



## Insights into the mechanism of SARS-CoV-2 main protease autocatalytic maturation from model precursors

Annie Aniana<sup>1</sup>, Nashaat T. Nashed <sup>1</sup>, Rodolfo Ghirlando <sup>2</sup>, Leighton Coates <sup>3</sup>, Daniel W. Kneller<sup>4,5</sup>, Andrey Kovalevsky <sup>4</sup>✉ & John M. Louis <sup>1</sup>✉

A critical step for SARS-CoV-2 assembly and maturation involves the autoactivation of the main protease (MPro<sup>WT</sup>) from precursor polyproteins. Upon expression, a model precursor of MPro<sup>WT</sup> mediates its own release at its termini rapidly to yield a mature dimer. A construct with an E290A mutation within MPro exhibits time dependent autoprocessing of the accumulated precursor at the N-terminal nsp4/nsp5 site followed by the C-terminal nsp5/nsp6 cleavage. In contrast, a precursor containing E290A and R298A mutations (MPro<sup>M</sup>) displays cleavage only at the nsp4/nsp5 site to yield an intermediate monomeric product, which is cleaved at the nsp5/nsp6 site only by MPro<sup>WT</sup>. MPro<sup>M</sup> and the catalytic domain (MPro<sup>1-199</sup>) fused to the truncated nsp4 region also show time-dependent conversion in vitro to produce MPro<sup>M</sup> and MPro<sup>1-199</sup>, respectively. The reactions follow first-order kinetics indicating that the nsp4/nsp5 cleavage occurs *via* an intramolecular mechanism. These results support a mechanism involving an N-terminal intramolecular cleavage leading to an increase in the dimer population and followed by an intermolecular cleavage at the C-terminus. Thus, targeting the predominantly monomeric MPro precursor for inhibition may lead to the identification of potent drugs for treatment.

<sup>1</sup>Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, DHHS, Bethesda, MD 20892-0520, USA. <sup>2</sup>Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, DHHS, Bethesda, MD 20892-0520, USA. <sup>3</sup>Second Target Station, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA. <sup>4</sup>Neutron Scattering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA. <sup>5</sup>Present address: New England Biolabs, 240 County Road, Ipswich, MA 01938-2723, USA. ✉email: [kovalevskyay@ornl.gov](mailto:kovalevskyay@ornl.gov); [johnl@niddk.nih.gov](mailto:johnl@niddk.nih.gov)

Specific cleavages of the polyproteins catalyzed by the virally encoded protease to release the mature structural and functional proteins are indispensable for viral assembly and maturation<sup>1–3</sup>. A single copy of the main protease of SARS-CoV2 (nsp5 or MPro) is translated from overlapping reading frames in two large polyproteins (pp) 1a and 1ab that encompass all non-structural proteins (nsp)s<sup>4–6</sup>. A critical step in the regulation of MPro is the formation of a fully active mature MPro *via* autoprocessing at its termini<sup>2,7</sup>. MPro of SARS-CoV and SARS-CoV-2 are 96% identical in sequence and thus are similar in structure and function<sup>8</sup>. Both enzymes are ~68 kDa homodimers with each protomer having one active site C145-H41 catalytic dyad<sup>7,9</sup>. The protomer is composed of 306 amino acids and organized in three domains, I–III. Domains I (residues 8–101) and II (102–184) exhibit a chymotrypsin-like fold and connected to the helical domain III (residues 201–306) through a long loop region (residues 185–200)<sup>9,10</sup>. Structural studies indicate that monomeric forms are similar in their tertiary fold to the wild-type dimer except for differences in the active site oxyanion loop, the N-finger region and its interface, and domain III orientation<sup>7,8,11</sup>. Upon dimerization, the oxyanion loop (residues 137–144) assumes an active wound conformation (E\*), whereas it is predominantly in an inactive unwound conformation (E) in the monomeric form<sup>7,11,12</sup>. Dimer interface mutants that are monomeric or predominantly monomeric bind to known inhibitors of wild-type MPro accompanied by the reorganization of the oxyanion loop to the active E\* conformation typical of the dimer<sup>11,13</sup>. Also, inhibitor or substrate favor binding to the dimeric form of the enzyme<sup>13,14</sup>.

The spatio-temporal regulation of MPro release from its polyprotein precursor *in vivo* is predicted to be complex because of the association of its flanking nsp4 and nsp6 proteins with membranes<sup>6,15–17</sup>. Studies aimed at understanding the early mechanisms of maturation of the viral proteases from polyproteins are also complicated by the fact that polyproteins contain multiple cleavage sites and undergo autoprocessing during expression and purification<sup>6</sup>. In addition, our attempts to refold MPro from chemical denaturants were unsuccessful for *in vitro* studies. In this study, we have engineered several model precursors containing mutations that restrict dimer formation and thus facilitate the accumulation and purification of intact precursors from *E. coli*. These precursors undergo time-dependent autoprocessing reaction mediating cleavages at the termini of MPro in *E. coli* and *in vitro* and thus, provide simple models to study the mechanisms of MPro autocatalytic maturation.

## Results

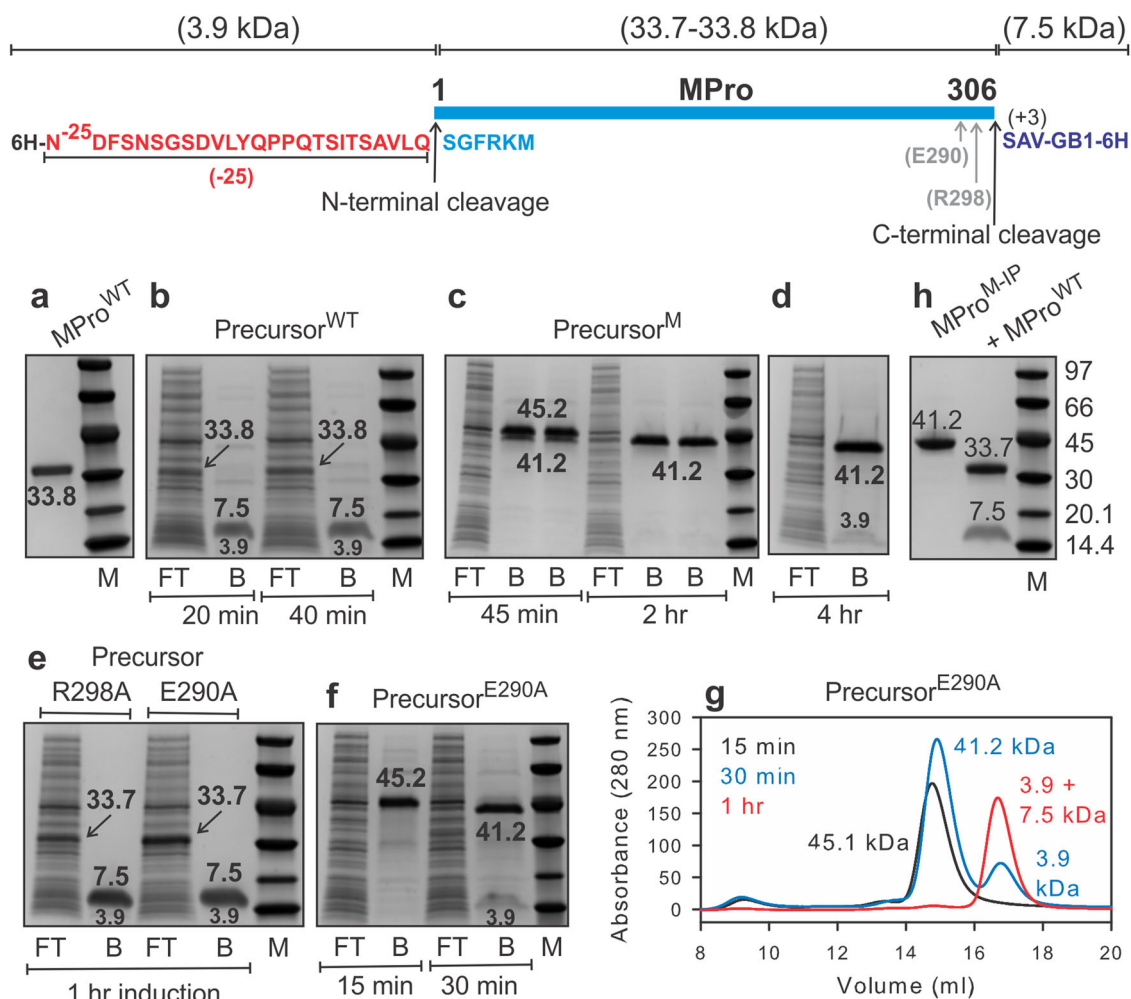
**Order of cleavage at the termini of MPro from model precursors in *E. coli*.** Our simplified strategy entails expressing MPro (nsp5) containing flanking region sequences of nsp4 and nsp6 as a precursor mimetic. Recent studies indicate that the addition of non-native residues to the N-terminus of MPro results in a drastic decrease in catalytic activity as compared to such residues appended to the C-terminus of MPro<sup>18</sup>. The decrease in catalytic activity is consistent with our earlier observation that the addition of six native nsp4 residues (TSAVLQ) to the N-terminus of MPro increases the dimer dissociation constant<sup>12</sup>. These results imply that the reorganization of the free N-terminal residues of MPro upon N-terminal autoprocessing (nsp4/nsp5 site) through specific key interactions with domain III leads to enhanced dimer stability and appearance of mature-like catalytic activity, similar to that previously described for SARS-CoV MPro<sup>7,19</sup>. Thus, our design strategy for MPro precursor contained 25 C-terminal residues of nsp4 appended to the N-terminus of MPro [<sup>(–25)</sup>MPro] to enable discerning the migration of the precursor from the mature MPro

upon autoprocessing by SDS-PAGE without compromising solubility. Likewise, extending the length of the flanking nsp4 even up to 102 residues (<sup>(–102)</sup>MPro) resulted in the production of soluble protein and with a  $K_{\text{dimer}}$  that is predicted to be like that of (<sup>(–25)</sup>MPro<sup>11,13</sup>. Constructs and their designations used in this study are shown in the corresponding main figures and listed in Figure S1. Since appending even 25 residues of nsp6 to MPro causes insolubility of the expressed protein, the soluble immunoglobulin binding domain B1 of Protein G [GB1<sup>20</sup>], preceded by 3 N-terminal residues of nsp6 (SAV) to retain nsp5/nsp6 cleavage site specificity, were used (Fig. 1). This strategy allows monitoring native cleavages at both termini of MPro as a fully soluble precursor in *E. coli* and *in vitro*. Expression of Precursor<sup>WT</sup> results in rapid autoprocessing at both sites within 20 min and no full-length precursor or intermediate product is observed (Fig. 1b). The products corresponding to mature MPro<sup>WT</sup> (33.8 kDa, Fig. 1a) and the terminal fragments were identified (Fig. 1b), and the purified terminal fragments were confirmed by mass spectrometry.

Since the catalytic activity inversely correlates with the dimer dissociation constant [ $K_{\text{dimer}}$ <sup>11,21</sup>], it is predictable that mutations of residues contributing to dimer formation would enable the production of a functional precursor for monitoring the autoprocessing reaction. Thus, interface residues E290 and R298 were targeted for mutations in Precursor<sup>WT</sup>. In earlier studies, we had shown that these mutations together increase the  $K_{\text{dimer}}$  of mature MPro by 5000-fold<sup>13</sup> and accordingly this mutant was termed MPro<sup>M</sup> (where <sup>M</sup> denotes monomer). Precursor<sup>R298A</sup> undergoes rapid autoprocessing at both termini, like that of Precursor<sup>WT</sup>, i.e., complete conversion within 1 h of expression to release the mature MPro<sup>R298A</sup>, 3.9 kDa and 7.5 kDa products (Figs. 1e, 2b). In contrast, Precursor<sup>E290A</sup> processing is observed to be much slower than Precursor<sup>WT</sup> and Precursor<sup>R298A</sup> allowing the transient observation of the time-dependent stepwise autoprocessing at both termini (see Fig. 1f). The expressed protein being fully soluble enabled the purification of the full-length protein and its product by nickel-affinity chromatography (NAC) expeditiously under native conditions. At 15 min, only Precursor<sup>E290A</sup> (45.1 kDa) is detectable in the bound (B) fraction, followed by the appearance of the intermediate species of 41.2 kDa and the 3.9 kDa corresponding to the products of the N-terminal cleavage at 30 min. The 41.2 kDa product is then converted to the mature MPro<sup>E290A</sup> and the 7.5 kDa product upon C-terminal cleavage at 1 h of induction (Fig. 1e). This is further confirmed by analyses of the bound fractions at 15, 30 and 60 min by size-exclusion chromatography (Fig. 1g) and mass spectrometry.

The double mutant E290A and R298A precursor (Precursor<sup>M</sup>) also exhibits its accumulation in *E. coli* and time-dependent autoprocessing reaction, but at a slower rate than that of Precursor<sup>E290A</sup>. Analysis of the fractions indicates cleavage at the N-terminus of MPro<sup>M</sup> leading to the accumulation of the intermediate product (MPro<sup>M-IP</sup>, 41.2 kDa, Fig. 1c) after 2 h of expression with no detectable product indicative of a C-terminal cleavage even after 4 h (Fig. 1d). The identity of MPro<sup>M-IP</sup> was verified by mass spectrometry (41,180 Da, calculated = 41,180 Da). Nevertheless, incubation of MPro<sup>M-IP</sup> with MPro<sup>WT</sup> at a ratio of 50:1 produced the previously characterized mature MPro<sup>M</sup><sup>13</sup> indicating that the C-terminal cleavage site is accessible for reaction (Fig. 1h).

**N-terminal autoprocessing of MPro precursor analogues in *E. coli*.** Like Precursor<sup>WT</sup>, expression of MPro<sup>WT</sup> with only the native nsp4/nsp5 cleavage site also results in the lack of accumulation of the precursor, with autoprocessing complete within 15 min of expression to release mature MPro<sup>WT</sup> [Fig. S2a<sup>11</sup>]. An active site C145A mutation abolishes this cleavage (Fig. S2a<sup>11</sup>).

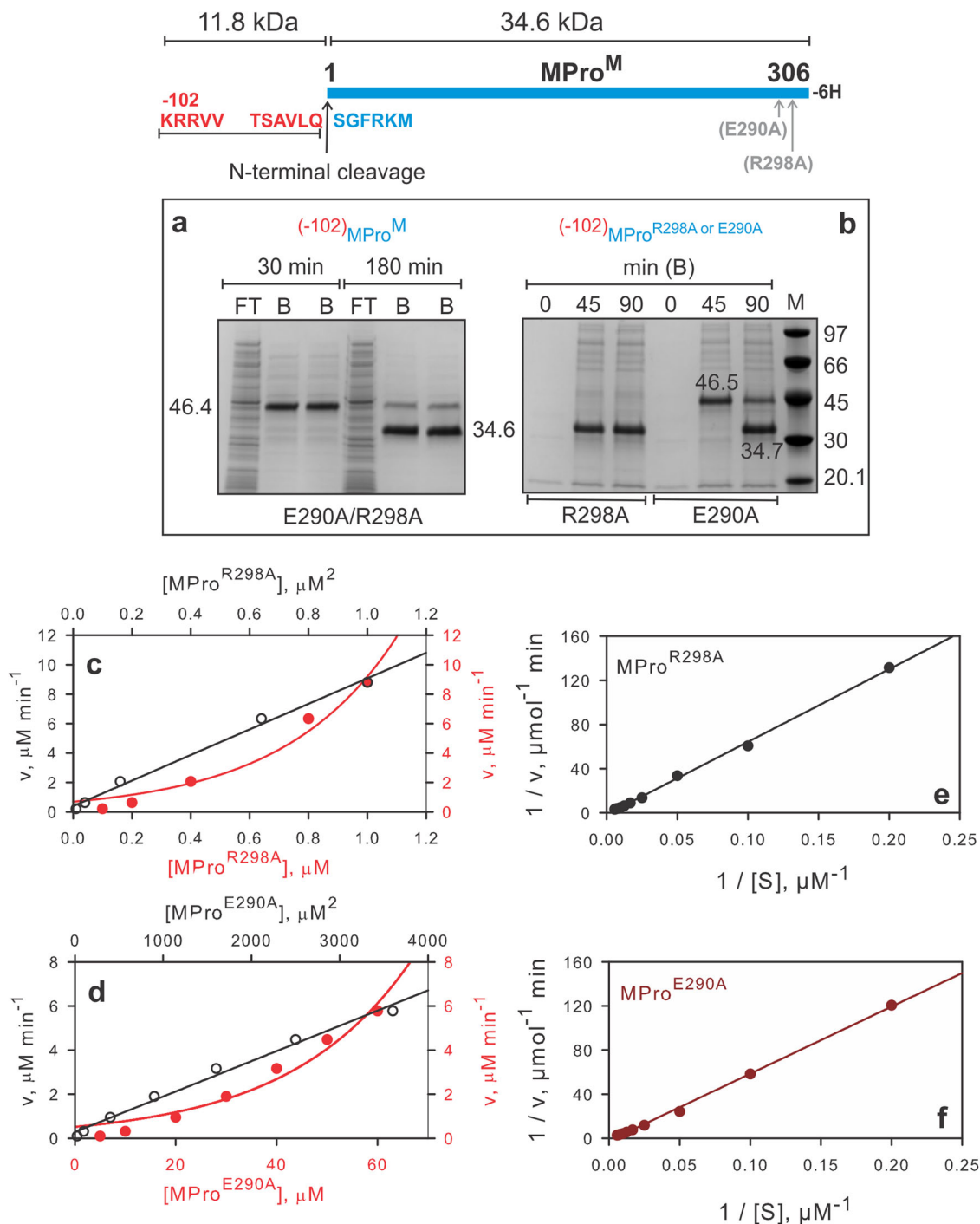


**Fig. 1 Autoprocessing of precursor mimetics of MPro containing native cleavage sites at both termini in *E. coli*.** The MPro model precursor comprising the nsp4/nsp5 and nsp5/nsp6 cleavage sites is shown on top. **a–f** 12 ml of cells were harvested at the time points indicated below the lanes, subjected to NAC and equal amounts (3–4  $\mu$ g) of the flow-through (FT) and bound (B) fractions were analyzed by SDS-PAGE. Molecular weight standards (M) are indicated in kDa. **a** Migration of purified mature MPro<sup>WT</sup> (33.8 kDa). **b** Autoprocessing of precursor<sup>WT</sup> results in 33.8 kDa mature MPro<sup>WT</sup> in the FT, and 3.9 kDa and 7.5 kDa products in the B fractions. Cleavage products were verified by mass spectrometry. **c** Autoprocessing of precursor<sup>M</sup> (45.2 kDa) results in the appearance of an intermediate product (41.2 kDa, MPro<sup>M-IP</sup>) upon cleavage at the N-terminus of MPro<sup>M</sup>, seen as a doublet together with the full-length protein at 45 min of induction, and only that of MPro<sup>M-IP</sup> at 2 hr of induction in the bound B fractions. **d** No C-terminal cleavage occurs even after 4 h of induction as indicated by the absence of the mature MPro<sup>M</sup> (33.7 kDa) in the FT. **e, f** Autoprocessing of Precursor<sup>R298A</sup> and Precursor<sup>E290A</sup>. **g** Fractionation of the bound fractions of Precursor<sup>E290A</sup> at 15, 30 (f) and 60 min (e) on Superose 12 (1  $\times$  30 cm column) in buffer A confirming the stepwise cleavage at the N-terminus first, followed by the cleavage at the C-terminus as seen on gels. The precursor and products of cleavage were verified by mass spectrometry. **h** Cleavage of the nsp5/nsp6 site of MPro<sup>M-IP</sup> by MPro<sup>WT</sup>.

We had previously reported the expression of a construct consisting of 25 amino acids of the nsp4 sequence appended to the N-terminus of the double mutant MPro<sup>M</sup> as a miniprecursor, (–25)MPro<sup>M</sup>, which undergoes time-dependent autoprocessing upon expression in *E. coli* [Fig. S2b<sup>11</sup>]. The 37.4 kDa precursor is converted to the 34.6 kDa and 2.8 kDa products at a much slower rate relative to the MPro<sup>WT</sup> precursor. Only the 34.6 kDa product (MPro<sup>M</sup>) containing the 6H-tag allows its detection together with the precursor after NAC. To better resolve the separation between the starting material and products of the autoprocessing reaction by SDS-PAGE, a similar set of constructs comprising 102 amino acids of nsp4 appended to the N-terminus of MPro<sup>WT</sup>, MPro<sup>E290A</sup>, MPro<sup>R298A</sup>, and MPro<sup>M</sup> were expressed in *E. coli* (Fig. 2). Like Precursor<sup>WT</sup> and its mutants, expression of (–102)MPro<sup>WT</sup> and (–102)MPro<sup>R298A</sup> (Fig. 2b) resulted in observing only the processed products released upon N-terminal autoprocessing. Time course of the autoprocessing of (–102)MPro<sup>WT</sup> parallels that of the wild-type

precursor shown in Figure S2A. Precursor<sup>(–102)MPro<sup>E290A</sup></sup> and (–102)MPro<sup>M</sup> exhibit transient accumulation of the full-length precursor followed by N-terminal cleavage to yield products (Fig. 2a, b) like that of Precursor<sup>E290A</sup> and Precursor<sup>M</sup> (and (–25)MPro<sup>M</sup>, Fig. S2c), respectively, undergoing N-terminal cleavage.

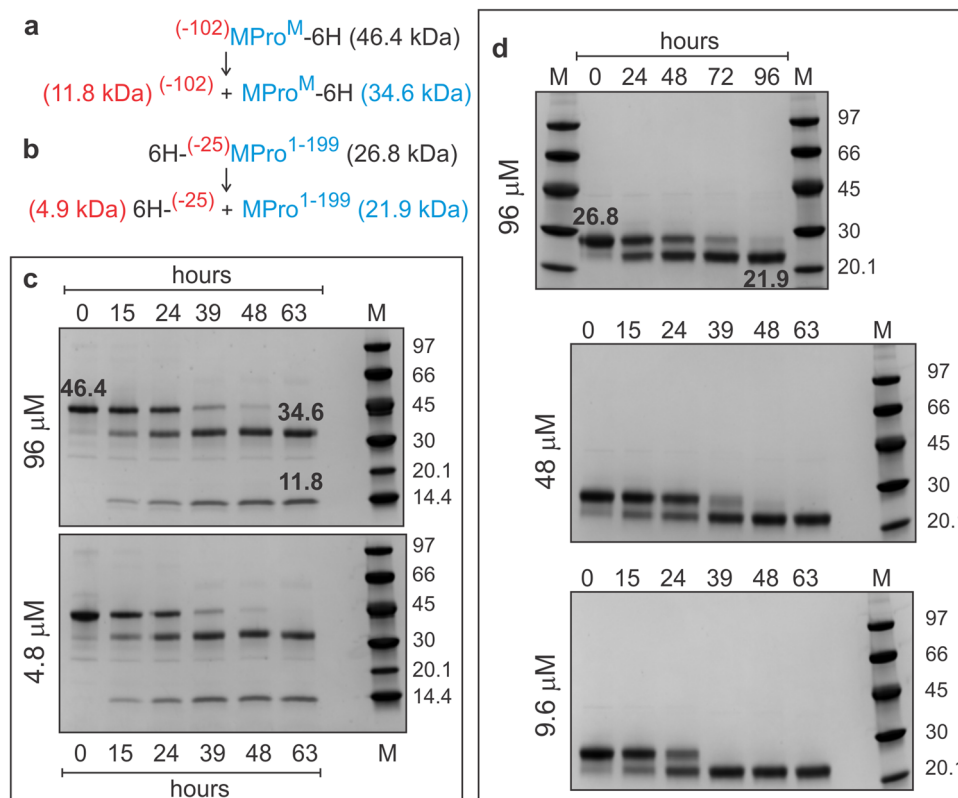
**N-terminal autoprocessing of MPro precursor analogues in vitro.** Interestingly, like (–25)MPro<sup>M</sup>, precursor<sup>(–25)MPro<sup>1–199</sup></sup> lacking the entire helical domain (residues 200–306), as a monomer, also undergoes time-dependent N-terminal autoprocessing upon expression in *E. coli* to produce MPro<sup>1–199</sup> [Fig. S2c<sup>11</sup>]. Transient accumulation of full-length precursors (–102)MPro<sup>M</sup>, (–25)MPro<sup>M</sup> and (–25)MPro<sup>1–199</sup> allowed their purification in a small scale to yield sufficient protein for in vitro analyses. Details of the reaction set-up are shown in Tables S1, S2. Samples collected at various time points were subjected to SDS-PAGE. Figure 3 shows



**Fig. 2 N-terminal autoprocessing of MPro<sup>M</sup> precursor and its single mutants in *E. coli* and characterization of products.** The N-terminal cleavage site is indicated with an upward black arrow. Precursor, products released upon cleavage at the N-terminus of MPro and molecular weight standards (M) are indicated in kDa. The gels show the time course of the autoprocessing reaction of MPro<sup>M</sup> (a) and single mutants R298A and E290A (b) with 102 amino acids of the nsp4 sequence appended to the N-terminus of MPro. Cells (12 ml) were harvested at the time points indicated, subjected to NAC, and analyzed by SDS-PAGE. FT and B denote flow-through and bound fractions, respectively, after NAC. The 11.8 kDa product is not observed in the bound fraction because it lacks the 6His-tag. **c-f** Catalytic activity of mature MPro<sup>R298A</sup> and MPro<sup>E290A</sup>. **c, d** Linear relationship between the rate vs the square of the protein concentration assayed with 200  $\mu\text{M}$  substrate. **e, f** Lineweaver-Burk plots for the hydrolysis of substrate by 0.2  $\mu\text{M}$  MPro<sup>R298A</sup> and 3.5  $\mu\text{M}$  MPro<sup>E290A</sup> in the presence of equimolar concentration of GC373. These concentrations represent the  $K_d$  estimated by ITC like that described for MPro<sup>M</sup><sup>13</sup>. The  $k_{\text{cat}}/K_m$  was calculated as described<sup>13</sup>, and compared with previously characterized constructs shown in Table 2<sup>11,13</sup>.

the time course of the reactions. For  $(-102)\text{MPro}^{\text{M}}$  and  $(-25)\text{MPro}^{199}$ , the starting material and the products are well separated on gels allowing quantifying the band intensities. The percent conversion at 15 and 24 h of incubation are calculated from each

initial protein concentration and shown in Table 1. In addition, precursor  $(-102)\text{MPro}^{\text{M}}$ -catalyzed hydrolysis of the FRET substrate was monitored. The intrinsic catalytic activity was estimated to be 0.0204  $\mu\text{M}/\text{h}$  when assayed with 50  $\mu\text{M}$  precursor and 200  $\mu\text{M}$



**Fig. 3** Time course of the autoprocessing of  $(-102)\text{MPro}^{\text{M}}$  and  $(-25)\text{MPro}^{1-199}$  as a function of decreasing protein concentration. **a**, **b** Precursor constructs showing the products released upon N-terminal autoprocessing. Time course of **(c)**  $(-102)\text{MPro}^{\text{M}}$  and **(d)**  $(-25)\text{MPro}^{1-199}$  autoprocessing. Reactions were initiated at the indicated concentration in buffer B at 28 °C (see Tables S1, S2 for details). Aliquots of the reaction were drawn at the indicated times and terminated by the addition of gel sample buffer. Samples (2.6–3.6 μg/lane) were subjected to SDS-PAGE and visualized by staining. Precursor, reaction products, and molecular weight standards (M) are indicated in kDa.

**Table 1** Time-dependent N-terminal autoprocessing of MPro in vitro.

Construct	$[\text{P}_0]$ μM	At 15 h		At 24 h	
		$\Delta[\text{P}]$ , μM	% Conversion	$\Delta[\text{P}]$ , μM	% Conversion
$(-102)\text{MPro}^{\text{M}}$	96	$36.9 \pm 2$	38.4	$53.7 \pm 3$	55.9
	4.8	$1.93 \pm 0.1$	40.2	$2.8 \pm 0.2$	59.2
$(-25)\text{MPro}^{1-199}$	96			$40.5 \pm 0.4$	42.2
	48	$15.4 \pm 0.2$	32.2	$20.7 \pm 0.2$	43.2
	9.6	$3.1 \pm 0.1$	32.4		

Band intensities shown in Fig. 3 were quantified using the ImageJ software (<https://imagej.nih.gov/ij/>).  $[\text{P}_0]$  denotes the precursor  $(-102)\text{MPro}^{\text{M}}$  (46.4 kDa) and  $(-25)\text{MPro}^{1-199}$  (26.8 kDa) concentrations at time zero and  $\Delta[\text{P}]$  is the amount of  $[\text{P}_0]$  converted to products upon incubation for 15 and 24 h, calculated from the sum of the band intensities at each time point.

substrate. After incubation of  $(-102)\text{MPro}^{\text{M}}$  for 24 h, the observed value is  $0.0384 \mu\text{M}/\text{h}$ , pointing to a 75% increase in catalytic activity attributable to the N-terminal cleavage of  $(-102)\text{MPro}^{\text{M}}$ .

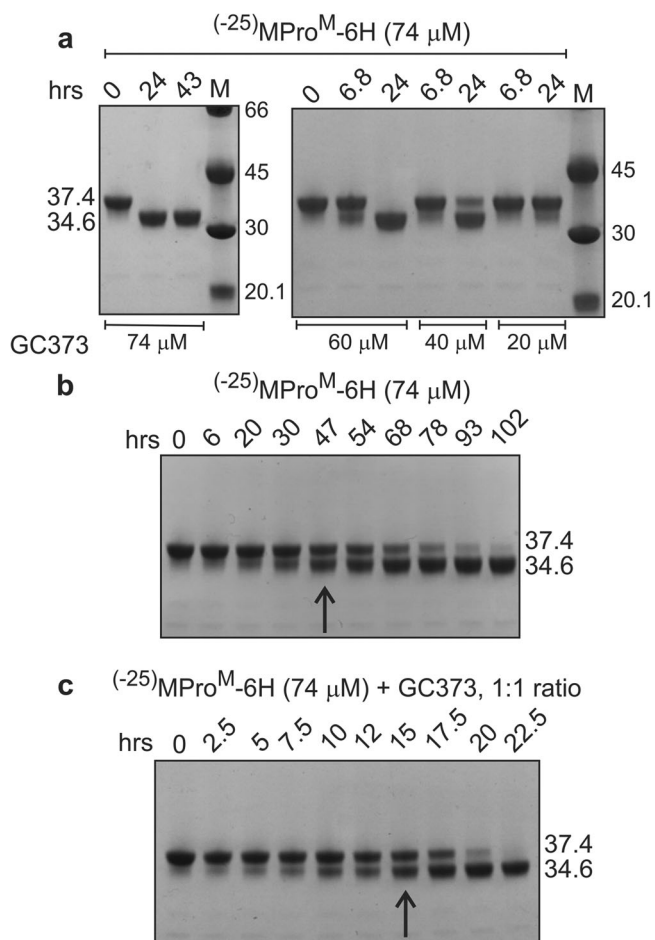
Like  $\text{MPro}^{\text{M}}$ <sup>13</sup>, the autoprocessing of  $(-25)\text{MPro}^{\text{M}}$  precursor is accelerated in a concentration-dependent manner by GC373 up to an equimolar concentration (Fig. 4a). Figure 4b, c show qualitatively that, similar amount of  $\text{MPro}^{\text{M}}$  is produced from  $(-25)\text{MPro}^{\text{M}}$  in 15 h in the presence of equimolar amount of GC373 as compared to that in 47 h without GC373.

#### Characterization of products of the autoprocessing reactions.

The intermediate product  $\text{MPro}^{\text{M-IP}}$  resulting from the N-terminal autoprocessing of Precursor<sup>M</sup>, was purified from cells expressed for 2.5 h as described in Methods. It is monomeric in solution as shown by SV-AUC up to a concentration of  $85 \mu\text{M}$

(Fig. 5a). It catalyzes the hydrolysis of the FRET substrate, and the rate of hydrolysis of the peptide substrate displays first-order kinetics in protein concentration, i.e., a linear dependency on the protein concentration at a constant substrate concentration (Fig. 5b). This further confirmed that the observed catalytic activity is that of a monomer, and no detectable amount of dimer is present in solution. Unlike mature  $\text{MPro}^{\text{M}}$ , the kinetics of  $\text{MPro}^{\text{M-IP}}$  catalyzed hydrolysis is similar to that of monomeric  $\text{MPro}^{1-199}$  [11 and Table 2]. However, like  $\text{MPro}^{\text{M}}$ <sup>13</sup>,  $\text{MPro}^{\text{M-IP}}$  binds to GC373 with concomitant dimer formation and enhanced catalytic activity up to an equimolar concentration of the  $\text{MPro}^{\text{M-IP}}$  and GC373 (Fig. 5c). Above the equimolar amount, the catalytic activity is inhibited by increasing the concentration of GC373. The kinetic parameter  $k_{\text{cat}}/K_{\text{m}}$  for  $\text{MPro}^{\text{M-IP}}$  in the absence and presence of GC373 as well as the binding constant to GC373 determined by SV-AUC and ITC are listed in Table 2 (Fig. 5d, S3, S4) along with those of  $\text{MPro}^{\text{WT}}$ ,  $\text{MPro}^{\text{M}}$  and  $\text{MPro}^{1-199}$ . Although the 7.5 kDa SAV-GB1-6H fragment corresponding to the product of the C-terminal nsp5/nsp6 cleavage of  $\text{MPro}^{\text{M-IP}}$  by  $\text{MPro}^{\text{WT}}$  is easily observed (Fig. 1h), surprisingly no cleavage is detected upon incubation of  $\text{MPro}^{\text{M-IP}}$  in the absence or presence of equimolar amount of GC373 up to 48 h.

The 6H-tag at the C-terminus facilitates the initial purification of the MPro products derived from the autoprocessing of  $(-102)\text{MPro}^{\text{E290A}}$  and  $(-102)\text{MPro}^{\text{R298A}}$  and was removed to produce mature  $\text{MPro}^{\text{E290A}}$   $\text{MPro}^{\text{R298A}}$  as previously described for  $\text{MPro}^{\text{WT}}$ <sup>9,13</sup>,  $\text{MPro}^{\text{R298A}}$  and  $\text{MPro}^{\text{E290A}}$  exhibit a dimer dissociation constant ( $K_{\text{dimer}}$ ) of  $7.1 \pm 0.4 \mu\text{M}$  and  $353 \pm 21 \mu\text{M}$ , respectively, compared to  $6.6 \text{ mM}$  for the double mutant  $\text{MPro}^{\text{M}}$  (Table 2 and Fig. S5). Purified mature  $\text{MPro}^{\text{E290A}}$  and  $\text{MPro}^{\text{R298A}}$



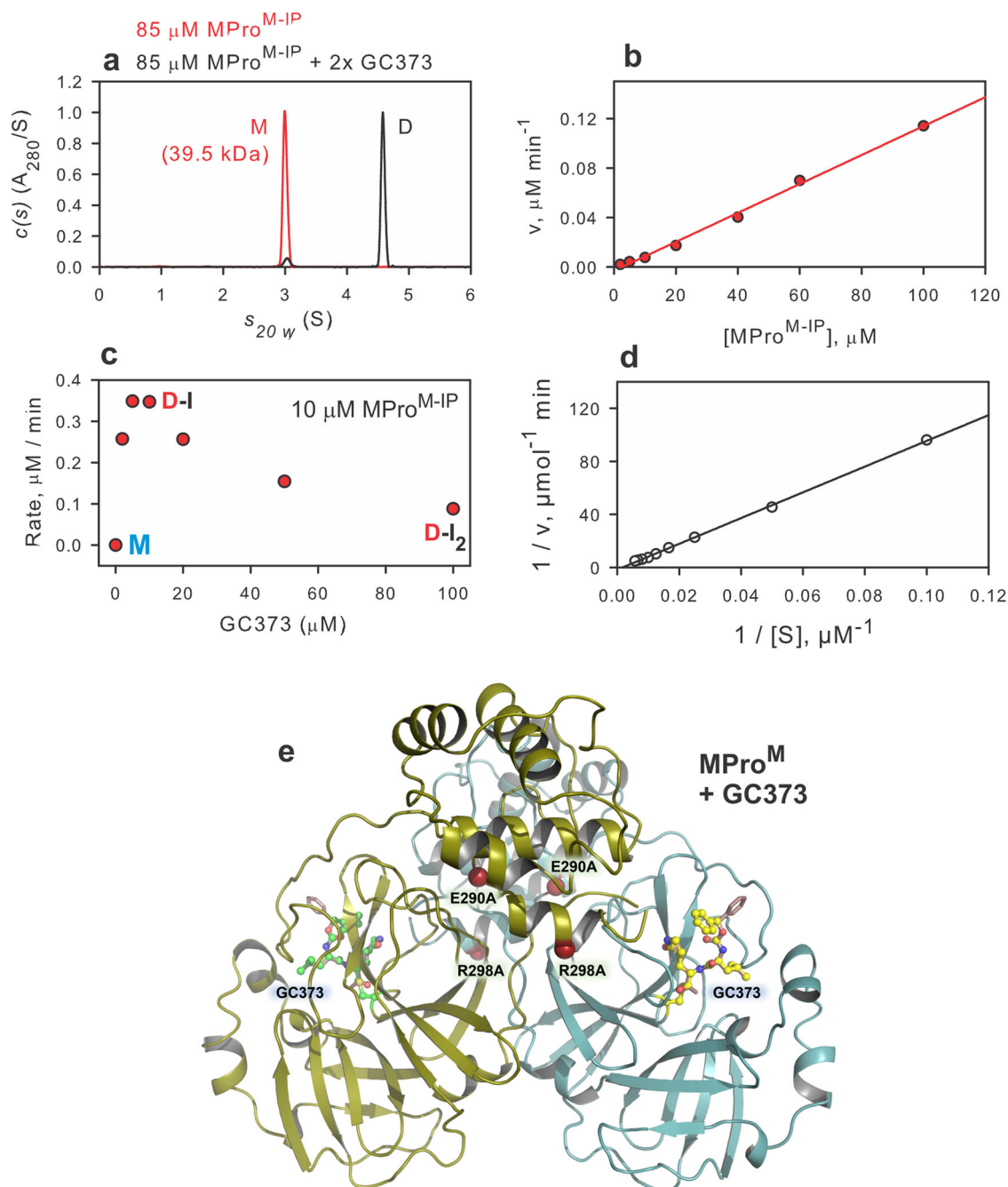
**Fig. 4** Time course of the autoprocessing of  $(-25)\text{MPro}^{\text{M}}$  in the absence and presence of GC373. **a** Autoprocessing monitored by varying the inhibitor GC373 concentration. Time course of autoprocessing in the absence (**b**) and presence (**c**) of GC373. Reactions were initiated at the indicated concentration in buffer B at 28 °C. Aliquots of the reaction were drawn at the indicated times, terminated by the addition of gel sample buffer, subjected (3.5  $\mu\text{g}/\text{lane}$ ) to SDS-PAGE and visualized by staining. Precursor, reaction product, and molecular weight standards (M) are indicated in kDa.

catalyze the hydrolysis of the peptide substrate, and the rate of hydrolysis is linearly dependent on the square of the protein concentration, i.e., second-order kinetics, indicating that the observed catalytic activity is that of a dimer like that of  $\text{MPro}^{\text{WT}}$  and  $\text{MPro}^{\text{M}}$  (Fig. 2c, d). The kinetic parameters  $k_{\text{cat}}/K_{\text{m}}$  of mature  $\text{MPro}^{\text{E290A}}$ ,  $\text{MPro}^{\text{R298A}}$ -catalyzed hydrolysis of substrate as well as the inhibitor GC373 binding constants are listed in Table 2 (Figs. 2e, f, S4a, b, S5c). The monomer-dimer distributions of  $\text{MPro}^{\text{R298A}}$  and  $\text{MPro}^{\text{E290A}}$  in the absence of inhibitor or substrate by SV-AUC are shown in Figure S5a, b.

**Room-temperature structure of GC373 bound  $\text{MPro}^{\text{M}}$ .** Since  $\text{MPro}^{\text{M}}$  is predominantly monomeric in solution, the effect of the mutation of residues E290A and R298A on the overall structure of the protein particularly in the active site region and the monomer-dimer interface regions that include the N- and C-termini is examined by X-ray crystallography. Attempts to grow crystals of monomeric  $\text{MPro}^{\text{M}}$  and  $\text{MPro}^{\text{M-IP}}$  were unsuccessful, although a crystal structure of the monomeric SARS-CoV-1  $\text{MPro}^{\text{R298A}}$  single mutant was attained previously<sup>22</sup>. However, we succeeded in obtaining a room-temperature crystal

structure of  $\text{MPro}^{\text{M}}$  in complex with the reversible covalent inhibitor GC373 at 1.75 Å resolution (Table 3).  $\text{MPro}^{\text{M}}$  crystallizes in the monoclinic unit cell (space group *I*2) with one protomer present in the asymmetric unit.  $\text{MPro}^{\text{M}}$ -GC373 complex is homodimeric, with the quaternary structure virtually identical to that of the  $\text{MPro}^{\text{WT}}$  [Fig. 5e<sup>10,23</sup>]. We modeled residues 1-303 in  $\text{MPro}^{\text{M}}$ -GC373 structure, whereas the rest of the C-terminal residues are disordered and are not visible in the electron density map. GC373 is covalently bonded to the sulfur of the catalytic Cys145 generating the hemithioacetal conjugate. Interestingly, the nucleophilic attack of the sulfur of Cys145 on the carbonyl carbon of GC373 is not stereospecific and occurs from both sides of the carbonyl carbon such that the hemithioacetal hydroxyl group is observed in two alternate orientations (Fig. 6a). In one orientation, the hydroxyl is directed into the oxyanion hole created by residues Leu141-Cys145, whereas in the other it faces the catalytic His41. In addition, the phenyl tail of GC373 occupies two alternate conformations. In the previous room temperature X-ray structure of  $\text{MPro}^{\text{WT}}$ -GC373 complex<sup>11</sup>, the ligand was found in one orientation, bound in a stereospecific fashion with the hemithioacetal hydroxyl directed into the oxyanion hole. GC373 makes five direct hydrogen bonds with  $\text{MPro}^{\text{M}}$  residues (Fig. 6b). The shortest and possibly the strongest hydrogen bond of 2.5 Å is formed between the ligand's P1  $\gamma$ -lactam and imidazole of His163 and is known to be a critical interaction for inhibitors with the  $\text{MPro}$  active site<sup>24-26</sup>. The remaining four hydrogen bond distances are significantly longer, 2.8–3.4 Å. It should be noted that the lack of stereospecificity in the reaction of the sulfur of C145 with the carbonyl carbon of GC373 is not caused by the mutations because (1) the isomeric hemithioacetals have been previously reported in the cryo-structure of  $\text{MPro}^{\text{WT}}$ -GC373 complex<sup>27</sup> and (2) the room temperature structure of  $\text{MPro}^{\text{E290A}}$ -GC373 complex shows only the S-configuration for the hemithioacetal carbon<sup>11</sup>.

The electron density map for the E290A mutation site located at the dimer interface is shown in Fig. 6c where interactions with the N-terminal finger (residues 1–6) occur. Four water molecules (W1–W4) are clearly visible in the vicinity of Ala290, making a tight hydrogen-bonded network with distances between the waters of 2.4–2.6 Å. The water structure intervenes between the stretch of residues 288–291 of protomer A and Arg4 and Lys5 of protomer B. W1 hydrogen bonds with Arg131 (protomer A), and W3 makes a hydrogen bond with Lys5 (protomer B). Thus, the waters mediate the connection between the two protomers. Arg4 of protomer B forms hydrogen bonds with the main chain carbonyl of Glu127 of protomer A. Evidently, there are no direct hydrogen bonds between the protomers in this region. When  $\text{MPro}^{\text{M}}$ -GC373 is superimposed with the room-temperature joint X-ray/neutron structure of  $\text{MPro}^{\text{WT}}$  [PDB ID 7JUN<sup>23</sup>,] or with the room-temperature X-ray structure of  $\text{MPro}^{\text{WT}}$ -GC373 complex [PDB ID 7UKK<sup>11</sup>,], several important structural differences relative to  $\text{MPro}^{\text{M}}$  can be noted (Figs. 6d and S6). In both  $\text{MPro}^{\text{WT}}$  structures, Glu290 of protomer A makes strong 2.6–2.8 Å hydrogen bond interactions with Arg4 of protomer B through a salt bridge structure. In addition, Glu288 of protomer A forms 2.9–3.0 Å hydrogen bonds with Lys5 of protomer B in the inhibitor-free and GC373-bound  $\text{MPro}^{\text{WT}}$  structures. In  $\text{MPro}^{\text{M}}$ -GC373, these hydrogen bonds are severed due to shifts of Glu288 of protomer A and Lys5 of protomer B away from each other, resulting in their separation of 5.5 Å. Moreover, to stabilize its position in the absence of the Glu290 carboxylic group, the Arg4 of the protomer B side chain flips towards the main chain of Glu127. W1 is conserved between inhibitor-free  $\text{MPro}^{\text{WT}}$  and  $\text{MPro}^{\text{M}}$ -GC373 structures and is replaced with the alternate conformation of Glu290 in  $\text{MPro}^{\text{WT}}$ -GC373 complex. W1 makes a 2.6 Å hydrogen bond with Glu290 in  $\text{MPro}^{\text{WT}}$ . Thus, because of the E290A mutation, three water molecules W2, W3, and W4 invade the dimer



**Fig. 5** Characterization of MPro<sup>M-IP</sup> and ribbon representation of room temperature structure of GC373 bound MPro<sup>M</sup> dimer. **a** Normalized sedimentation velocity absorbance  $c(s)$  distributions in the absence (red) and presence (black) of GC373. **b** Linear relationship between the rate of catalyzed hydrolysis vs concentration. **c** A plot of the rate as a function of increasing GC373 mixed with a constant amount of MPro<sup>M-IP</sup>. **d** Lineweaver-Burk plot for the hydrolysis of substrate by 10  $\mu\text{M}$  MPro<sup>M-IP</sup> in the presence of equimolar concentration of GC373. The  $k_{\text{cat}}/K_m$  was calculated as described<sup>13</sup> and listed in Table 2. **e** Room temperature structure of GC373 bound MPro<sup>M</sup> (PDB 8FIG). The positions of mutations E290A and R298A used to create MPro<sup>M</sup> with a very high  $K_{\text{dimer}}$ <sup>13</sup> and GC373 are shown.

interface disrupting several hydrogen bonds between the protomers and contributing to the increase in the  $K_{\text{dimer}}$ .

The electron density map for the MPro<sup>M</sup> residues around the R298A mutation site located where C-terminal residues interact with the other protomer and specifically with Ser139–Leu141 turn at the start of the oxyanion loop is shown in Fig. 6e. There is one direct interprotomer hydrogen bond of 2.8 and 2.7 Å from Gln299 of protomer A to Ser139 of protomer B in inhibitor-free MPro<sup>WT</sup> and MPro<sup>WT</sup>-GC373, respectively. This hydrogen bond is retained in MPro<sup>M</sup>-GC373 compared to MPro<sup>WT</sup> when the

structures are superimposed. In both MPro<sup>WT</sup> structures, Arg298 is situated more than 5 Å from the nearest residue of the other protomer. However, it makes hydrophobic interactions with the C-terminal residues from the same protomer, which are lost when this Arg is mutated to Ala. In both MPro<sup>WT</sup>, the C-terminal residues starting from Cys300 located at the end of a short  $\alpha$ -helix of residues 293–300 make a 90° turn towards the interface between the catalytic domains. In MPro<sup>M</sup>-GC373 instead, apparently caused by the R298A mutation, the C-terminus flips in the opposite direction towards the helical domain so that

**Table 2 Catalytic activity, dimer dissociation and GC373 binding constants of mature MPro and its precursor analogues.**

Construct	$k_{cat}/K_m$ ( $\mu\text{M}^{-1} \text{min}^{-1}$ )	$k_{cat}/K_m K_{dimer}^a$ ( $\mu\text{M}^{-2} \text{min}^{-1}$ )	$K_{dimer}$ ( $\mu\text{M}$ )	GC373 binding ( $K_d$ , $\mu\text{M}$ )	Catalytic/Binding species
<b>Wild-type</b>					
MPro <sup>WT</sup> <sup>b</sup>	0.6 ± 0.05		1.3 ± 0.2	0.15 ± 0.03 <sup>c</sup>	Dimer
<b>Substitution mutants</b>					
MPro <sup>R298A</sup>	0.31 ± 0.01 <sup>d</sup>	0.0435 ± 0.002	7.13 ± 0.40	0.2 ± 0.04 <sup>c</sup>	Dimer
MPro <sup>E290A</sup>	(2.82 ± 0.04) 10 <sup>-3</sup> <sup>d</sup>	(8.0 ± 0.40) 10 <sup>-6</sup>	353 ± 21	3.5 ± 0.5 <sup>c</sup> 4.8 ± 0.6 <sup>e</sup>	Dimer
MPro <sup>M</sup> <sup>b</sup>	(3.4 ± 0.13) 10 <sup>-3</sup>	(5.0 ± 0.15) 10 <sup>-7</sup>	6600	6.13 ± 0.3 <sup>c</sup> 6.7 ± 0.2 <sup>e</sup>	Dimer
<b>Deletion mutant</b>					
MPro <sup>1-199</sup> <sup>b</sup>	(1.15 ± 0.03) 10 <sup>-6</sup>			32 ± 5 <sup>c</sup>	Monomer
<b>Precursor analogues</b>					
-102MPro <sup>M</sup>				19.2 ± 4.4 <sup>c</sup>	Monomer
MPro <sup>M-IP</sup>	(6.04 ± 0.15) 10 <sup>-6</sup> (9.40 ± 0.90) 10 <sup>-4</sup> <sup>d</sup>			6.02 ± 1.2 <sup>c</sup> 8.5 ± 1.5 <sup>e</sup>	Monomer Dimer

<sup>a</sup> $k_{cat}/K_m K_{dimer}$  is the slope of the linear plot of the rate vs the square of the protein concentration.

<sup>b</sup>cited from references <sup>11,13</sup>.

<sup>c</sup>determined by ITC.

<sup>d</sup> $k_{cat}/K_m$  was determined in the presence of the equimolar amount of enzyme and GC373 as described for MPro<sup>M</sup> in Nashed et al.<sup>13</sup>.

<sup>e</sup>determined by SV-AUC.

**Table 3 Crystallographic data collection and refinement statistics.**

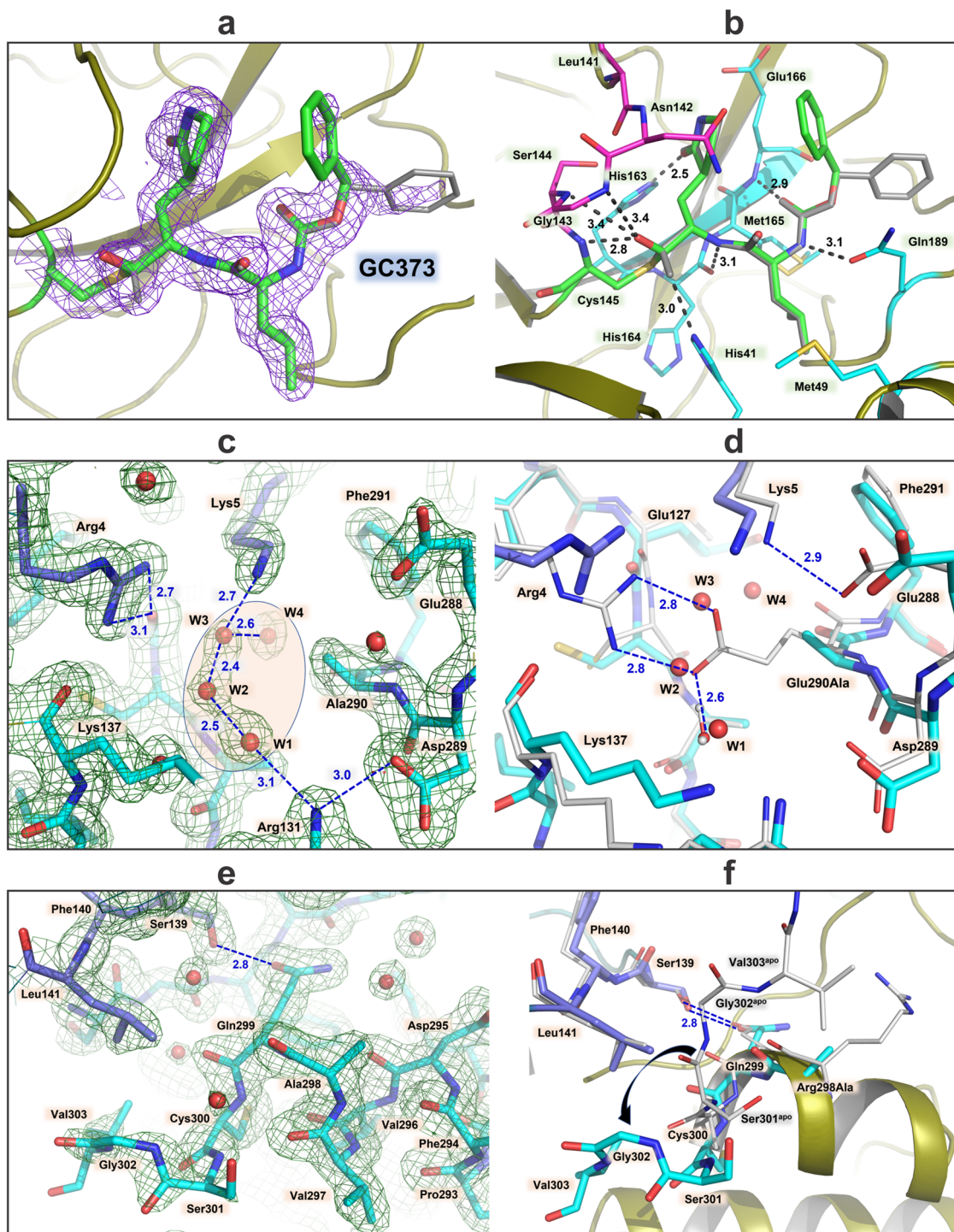
Data collection:	MPro <sup>M</sup> -GC373 PDB ID 8FIG X-ray (in-house)
Diffractometer	Rigaku HighFlux, DECTRIS Eiger R 4 M
Space group	I2
Wavelength (Å)	1.5406
<b>Cell dimensions:</b>	
a, b, c (Å)	52.24, 80.76, 92.52
α, β, γ (°)	90, 95.6, 90
Resolution (Å)	60.71-1.75 (1.81-1.75)
No. reflections unique	36987 (3583)
R <sub>merge</sub>	0.065 (0.635)
CC <sub>1/2</sub>	0.996 (0.580)
<I / σI >	18.7 (1.5)
Completeness (%)	95.9 (92.8)
Redundancy	5.5 (4.7)
<b>Refinement:</b>	
R <sub>work</sub> / R <sub>free</sub>	0.1647 / 0.1847
Resolution (Å)	60.71-1.75
<b>B-factors (Å<sup>2</sup>)</b>	
Protein	37.16
Ligand	31.2
Water	43.34
<b>R.M.S. deviations</b>	
Bond lengths (Å)	0.005
Bond angles (°)	0.804
All atom clash score	1.26

Data reduction and refinement statistics for the room temperature X-ray crystal structure of SARS-CoV-2 MPro<sup>M</sup>-GC373 complex. Values in parentheses are for the highest-resolution shell.

Val303 side chain, the last residue visible in the electron density, is as close as 4 Å to Ser1 side chain of the same protomer (Figs. 6f and S6). Although C-terminal residues do not make hydrogen bonds at the dimer interface, they interact through hydrophobic interactions with the other protomer in MPro<sup>WT</sup>. These interactions are lost in MPro<sup>M</sup>-GC373 due to the drastic conformational change of the C-terminus in the double mutant.

## Discussion

All precursor analogues used in this study autocatalyze the hydrolytic cleavage at the N-terminus (nsp4/nsp5 site) of MPro including the monomeric (-25)MPro<sup>1-199</sup> which lacks the entire helical domain (domain III). *E. coli* expression of constructs comprising MPro<sup>WT</sup> fused to native cleavage sites at the N- and C-termini results in processed mature MPro<sup>WT</sup> without observing any full-length or intermediate products, consistent with previous reports<sup>9,11,28</sup>. Similarly, full-length precursors are not observed upon expression of Precursor<sup>R298A</sup> and (-102)MPro<sup>R298A</sup>. In contrast, full-length precursors accumulate in *E. coli* upon expression of Precursor<sup>E290A</sup>, Precursor<sup>M</sup>, (-102)MPro<sup>E290A</sup>, (-102)MPro<sup>M</sup>, (-25)MPro<sup>M</sup>, and (-25)MPro<sup>199</sup> that undergo time-dependent autoprocessing reaction at the N-terminus (Figs. 1, 2 and S2). We isolated the full-length precursors and the products of the autoprocessing reactions from *E. coli*. The results in Table 2 show the effect of mutations on the kinetic parameter  $k_{cat}/K_m$ , binding constants to the known covalent inhibitor GC373, and the dimer dissociation constants. Mutations E290A and R298A affect both  $k_{cat}/K_m$  and  $K_{dimer}$  to a different extent. MPro<sup>R298A</sup> exhibits a  $k_{cat}/K_m$  of 0.31  $\mu\text{M}^{-1} \text{min}^{-1}$ , i.e., about half that of MPro<sup>WT</sup>, and a  $K_{dimer}$  of 7.13  $\mu\text{M}$ , i.e., about 5.5 times larger than that of MPro<sup>WT</sup>. E290A mutation has a much larger effect on  $k_{cat}/K_m$  and  $K_{dimer}$  compared to the R298A mutation. The observed  $k_{cat}/K_m$  for MPro<sup>E290A</sup> is within the range of the double mutant MPro<sup>M</sup>. The observed  $K_{dimer}$  of 353  $\mu\text{M}$  is about 19 times smaller than that of MPro<sup>M</sup> and about 270 times larger than that of MPro<sup>WT</sup> (Table 2). These results are consistent with the observed accumulation profile of the precursor in *E. coli* and highlight the role of dimer formation in the autoprocessing reaction. Precursor comprising MPro<sup>R298A</sup> does not accumulate in *E. coli* and only the fully processed mature enzyme upon cleavages at both termini is observed. While the intrinsic catalytic activity ( $k_{cat}/K_m$ ) of MPro<sup>E290A</sup> is nearly identical to that of MPro<sup>M</sup>, the precursor comprising MPro<sup>E290A</sup> undergoes stepwise autoprocessing at both termini at a much faster rate because of its smaller  $K_{dimer}$ . In contrast, precursors comprising MPro<sup>M</sup> and MPro<sup>1-199</sup> accumulate and undergo autoprocessing in the monomeric form at a much slower rate in cells which allows their purification for in vitro reactions. It is intriguing, however, that the monomeric precursor comprising MPro<sup>M</sup> with only the nsp5/nsp6 cleavage



**Fig. 6** Binding of GC373 to MPro<sup>M</sup> and the effects of mutations on the dimer interface. **a** 2F<sub>o</sub>-F<sub>c</sub> electron density of GC373 covalently bonded to C145 is shown at 1.5  $\sigma$  level. GC373 binds in two alternate conformations shown with green carbon atoms and in gray with the refined occupancies of 51/49%, respectively. **b** Hydrogen bonds formed by GC373 with the MPro<sup>M</sup> active site residues. Active site and oxyanion hole residues are colored by atom type with cyan and magenta carbon atoms, respectively. **c** 2F<sub>o</sub>-F<sub>c</sub> electron density for the interprotomer interface near E290A mutation site shown at 1.5  $\sigma$  level. Protomer A carbon atoms are colored cyan, whereas protomer B carbon atoms are colored slate. **d** Superposition of MPro<sup>M</sup> (cyan and slate carbon atoms) and MPro<sup>WT</sup> (gray carbon atoms) near E290A mutation site demonstrating the loss of the E290...R4' salt bridge. **e** 2F<sub>o</sub>-F<sub>c</sub> electron density for the interprotomer interface near R298A mutation site is shown at 1.5  $\sigma$  level. **f** Superposition of MPro<sup>M</sup> and MPro<sup>WT</sup> near R298A mutation site showing a conformational reorientation of the C-terminal residues (black curved arrow). Atoms are colored in the same way as in panel **d**. All distances are shown in Angstroms.

site, even in its dimer form, failed to promote cleavage at the C-terminus of MPro<sup>M</sup> despite the accessibility for cleavage of this site by provided MPro<sup>WT</sup>.

Similarly, E290A and R298A mutations affect the binding of the transition state analog GC373 to varying extent. While the

binding constant of GC373 to MPro<sup>R298A</sup> is within the experimental error to that of MPro<sup>WT</sup>, MPro<sup>E290A</sup> is 23 times larger relative to MPro<sup>WT</sup> and comparable to that of MPro<sup>M</sup>. Furthermore, the mutant enzymes bind the aldehyde inhibitor GC373 with a binding constant inversely proportional to  $k_{cat}/K_m$

which is consistent with previously reported studies [see Table 2 and ref. 11,13].

Room temperature crystal structure of MPro<sup>M</sup> shows that E290A and R298A mutations destabilize the dimer by excluding all direct hydrogen bonding between the protomers in the vicinity of residue 290 and retaining one hydrogen bond near residue 298. In the presence of the inhibitor, the dimer is stabilized through hydrophobic interactions, bridged by hydrogen bonding through water molecules and a H-bond between Gln299 of protomer A with Ser139 of protomer B. Destabilization of the dimer form is reflected in the catalytic activity, the  $K_{\text{dimer}}$ , binding of the covalent inhibitor GC373 (Table 2) as well as the accumulation of the full-length precursor upon expression in *E. coli*. Both mature MPro mutants, MPro<sup>E290A</sup> and MPro<sup>R298A</sup>, follow second-order kinetics in protein concentration for catalyzing the hydrolysis of the peptide substrate (Fig. 2c, d) indicating that the observed catalytic activity is that of a dimer, like that of MPro<sup>WT</sup> and MPro<sup>M</sup> and in contrast to that of the monomeric catalytic activity of MPro<sup>1-199</sup><sup>11</sup>. Recently, we reported the room temperature X-ray structures of MPro<sup>1-199</sup> and MPro<sup>1-196</sup>-GC373 complexes<sup>11</sup>. The structure of the monomeric MPro<sup>1-199</sup> is nearly identical to that of the dimeric wild-type except for the oxyanion-loop being in an unwound conformation (inactive E-state), whereas MPro<sup>1-196</sup>-GC373 covalent complex displays a wound oxyanion loop (active E\*-state) typical of the dimeric form of inhibitor-free mature MPro. The observed catalytic activities of MPro<sup>M</sup> and MPro<sup>1-199</sup> indicated that the E-E\* equilibrium is dynamic in which the E state predominates in the monomer while the E\* is the catalytically active species and favored in the dimer<sup>11-13</sup>.

Although we were unable to isolate Precursor<sup>E290A</sup> in sufficient purity and quantity for in vitro analysis due to the relatively fast autoprocessing reaction, small-scale isolation of the full-length precursor and its products enabled verification by mass spectrometry. Precursor<sup>E290A</sup> exhibits time-dependent autoprocessing *via* initial cleavage at the N-terminus to produce MPro<sup>E290A-IP</sup> followed by the cleavage at the C-terminus to produce mature MPro<sup>E290A</sup> (Fig. 1e-g). This order of cleavage is consistent with the observed cleavage order of a construct comprising 10 amino acid of the native flanking sequence at the N- and C-termini of MPro of SARS-CoV having an active site C145A mutation catalyzed intermolecularly by mature MPro<sup>WT</sup><sup>29</sup>.

As indicated above, <sup>(-102)</sup>MPro<sup>M</sup> and <sup>(-25)</sup>MPro<sup>199</sup> undergo time-dependent autoprocessing at the N-terminus to produce MPro<sup>M</sup> and MPro<sup>199</sup> in vitro (Fig. 3). The results in Table 1 show that the percent conversion of the precursor to products at a given time is independent of the initial protein concentration indicating that the reactions follow first-order kinetics, i.e., unimolecular. The percent conversion of the two precursors at a given time are similar, if not identical. Thus, the cleavage at the N-terminus of MPro occurs *via* an intramolecular mechanism from a monomer in these two model precursors. Importantly, the addition of GC373 to the precursor analogue, <sup>(-25)</sup>MPro<sup>M</sup>, increases the rate of autoprocessing and the increase in rate of autoprocessing is dependent on the GC373 concentration. This attribute is similar to that reported for the activation of MPro<sup>M</sup> by GC373 and can be explained by a similar mechanism<sup>13</sup>. In that mechanism, MPro<sup>M</sup> is predominantly in the monomer form with an oxyanion loop in the E conformation. Upon binding GC373, MPro<sup>M</sup> forms a dimer concomitant with the reorganization of the oxyanion loop to the active E\* conformation. In the presence of the equimolar amount of protein and GC373, three major protein species are present in equilibrium: monomer, dimer with one of the active sites occupied by GC373, and a dimer with both active sites occupied by GC373. The increase in catalytic activity is attributed to the presence of the population of a dimer with one

active site occupied by GC373, leaving the other active site with an oxyanion loop in the active E\* conformation for catalytic function. Thus, the increase in the reaction rate of <sup>(-25)</sup>MPro<sup>M</sup> autoprocessing in the presence GC373 must be due to the formation of a dimer. Accordingly, the autoprocessing of the model precursor is initiated by N-terminal intramolecular cleavage from a monomer or a dimer. The enhanced dimer stability in the context of a wild-type precursor, relative to MPro<sup>M</sup>, suggests that the initial cleavage at the N-terminus most likely occurs from the dimer because it is more reactive than the monomer. This conclusion is in accordance with the precursor accumulation profile in *E. coli* described above.

N-terminal cleavage of <sup>(-102)</sup>MPro<sup>M</sup> to produce MPro<sup>M</sup> is accompanied by an increase in catalytic activity as well as the binding affinity to GC373 increasing by about 3-fold (Table 2, Fig. S4c). As previously reported, the observed catalytic activity of MPro<sup>M</sup> is that of a dimer indicating that N-terminal cleavage will favor dimer formation. The involvement of the N-finger in dimer formation has been examined in detail by structural and mutational studies. Deletion and mutational studies of the N-finger residues 1-9 show drastic effect on dimer formation and catalytic activity, relative to MPro<sup>WT</sup>, consistent with the critical interfaces involving the N-finger with domains II and III<sup>8,11</sup>. In recent studies, the effect of flanking nsp4 residues, contrary to deleting the N-terminal residues of MPro, on  $K_{\text{dimer}}$  was also examined<sup>12</sup>. The active site H41A mutation to restrict N-terminal autoprocessing enabled the analyses of such precursor mimetics. Even the addition of six residues of the native nsp4 sequence to mature MPro<sup>H41A</sup> (<sup>(-6)</sup>MPro<sup>H41A</sup>) increases the  $K_{\text{dimer}}$  by >85-fold. Structural studies revealed asymmetry, semi-open conformation and disorder of the terminal residues of the dimeric <sup>(-6)</sup>MPro<sup>H41A</sup><sup>12</sup>. Thus, the product of the autoprocessing reaction to liberate the free N-terminus of MPro is critical for dimer formation.

The product MPro<sup>M-IP</sup> resulting from the autoprocessing of Precursor<sup>M</sup> containing the nsp5/nsp6 cleavage site at the C-terminus of MPro<sup>M</sup> exhibits catalytic activity, which is that of a monomer and is activated and inhibited by GC373 like that of mature MPro<sup>M</sup>. Activation by GC373 is concomitant with dimer formation. However, it does not autocatalyze the nsp5/nsp6 junction indicating that the monomer is incapable of this cleavage either in the absence, or presence of GC373 in its dimer form. But treatment of the monomeric MPro<sup>M-IP</sup> with MPro<sup>WT</sup> produces MPro<sup>M</sup> indicative of the site being accessible for cleavage in its monomeric form. Failure of MPro<sup>M-IP</sup> to process its own C-terminus could be due to its inability to form a dimer. The requirement of a dimer for C-terminal cleavage is consistent with the observations in crystal structures showing the C-terminal residues of one dimer of MPro<sup>C145A</sup> bound to the active site of a second dimer<sup>30-32</sup>. It is unclear as to why MPro<sup>M-IP</sup> fails to undergo C-terminal cleavage in the presence of equimolar amount of GC373 when both monomer and the enzymatically active dimer coexist. This may point to the dynamics of this system which is incapable of forming a productive reaction complex. It is worth noting, however, that lack of C-terminal autoprocessing is limited only to MPro<sup>M-IP</sup> having both mutations E290A and R298A, whereas MPro<sup>E290A-IP</sup> exhibits C-terminal cleavage likely due to its ability to form a significant population of dimer even in the absence of inhibitor GC373.

Transient accumulation of Precursor<sup>E290A</sup> and its conversion to MPro<sup>E290A-IP</sup> and mature MPro<sup>E290A</sup> indicates that cleavage at the N-terminus precedes the C-terminal cleavage. A proposed mechanism for MPro autoprocessing from a model precursor is shown in Fig. 7. In this mechanism, the MPro region of the monomeric precursor exhibits a tertiary fold like that of the

mature subunit with an oxyanion loop predominantly in the E-state. This inactive state is in equilibrium with a minor population having the active E\* conformation, which undergoes slow unimolecular N-terminal autoprocessing. The monomeric precursor is also in equilibrium with its homodimer. This is supported by the observation that the presence of N- and C-terminal flanking sequences increase the  $K_{dimer}$  significantly but do not exclude the precursor from becoming a dimer. In the transient precursor dimer, oxyanion loop dynamics favor an E\* conformation leading to a faster rate of cleavage of the N-terminal site than that of the monomeric precursor. Results presented here show that the monomer and dimer precursor forms can autocatalyze N-terminal cleavage and that the dimer undergoes significantly faster reaction, compared to the monomer. The observed first-order kinetics of the monomeric model precursors indicate that the cleavage at the N-terminus proceeds *via* an intramolecular mechanism. N-terminal cleavage of a monomer or a dimer would lead to an increase in the dimer population as the newly formed free N-finger adopts a native-like interface with the second protomer. A heterodimer comprising a free N-terminus is likely to be more stable than the precursor homodimer and hence, being more reactive leads to an even faster rate of cleavage of the second N-terminus resulting in further stabilization of the dimer interface and decrease in  $K_{dimer}$ . It is likely that the precursor protomer containing the N-terminal flanking sequence is responsible for the cleavage of its own N-terminus because it conforms with the correct orientation for substrate binding visualized in structures<sup>33,34</sup>. Since the N-terminal flanking sequence has a larger effect on the  $K_{dimer}$  as well as the GC373 binding constant (Table 2 and Fig. S4c) than the C-terminal flanking sequence<sup>29</sup>, upon N-terminal cleavage, a more stable

population of homo- and heterodimers favoring the E\* conformation of the oxyanion loop leads to the processing of the C-terminal nsp5/nsp6 site *via* an intermolecular mechanism to release the mature enzyme.

The above mechanism of activation is consistent with reports on the self-cleavage of picornaviral replicase precursors<sup>35,36</sup> and early cleavages of 3CL protease of coronavirus MHV-A59<sup>37</sup>. Both nsp4 and nsp6 are shown to have multiple transmembrane domains and thus, MPro is suggested to be anchored on both sides to membranes. The resulting major products nsp5-nsp10 and nsp5-nsp16 generated from pp1a and pp1ab, respectively, by cleavage at the N-terminus may be transported through membrane association<sup>17</sup> for processing of the nsp5/nsp6 site through a concentration-dependent intermolecular mechanism at a later stage of viral polyprotein maturation.

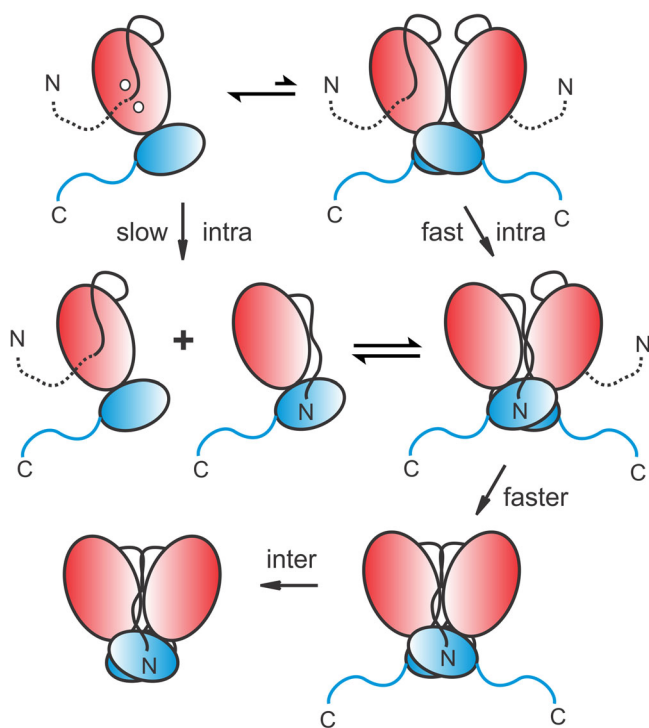
Also, the proposed mechanism of MPro autoprocessing is strikingly similar to that of the dimeric aspartic acid protease of HIV-1 from its precursor<sup>1,3,38</sup>. The main difference lies in the cleavage at the N-terminus of HIV protease occurring from the dimer form because each of the monomer possesses only half of the catalytic residues of the active site. Thus, it appears that intramolecular cleavage at the N-terminus followed by intermolecular cleavage at the C-terminus is a general mechanism for autoprocessing of proteases of RNA viruses from their polyprotein precursor.

## Methods

**Construction and design of MPro constructs.** Expression and purification of MPro<sup>WT</sup> (GenBank ID: MN908947.3), MPro<sup>M</sup> and MPro<sup>1-199</sup> were described before<sup>11,13,39</sup>. New constructs prepared for this work are: Precursor<sup>WT</sup>, Precursor<sup>M</sup> and its single mutants, precursor<sup>(-102)</sup>MPro<sup>WT</sup>, precursor<sup>(-102)</sup>MPro<sup>M</sup> and its single mutants, precursor<sup>(-25)</sup>MPro<sup>1-199</sup>, mature MPro<sup>R298A</sup> and mature MPro<sup>E290A</sup>. All constructs were synthesized and cloned into pJ414 vector (ATUM, Newark, CA). Designations and amino acid sequences of the constructs are listed in Figure 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 at 37 °C. Proteins were purified from the cell lysate by nickel-affinity chromatography (NAC, step 1). The bound fraction was subjected to isocratic fractionation on Superose-12 column (step 2, Cytiva Life Sciences) in a final buffer of 25 mM Tris-HCl, pH 7 or 7.2, 150 mM NaCl and 1 mM TCEP (buffer A). Peak fractions were pooled and concentrated to the desired concentration using Amicon Ultra – 15 or 0.5 ml centrifugal filters (Merck Millipore Ltd.) and stored in aliquots at –30 °C and for long-term storage at –80 °C. Purity was verified both by SDS-PAGE on 4–20% gradient mini-protean TGX pre-cast gel (Bio-Rad) and reverse-phase liquid chromatography with in-line electrospray ionization mass spectrometry<sup>11</sup>. Protein concentrations were measured before storage and prior to the experiment at least in duplicate based on the extinction coefficient (Fig. S1) at 280 nm.

Purifications of mature MPro<sup>R298A</sup> and MPro<sup>E290A</sup> were carried out with an additional HRV-3C protease cleavage step to remove the C-terminal 6His-tag as described previously for MPro<sup>WT</sup> and MPro<sup>M</sup><sup>11</sup>. To isolate precursors<sup>(-102)</sup>MPro<sup>M</sup> and<sup>(-25)</sup>MPro<sup>1-199</sup>, cell cultures (<100 ml) were induced for <30 min and chilled immediately prior to harvesting. Cell pellets were subjected to the same purification scheme as above not exceeding a total time of 4 h at 4 °C due to autoprocessing at native conditions. The final column step on Superose-12 was carried out in buffer B (25 mM Tris-HCl, pH 7, 50 mM NaCl and 1 mM TCEP). The full-length



**Fig. 7 Proposed mechanism of MPro autoprocessing.** Catalytic (domains I and II) and helical (domain III) regions are shown as red and blue ovals, respectively. Solid black lines denote the N-terminal residues (N-finger) of MPro. Dashed black and solid blue lines represent the truncated nsp4 and nsp6 regions flanking the N- and C-termini of MPro. Catalytic dyad H41 and C145 residues are indicated as white circles in the top left monomer cartoon.

protein, representing only a small fraction of the total load on the column, with minimal contamination of the products were pooled and stored in aliquots at  $-70^{\circ}\text{C}$ .

**Autoprocessing.** Time course of the autoprocessing reaction was carried out as follows. Cells (12 ml) were harvested at the indicated time points, chilled on ice, and subjected to NAC on spin columns. Equal volumes of the bound fraction were analyzed by SDS-PAGE. But for side-by-side comparison of the flow through (FT) and bound (B) fractions and monitoring the autoprocessing reactions in vitro (Tables S1, S2), equal amounts of proteins<sup>11</sup> were used for gel fractionation. Molecular weight markers (in kDa) are shown for all gel panels except in few panels when a protein of similar mass (Figs. 2 and S1) or protein from the same stock solution are analyzed (Fig. 4). Uncropped gel images of Figs. 1–4 and S2 are shown in Fig. S7.

**Enzyme kinetics.** Activity assays using the FRET substrate Dabcyl-KTSAVLQ/SGFRKM-E(Edans)-NH<sub>2</sub>, where (I) denotes the scissile peptide bond, were performed in a total volume of 100  $\mu\text{l}$  in buffer B (25 mM Tris-HCl, pH 7, 50 mM NaCl and 1 mM TCEP) at  $28^{\circ}\text{C}$  as previously described<sup>9,13</sup>. Assays were carried out with equimolar amount of GC373 and enzyme and varying the substrate concentration from 5–200  $\mu\text{M}$ .  $k_{\text{cat}}/K_{\text{m}}$  was calculated from plots of  $1/v$  vs  $1/S$  and the GC373 binding constant was determined by ITC. Plots of rate vs the square of the protein concentration at 200  $\mu\text{M}$  were linear, and the slope of the line is  $2 k_{\text{cat}}[S]/K_{\text{dimer}}(K_{\text{m}} + [S])$ . Since the  $K_{\text{m}}$  value is about the same as the substrate concentration used, the slope of the line is simplified to  $k_{\text{cat}}/K_{\text{m}}K_{\text{dimer}}$  that allowed the calculation of  $K_{\text{dimer}}$ . For details, see reference<sup>13</sup>. The substrate was custom synthesized (Biomatik, Ontario, Canada), and GC376 was purchased from Selleckchem, Houston, TX.

**Sedimentation velocity analytical ultracentrifugation (SV-AUC) and Isothermal titration calorimetry (ITC).** SV-AUC (in buffer B at  $25^{\circ}\text{C}$ ) and ITC in buffer C (25 mM Tris-HCl, pH 7.2, 20 mM NaCl and 1 mM TCEP) at  $28^{\circ}\text{C}$  were carried out as described in references<sup>11–13</sup>.

**Protein crystallization and room-temperature X-ray crystallography.** MPro<sup>M</sup> protein sample was concentrated to 7 mg/ml. GC376 stock was prepared at 10 mM GC376 in 25 mM Tris-HCl, pH 7.2, 20 mM NaCl and 1 mM TCEP for crystallization purposes and stored at  $-30^{\circ}\text{C}$ . GC376 is converted to the reactive aldehyde GC373 when mixed with an aqueous solution. For co-crystallization, MPro<sup>M</sup> was mixed with GC376 at 1:5 molar ratio and allowed to incubate at room temperature for at least 30 min before setting up crystal trays. Crystals of GC373-bound MPro<sup>M</sup> were grown at  $14^{\circ}\text{C}$  by sitting drop vapor diffusion methodology 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  $\mu\text{L}$  at 1:1 ratio were seed struck using the crystals of the native MPro in complex with a covalent ligand NBH2 as described in references<sup>26,39</sup>. The crystals suitable for X-ray diffraction measurements were mounted in MiTeGen (Ithaca, NY) room-temperature capillary setups for data collection.

Room temperature X-ray crystallographic data were collected on a Rigaku HighFlux HomeLab instrument equipped with a MicroMax-007 HF X-ray generator, Osmic VariMax optics, and a DECTRIS Eiger R 4 M hybrid photon counting detector. X-ray diffraction data were integrated using the CrysAlis Pro software suite (Rigaku Inc., The Woodlands, TX), then reduced and scaled using Aimless<sup>40</sup> from the CCP4 suite<sup>41</sup>. Structures were solved by molecular replacement using Phaser<sup>42</sup>. MPro<sup>WT</sup>-GC373 complex

structure (PDB code 7UKK)<sup>11</sup>, was used as a search model to solve the structure of MPro<sup>M</sup>-GC373 complex. The model was iteratively refined with *phenix.refine* from the PHENIX suite<sup>43</sup> and COOT<sup>44</sup>. Geometry validation was aided by Molprobrity<sup>45</sup>. GC373 restraints were generated with eLBOW<sup>46</sup> using geometry optimized by quantum chemical calculations in Gaussian16 at B3LYP/6-31 g(d,p) level of theory<sup>47</sup>. Final data collection and refinement statistics can be found in Table 3.

**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 substrate 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 ( $K_{\text{dimer}}$ ) and the binding constant of the inhibitor to the enzyme ( $K_{\text{d}}$ ), respectively.  $K_{\text{dimer}}$  and  $K_{\text{d}}$  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 constants were determined by Lamm equation 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.

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

#### Data availability

The structure and corresponding structure factors have been deposited into the protein data bank with the PDB accession code 8FI9 for MPro<sup>M</sup>/GC373 complex. Source data files are provided in Supplementary Data 1, 2, and Supplementary information. All other data that support this study are available from the corresponding authors upon reasonable request.

Received: 11 July 2023; Accepted: 16 October 2023;

Published online: 13 November 2023

#### References

1. Louis, J. M., Weber, I. T., Tozser, J., Clore, G. M. & Gronenborn, A. M. HIV-1 protease: maturation, enzyme specificity, and drug resistance. *Adv. Pharm.* **49**, 111–146 (2000).
2. Groneberg, D. A., Hilgenfeld, R. & Zabel, P. Molecular mechanisms of severe acute respiratory syndrome (SARS). *Respir. Res.* **6**, 8 (2005).
3. Louis, J. M., Ishima, R., Torchia, D. A. & Weber, I. T. HIV-1 protease: structure, dynamics, and inhibition. *Adv. Pharm.* **55**, 261–298 (2007).
4. Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M. & Bergeron, J. R. C. Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. *Front. Mol. Biosci.* **7**, 605236 (2020).
5. Wang, C. et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* **92**, 667–674 (2020).
6. Yoshimoto, F. K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *Protein J.* **39**, 198–216 (2020).
7. Xia, B. & Kang, X. Activation and maturation of SARS-CoV main protease. *Protein Cell* **2**, 282–290 (2011).
8. Goyal, B. & Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **22**, 297–305 (2020).
9. 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).

10. Kneller, D. W. et al. Structural plasticity of SARS-CoV-2 3CL M(pro) active site cavity revealed by room temperature X-ray crystallography. *Nat. Commun.* **11**, 3202 (2020).
11. Nashed, N. T. et al. Autoprocessing and oxyanion loop reorganization upon GC373 and nirmatrelvir binding of monomeric SARS-CoV-2 main protease catalytic domain. *Commun. Biol.* **5**, 976 (2022).
12. Kovalevsky, A. et al. Unmasking the Conformational Stability and Inhibitor Binding to SARS-CoV-2 Main Protease Active Site Mutants and Miniprecursor. *J. Mol. Biol.* **434**, 167876 (2022).
13. Nashed, N. T., Aniana, A., Ghirlando, R., Chiliveri, S. C. & Louis, J. M. Modulation of the monomer-dimer equilibrium and catalytic activity of SARS-CoV-2 main protease by a transition-state analog inhibitor. *Commun. Biol.* **5**, 160 (2022).
14. Cheng, S. C., Chang, G. G. & Chou, C. Y. Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. *Biophys. J.* **98**, 1327–1336 (2010).
15. Oostra, M. et al. Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J. Virol.* **81**, 12323–12336 (2007).
16. Oostra, M. et al. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* **82**, 12392–12405 (2008).
17. Ricciardi, S. et al. The role of NSP6 in the biogenesis of the SARS-CoV-2 replication organelle. *Nature* **606**, 761–768 (2022).
18. Rong, Y., Zhang, C., Gao, W. C. & Zhao, C. Optimization of the expression of the main protease from SARS-CoV-2. *Protein Expr. Purif.* **203**, 106208 (2023).
19. Chen, S., Jonas, F., Shen, C. & Hilgenfeld, R. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. *Protein Cell* **1**, 59–74 (2010).
20. Gronenborn, A. M. et al. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657–661 (1991).
21. Lin, P. Y., Chou, C. Y., Chang, H. C., Hsu, W. C. & Chang, G. G. Correlation between dissociation and catalysis of SARS-CoV main protease. *Arch. Biochem. Biophys.* **472**, 34–42 (2008).
22. Shi, J., Sivaraman, 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).
23. Kneller, D. W. et al. Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography. *J. Biol. Chem.* **295**, 17365–17373 (2020).
24. Owen D. R. et al. An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science* **274**, 1586–1593 (2021).
25. Qiao, J. et al. SARS-CoV-2 M(pro) inhibitors with antiviral activity in a transgenic mouse model. *Science* **371**, 1374–1378 (2021).
26. Kneller, D. W. et al. Covalent nirmatrelvir- and boceprevir-derived hybrid inhibitors of SARS-CoV-2 main protease. *Nat. Commun.* **13**, 2268 (2022).
27. Ma, C. et al. Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. *Cell Res.* **30**, 678–692 (2020).
28. Xue, X. et al. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.* **366**, 965–975 (2007).
29. Hsu, M. F. et al. Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* **280**, 31257–31266 (2005).
30. Noske, G. D. et al. A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process. *J. Mol. Biol.* **433**, 167118 (2021).
31. Muramatsu, T. et al. SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity. *Proc. Natl Acad. Sci. USA* **113**, 12997–13002 (2016).
32. Lee, J. et al. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nat. Commun.* **11**, 5877 (2020).
33. 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. *IUCr* **8**, 973–979 (2021).
34. MacDonald, E. A. et al. Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **7**, 2591–2595 (2021).
35. Palmenberg, A. C. & Rueckert, R. R. Evidence for intramolecular self-cleavage of picornaviral replicase precursors. *J. Virol.* **41**, 244–249 (1982).
36. Blair, W. S. & Semler, B. L. Self-cleaving proteases. *Curr. Opin. Cell Biol.* **3**, 1039–1045 (1991).
37. Lu, X., Lu, Y. & Denison, M. R. Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. *Virology* **222**, 375–382 (1996).
38. Louis, J. M., Nashed, N. T., Parriss, K. D., Kimmel, A. R. & Jerina, D. M. Kinetics and mechanism of autoprocessing of the Gag-Pol polyprotein. *Proc. Natl Acad. Sci. USA* **91**, 7970–7974 (1994).
39. Kneller, D. W. et al. Structural, Electronic, and Electrostatic Determinants for Inhibitor Binding to Subsites S1 and S2 in SARS-CoV-2 Main Protease. *J. Med. Chem.* (2021).
40. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr D. Biol. Crystallogr* **69**, 1204–1214 (2013).
41. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D. Biol. Crystallogr* **67**, 235–242 (2011).
42. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr* **40**, 658–674 (2007).
43. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D. Struct. Biol.* **75**, 861–877 (2019).
44. Casanal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* **29**, 1069–1078 (2020).
45. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D. Biol. Crystallogr* **66**, 12–21 (2010).
46. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D. Biol. Crystallogr* **65**, 1074–1080 (2009).
47. Frisch M. J. et al. Gaussian 16, Revision B.01. Gaussian Inc., Wallingford, CT. (2016).

## Acknowledgements

This research used resources at the Spallation Neutron Source, and the High Flux Isotope Reactor, which are DOE Office of Science User Facilities operated by the Oak Ridge National Laboratory. The Office of Biological and Environmental Research supported research at ORNL's Center for Structural Molecular Biology (CSMB), a DOE Office of Science User Facility. ORNL is managed by UT-Battelle LLC for DOE's Office of Science, the single largest supporter of basic research in the physical sciences in the United States. We thank John Lloyd and the NIDDK mass spectrometry core facility. This work was supported by the Intramural Research Program of National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), NIH.

## Author contributions

N.T.N., A.K. and J.M.L. designed the study; N.T.N., A.A. and J.M.L. carried out enzyme 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. performed SV-AUC and analyzed data; J.M.L. performed ITC and analyzed data; N.T.N., A.K. and J.M.L. wrote the paper with help from all co-authors.

## Funding

Open Access funding provided by the National Institutes of Health (NIH).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-023-05469-8>.

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

**Peer review information** *Communications Biology* thanks Yu Chen and Iñaki Tuñón for their contribution to the peer review of this work. Primary Handling Editors: Min Zhuang, Gene Chong, and David Favero.

**Reprints and permission information** is 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, 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 2023