

Statistics and reproducibility. Expressed proteins were verified both by DNA sequencing and mass spectrometry. The reproducibility of enzyme kinetics was tested at least 2–3 times with freshly prepared enzyme and stock solutions of the enzyme and inhibitor. Once this was determined to provide consistent reaction rates within an error limit of 5%, the final experiment for the data displayed in the manuscript was carried out in duplicate and 4 reads per well for each time point. The mean of the data points was used for fitting. The same stock solutions of enzyme and inhibitor were used for SV-AUC and ITC analyses to determine the dimer dissociation constant (Kd, native) and the binding constant of the inhibitor to the enzyme (Kd, mutant). Kd, native and Kd, mutant and molecular mass were determined with multiple protein constructs (Fig. S1) and concentrations. Each ITC experiment was carried out with a minimum of 20 injections. The apparent dimer dissociation constant (Kd) was estimated from the nuclear magnetic resonance modeling of the absorbance and interference data. X-ray diffraction data and refinement statistics are shown. Gel images are best representative for each of the construct analyzed.
Table 2 Crystallographic data collection and refinement statistics.

|                  | MPro1<sup>199</sup> | MPro1<sup>196-596-GC373</sup> | MPro1<sup>196-NMV</sup> | MProWT<sup>196-GC373</sup> |
|------------------|----------------------|-------------------------------|--------------------------|-------------------------------|
| Data collection  | X-ray (in-house)     | X-ray (in-house)              | X-ray (in-house)         | X-ray (in-house)              |
| Diffractometer   | Rigaku HighFlux, Eiger R 4M | P2<sub>1</sub>, P2<sub>1</sub> | P2<sub>1</sub>, P2<sub>1</sub> | P2<sub>1</sub> |
| Space group      | P2<sub>1</sub>, P2<sub>1</sub>, P2<sub>1</sub> | P2<sub>1</sub>, P2<sub>1</sub> | P2<sub>1</sub>, P2<sub>1</sub> | P2<sub>1</sub> |
| Wavelength (Å)   | 1.5406               | 1.5406                        | 1.5406                   | 1.5406                        |
| Cell dimensions: | a, b, c (Å)          | 53.34, 63.09, 121.35          | 52.79, 61.93, 58.94     | 52.73, 62.61, 58.96          |
|                  | α, β, γ (°)          | 90, 90, 90                   | 90, 97.8, 90            | 90, 99.1, 90                 |
|                  | Resolution (Å)       | 121.35–2.25 (2.33–2.25)      | 61.93–1.80 (1.87–1.80)  | 62.60–1.85 (1.92–1.85)       |
| No. reflections  | 20,094 (2012)        | 33,340 (3207)                | 32,399 (315)            | 18,348 (1840)                |
| R<sub>work</sub> | 0.137 (0.827)        | 0.038 (0.212)                | 0.072 (0.490)           | 0.097 (0.375)                |
| CC<sub>r</sub>   | 0.952 (0.798)        | 0.996 (0.945)                | 0.997 (0.650)           | 0.994 (0.854)                |
| Completeness (%) | 99.7 (99.6)          | 95.1 (91.5)                  | 99.3 (95.7)             | 99.9 (99.9)                  |
| Redundancy       | 7.0 (7.1)            | 4.8 (3.4)                    | 5.1 (3.5)               | 5.0 (4.8)                    |
| Refinement       | R<sub>work</sub> / R<sub>free</sub> | 0.2043/0.2502 | 0.1517/0.1821 | 0.1716/0.2067 |
| B-factors (Å²)  | Protein              | 52.31                        | 23.11                    | 28.05                        |
|                  | Ligand               | N/A                          | 24.20                    | 27.67                        |
|                  | Water                | 43.71                        | 34.20                    | 37.95                        |
|                  | R.M.S. deviations    | 0.004                        | 0.007                    | 0.009                        |
|                  | Bond lengths (Å)     | 0.677                        | 0.958                    | 1.090                        |
|                  | Bond angles (°)      | 4.84                         | 2.21                     | 2.87                         |
|                  | All atom clash score | 1.88                         | 1.88                     | 1.88                         |

Data reduction and refinement statistics for the room temperature X-ray crystal structures of SARS-CoV-2 MPro1<sup>199</sup> and MPro1<sup>196-196</sup> in complex with GC373 and NMV, respectively, and MProWT<sup>196</sup> in complex with GC373. Values in parentheses are for the highest-resolution shell.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Source data files are provided in Supplementary Data 1 and 2.

Received: 23 June 2022; Accepted: 29 August 2022; Published online: 16 September 2022

References
1. Vršanský, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. Nat. Rev. Microbiol. 19, 155–170 (2020).
2. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. Nature 579, 265–269 (2020).
3. Zhang, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science 368, 409–412 (2020).
4. Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M. & Bergeron, J. R. C. Structural variation analysis. Front. Mol. Biosci. 7, 605236 (2020).
5. Wang, C. et al. The establishment of reference sequence for SARS-CoV-2 and characterization of SARS-CoV-2: where we are, and where we need to be. J. Virol. 89, 2652–2658 (2020).
6. Gharehamanpour, M. M. et al. Identification of 14 known drugs as inhibitors of the main protease of SARS-CoV-2. BioRxiv (2020).
7. Baker, J. D., Uhrich, R. L., Kraemer, G. C., Love, J. E. & Kraemer, B. C. A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. PLoS One 16, 1–13 (2021).
8. Kneller, D. W. et al. Structural, electronic, and electrostatic determinants for inhibitor binding to s1 and s2 in SARS-CoV-2 main protease. J. Med. Chem. 64, 17366–17383 (2021).
9. Owen, D. R. et al. An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. Science 374, eab4784 (2021).
27. Shi, J., Swaraman, J. & Song, J. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. J. Virol. 82, 4620–4629 (2008).
28. Chen, S. et al. Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. J. Biol. Chem. 283, 554–564 (2008).
29. Hu, T. et al. Two adjacent mutations on the dimer interface of SARS coronavirus 3C-like protease cause different conformational changes in crystal structure. Virology 388, 324–334 (2009).
30. Cheng, S. C., Chang, G. G. & Chou, C. Y. Mutation of Glu-166 blocks the irreversible hydrolytic metalloenzymes by a unified mechanistic approach. J. Am. Chem. Soc. 24, 145–152 (1991).
31. Wu, C. G. et al. Mechanism for controlling the monomer-dimer conversion of SARS coronavirus main protease. Biochemistry 43, 14958–14970 (2004).
32. Chou, C. Y. et al. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. Biochemistry 43, 14958–14970 (2004).
33. Kneller, D. W. et al. Malleability of the SARS-CoV-2 3C-like protease active-site cavity facilitates binding of clinical antivirals. Structure 28, 1313–1320.e1313 (2020).
34. Noske, G. D. et al. A crystallographic snapshot of SARS-CoV-2 main protease maturation process. J. Mol. Biol. 433, 167118 (2021).
35. Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S. C. & Mesecar, A. D. Evaluating the 3C-like protease activity of SARS-CoV-2 protease recombinations for standardized assays for drug discovery. Virus Res. 133, 63–73 (2008).
36. Wang, Y. C. et al. Structural basis of SARS-CoV-2 main protease inhibition by a broad-spectrum antiviral drug. Am. J. Cancer Res. 10, 2535–2540 (2020).
37. Vuong, W. et al. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. Nat. Commun. 11, 4282 (2020).
38. Chou, J. Development of artificial hydrolytic metalloenzymes by a unified mechanistic approach. J. Am. Chem. Soc. 24, 145–152 (1991).
39. Kneller, D. W. et al. Unusual zwiterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. J. Biol. Chem. 295, 17365–17373 (2020).
40. Kneller, D. W., Zhang, Q., Coates, L., Louis, J. M. & Kovalevsky, A. Michaelis-like complex of SARS-CoV-2 main protease visualized by room-temperature X-ray crystallography. IUCrJ 9, 973–979 (2021).
41. Chaturvedi, R. K., MacMahon, A. E. & Schmir, G. Hydrolysis of thioimidate thiopeptide and thionester analogs of specific substrates. An effect on kcat for f dic substrates. An effect on kcat for f dic substrates. Kinetics; A.A. and J.M.L. constructed, expressed, and purified proteins; D.W.K. and A.K. crystallized the protein and inhibitor complexes; D.W.K., L.C. and A.K. collected X-ray data, reduced the data and refined the structures; R.G. and J.M.L. performed SV-AUC and SEC-MALS and analyzed data; J.M.L. performed ITC and analyzed data; N.T.N., A.K. and J.M.L. analyzed data and wrote the paper with help from all co-authors.

Competing interests: The authors declare no competing interests.

Supplementary information: the online version contains supplementary material available at https://doi.org/10.1038/s42003-022-03910-y.

Correspondence and requests for materials should be addressed to Andrey Kovalevsky or John M. Loun.

Peer review information: Communications Biology thanks the anonymous reviewers for their contribution to the peer review of this work. Chief Handling Editor: Gene Chong.

Reprints and permission information: are available at http://www.nature.com/reprints

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access: This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

This is a U.S. Government work and not under copyright protection in the US. Foreign copyright protection may apply 2022