A Crystallographic Snapshot of SARS-CoV-2 Main Protease Maturation Process

G. D. Noske [1‡], **A. M. Nakamura** [1‡], **V. O. Gawriljuk** [1], **R. S. Fernandes** [1],
G. M. A. Lima [2†], **H. V. D. Rosa** [1], **H. D. Pereira** [1], **A. C. M. Zeri** [3], **A. F. Z. Nascimento** [3],
M. C. L. C. Freire [1], **D. Fearon** [4,5], **A. Douangamath** [4,5], **F. von Delft** [4,5,6,7],
G. Oliva [1*] **and A. S. Godoy** [1*]

1 - *Institute of Physics of Sao Carlos,* University of Sao Paulo, Av. Joao Dagnone, 1100, Jardim Santa Angelina, Sao Carlos 13563-120, Brazil

2 - *BioMAX,* MAX IV Laboratory, Fotongatan 2, Lund 224 84, Sweden

3 - *Brazilian Synchrotron Light Laboratory (LNLS),* Brazilian Center for Research in Energy and Materials (CNPEM), Zip Code 13083-970 Campinas, Sao Paulo, Brazil

4 - *Diamond Light Source Ltd.,* Harwell Science and Innovation Campus, Didcot OX11 0QX, UK

5 - *Research Complex at Harwell,* Harwell Science and Innovation Campus, Didcot OX11 0FA, UK

6 - *Centre for Medicines Discovery,* University of Oxford, Old Road Campus, Roosevelt Drive, Headington OX3 7DQ, UK

7 - *Department of Biochemistry,* University of Johannesburg, Auckland Park 2006, South Africa

Abstract

SARS-CoV-2 is the causative agent of COVID-19. The dimeric form of the viral M^{pro} is responsible for the cleavage of the viral polyprotein in 11 sites, including its own N and C-terminus. The lack of structural information for intermediary forms of M^{pro} is a setback for the understanding its self-maturation process. Herein, we used X-ray crystallography combined with biochemical data to characterize multiple forms of SARS-CoV-2 M^{pro}. For the immature form, we show that extra N-terminal residues caused conformational changes in the positioning of domain-three over the active site, hampering the dimerization and diminishing its activity. We propose that this form preludes the cis and trans-cleavage of N-terminal residues. Using fragment screening, we probe new cavities in this form which can be used to guide therapeutic development. Furthermore, we characterized a serine site-directed mutant of the M^{pro} bound to its endogenous N and C-terminal residues during dimeric association stage of the maturation process. We suggest this form is a transitional state during the C-terminal trans-cleavage. This data sheds light in the structural modifications of the SARS-CoV-2 main protease during its self-maturation process.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19, a highly infectious disease that rapidly spreads causing a global pandemic. SARS-CoV-2 is an enveloped RNA virus belonging to the β-lineage of coronaviruses, which includes SARS-CoV and Middle East (MERS-CoV) respiratory viruses.[1–3] The viral genome is a single-stranded positive RNA comprising about 30,000 nucleotides, that shares 82% sequence identity with SARS-CoV.[4] The replicase gene (ORF1ab) encodes two over-

lapping polyproteins (pp1a and pp1ab) that are required for viral replication and transcription.[5]

The main protease (M[pro]), also known as 3C-like protease (3CL[pro]) is a viral cysteine protease specific for glutamine at the S1 subsite, showing variable recognition preferences at S2 (Leu/Phe/Met/Val) and S2′ subsites (Ser/Ala/Gly/Asn).[6] M[pro] is responsible for the maturation of pp1a and pp1ab in at least 11 characterized sites, including its auto-processing at the N and C terminus, which is essential for its activity and dimerization.[4,7,8] Due to its essential role in viral replication, M[pro] is one of the most well characterized non-structural proteins of SARS-CoV-2. In addition, its unique features of cleavage site recognition and the absence of closely related homologues in humans, identify M[pro] as a major target for antiviral drug development.[4,9,10]

Although M[pro] activity is crucial to viral biology, its self-maturation process is still poorly understood. Several biochemical and crystallographic studies on native and mutated forms of SARS-CoV M[pro] tried to elucidate its maturation mechanism,[11] by evaluating if the N and C-terminus processing occurs within a dimer (*cis*-cleavage) or between two distinct dimers (*trans*-cleavage). The first 2005 model suggested that M[pro] probably forms a small amount of active dimer after autocleavage that immediately enables the catalytic site to act on other cleavage sites in the polyprotein.[12] In 2010, based on the observation that dimerization of mature M[pro] is enhanced by the presence of substrates, Li and colleagues proposed that after the translation, two M[pro] protomers form a transient dimer which is stabilized by binding the N-terminal site of its substrate (another M[pro] in polyprotein) and further cleave to free its N-terminus.[13] In addition, Chen et al. suggested that the N-terminal autocleavage might only need two immature forms of M[pro] in monomeric polyproteins to form an intermediate dimer that is not related to the active dimer of the mature enzyme.[14]

Herein, we used X-ray crystallography integrated with biochemical techniques to investigate the self-maturation process of SARS-CoV-2 M[pro]. The construct of M[pro] containing N-terminal insertions produced an immature form of the enzyme (IMT M[pro]), unable to form a dimer, that showed a reduced enzymatic activity. We used fragment screening to probe new cavities for drug development in this construct. The inactive mutant C145S with inserted native N-terminal residues (C145S M[pro]) produced a form of the protein that behaves as monomers, dimers, trimers and tetramers in solution. Crystals of the tetrameric form revealed details of the dimeric association of M[pro] during self-processing of its N and C-terminal residues. All forms of the enzyme revealed important conformation changes of the enzyme during maturation, which can guide direct-acting drug development.

Activity and biochemical characterization

A general strategy to produce SARS-CoV-2 M[pro] is to maintain its self-cleavage N-terminal portion and add the HRV-3C cleavage site with a histidine-tag at the C-terminal portion. We successfully used ammonium sulfate precipitation followed by ion exchange chromatography to obtain pure M[pro], simplifying the protocol to one that takes less than 8 h and with a final yield of ~2.5 mg/L of culture. The SARS-CoV-2 IMT M[pro] was obtained by adding a non-cleavable sequence (Gly-Ala-Met) at the N-terminal Ser1 of M[pro], and purified by a similar protocol. The SARS-CoV-2 IMT M[pro] was produced as a soluble protein, yielding ~80 mg/L of culture. To further investigate the role of N-terminal residues in the maturation of M[pro], we designed a construct containing the mutated C145S residue with its native cleavage peptide of M[pro] (Ser-4, Ala-3, Val-2, Leu-1, Gln-0↓) at the N-terminal of Ser1 (Figures 1(a) and S1(a)). During gel filtration, two M[pro] peaks were identified with mass consistent with a monomer and a tetramer (Figure S1).

M[pro] and IMT M[pro] demonstrate to be active and able to recognize and cleave the fluorogenic substrate (Figure 1(b)), with K_m values of 16.4 ± 2.0 μM and 34.3 ± 2.2 μM, respectively. IMT M[pro] exhibited only 0.6% of the catalytic efficiency compared with mature M[pro] (Table 1). As previously reported, the M[pro] N-terminal is fundamental for dimerization and any additional residues would reduce or even abolish its activity.[4,15–18] As expected, C145S M[pro] has only shown residual activity (Figure 1(b)). All three M[pro] constructs exhibited similar thermal-stability profiles, indicating similar folding (Figure 1(c)).

Analysis in solution using SEC-MALS suggests that M[pro] behaves as a dimer in the tested conditions, as expected (Figure 1(c)).[10] For IMT M[pro], the additional residues at N-terminal seem to prevent dimerization completely (Figure 1(c)). For C145S M[pro], however, the additional residues allow the protein to adopt multiple conformational states ranging from monomers to tetramers (Figure 1(c)).

C145S M[pro] in solution characterization

Despite the site-direct mutagenesis of the C145S M[pro], this enzyme exhibited residual proteolytic activity which allowed us to observe the self-processing of the monomeric peak of C145S M[pro] by SDS-PAGE in the course of two days (Figure 1(e)). By quantifying the mass intensity, we estimate that at the end of two days about 30% of the protein was self-cleaved after incubation (Figure 1(e)). By using SEC-MALS, we also monitored the formation of dimers by monomeric C145S M[pro] sample after 0 h, 24 h, 48 h and 72 h incubation at room temperature (Figure 1(f)). At 0 h, the mass recovery ratio between monomers/dimers was 14.9, which decreased to 1.04 at 24 h,
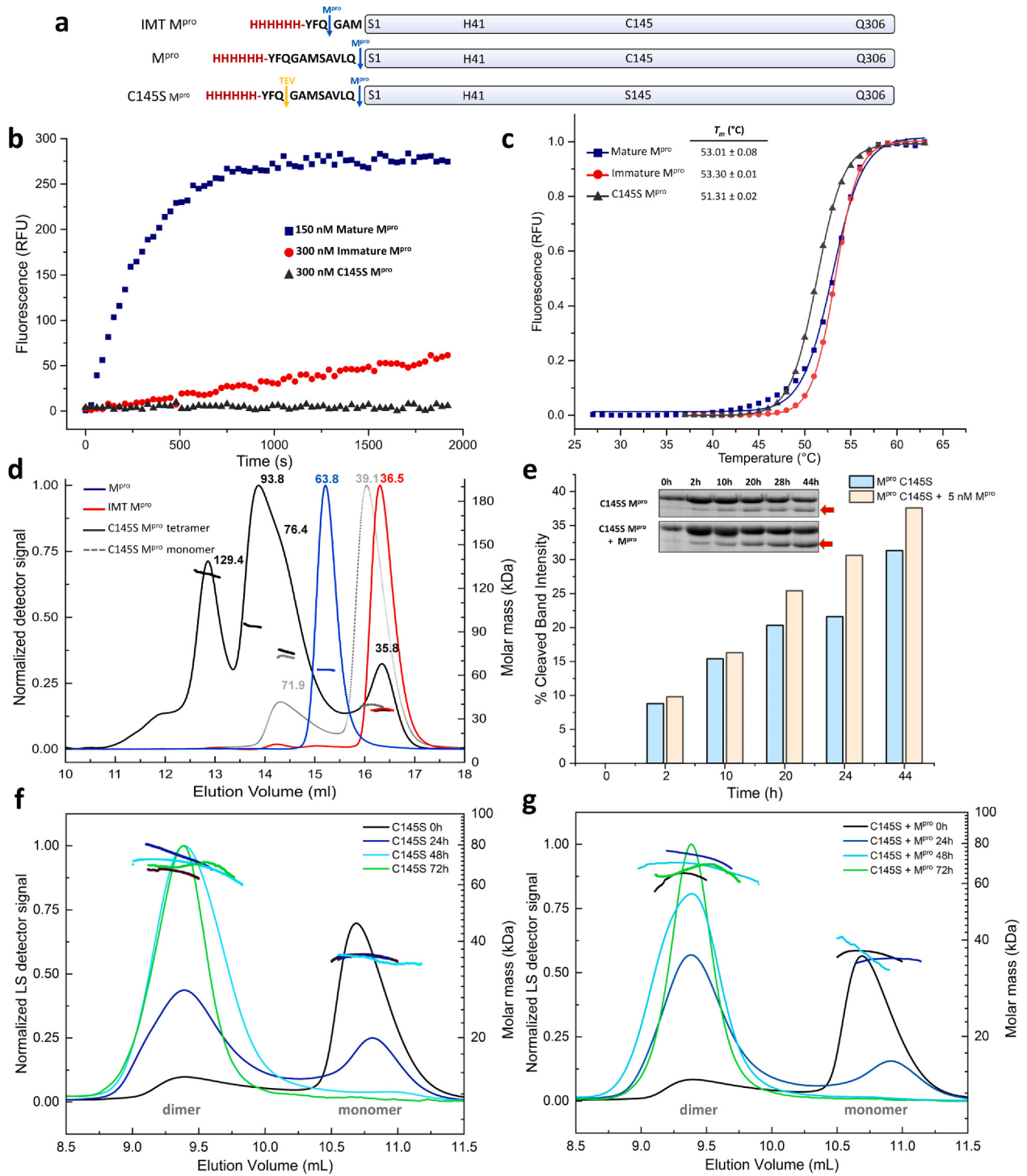
Table 1 Kinetic parameters of M^{pro} constructs. Relative efficiency is the K_m/k_{cat} of constructs relative to M^{pro}

	K_m (μM)	V_{max} (RFU.s^{-1})	k_{cat} (s^{-1})	Relative efficiency
M^{pro}	16.4 ± 2.0	1.05 ± 0.04	28.0 ± 0.1	1
IMT M^{pro}	34.3 ± 2.2	0.104 ± 0.003	0.35 ± 0.07	0.006

and 0.09 at 48 h incubation, with complete degradation of the monomer peak after 72 h (Table 2). This data indicated that the cleavage of the N-terminal is directly proportional to the formation of dimers in solution, highlighting the importance of the N-terminal processing for the assembly of the M^{pro}.

To investigate the effect of M^{pro} in the N-terminal processing, we monitored the effect of adding M^{pro} to the C145S M^{pro} samples in a ratio of 1:6000. On the SDS-Page, we can see that the sample containing M^{pro} showed an increase ratio of protein cleavage after 20 h when compared with the previous experiment (Figure 1(e)). At the SEC-MALS, the mass recovery ratio between monomers/dimers for this sample at 0 h was 12.9, them 0.5 at 24 h, and 0.02 at 48 h incubation, also with complete degradation of the monomer peak after 72 h (Figure 1(g) and Table 2). The data suggests the addition of M^{pro} to the C145S M^{pro} sample at 1:6000 ratio increased the speed of N-terminal processing and dimer formation by the order of 50% after 24 h.

Crystal structure of M^{pro} in monoclinic and orthorhombic crystal system

M^{pro} was crystallized in the monoclinic crystal system in several conditions and its X-ray structure was determined at 1.46 Å in $C2_1$ space group, as the majority of the PDB deposits. All 306 residues were refined at the electron density to a final R_{work}/R_{free} of 0.16/0.18, with 99% of Ramachandran in favored positions (Table S2). The crystal asymmetric unit contains one monomer which could be symmetry expanded to the biological dimer, following the same pattern of

the majority of known structures deposited in PDB (r.m.s.d of 0.2 Å vs PDB 5RGG, for all Cα 306). The M^{pro} protomers are formed by three domains (DI, DII and DIII), with its catalytic region located between the beta-barrels comprising DI and DII[4] (Figure S2). Using seeds from IMT M^{pro}, we were able to obtain a new crystal system in orthorhombic space group $P2_12_12_1$ at final resolution of 1.86 Å. This structure was refined to final R_{work}/R_{free} of 0.19/0.22 and 98.33% of Ramachandran in favored positions (Table S2). This crystal system shows the full dimer in the asymmetric unit, and its packing appears to offer advantages for soaking compounds in M^{pro} active site when compared with the canonical $C2_1$ form, especially for compounds targeting subsites S3-S4, which are less constrained by crystal packing in the orthorombic form (Figure S14). This is being explored by the COVID Moonshot initiative, and will be latter described in a separated manuscript.

Crystal structure of IMT M^{pro}

The crystal structure of IMT M^{pro} at 1.6 Å was determined using 3 merged datasets (Figures S3, S4, Table S1) in $P2_12_12_1$ space group, with two molecules in the asymmetric unit, packed in similar shape to the known biological unit of M^{pro}. The structure was refined to a final R_{work}/R_{free} of 0.20/0.22, with 97% of Ramachandran in favored positions (Table S2). In the recent published structures of GM-M^{pro}, both apo and ligand-complexes exhibited minor differences with the mature form.[15] However, in our structure there are distinguishable differences in the overall structure, especially in the position of DIII helices (Figure 2). Although IMT M^{pro} asymmetric unit resembles the

Figure 1. (a) Schematic showing different constructs of M^{pro}. (b) Time-course reactions of M^{pro} constructs against fluorogenic peptide substrate. (c) Differential scanning fluorimetry of M^{pro} constructs. M^{pro} is shown as blue squares, IMT M^{pro} is shown as red spheres and C145S M^{pro} is shown as black triangles (d) SEC elution profiles with overlaid calculated molar mass from elution peaks. M^{pro} (blue) elutes as a single peak with a calculated molecular mass consistent with a dimer. IMT M^{pro} (red) exhibits a single peak with a mass compatible with a monomer in solution. The monomeric SEC peak of C145S M^{pro} (grey) elutes as an equilibrium between dimers and monomers in solution. The tetrameric SEC peak of C145S M^{pro} (black) contains peaks of monomers, dimer, trimers and tetramers. (e) SDS-PAGE of N-terminal cleavage over time from C145S M^{pro}. At top, reaction containing 10 μM C145S M^{pro}, and at the bottom the same reaction supplemented with 5 nM M^{pro}. Red arrows are pointing to the band of cleaved M^{pro}. The bar graph shows the relative band intensity of cleaved M^{pro} overtime for both reactions in blue and salmon, respectively. (f) SEC elution profiles of monomeric peak of C145S M^{pro} over time. (g) SEC elution profiles of monomeric peak of C145S M^{pro} over time supplemented with 10 nM M^{pro}. In SEC-MAL graphs, curves correspond to the change in the normalized scattered light intensity at 90° (lines) and calculated molar mass of the corresponding peak (dots) are given for each peak.

Table 2 Molecular Mass and Mass Recovery percentages for each observed peak in the SEC-MALS profiles of SARS-CoV-2 M^pro constructs. Mass Recovery Ratio is the ratio of Mass Recovery of Peak 1 and Peak 2

Protein	Time (h)	Peak 1 - Monomer		Peak 2 - Dimer		Mass Recovery Ratio (Peak 1/Peak 2)
		Molar Mass (kDa)	Mass Recovery (%)	Molar Mass (kDa)	Mass Recovery (%)	
C145S M^pro	0	36.0 ± 0.1	74.5	65.7 ± 0.2	5.0	14.9
	24	35.6 ± 0.1	29.6	73.7 ± 0.1	28.4	1.04
	48	34.9 ± 0.1	7.4	69.1 ± 0.1	84.5	0.09
	72	–	–	68.7 ± 0.2	87.3	–
C145S M^pro + M^pro	0	36.4 ± 0.1	75.0	63.0 ± 0.1	5.8	12.9
	24	34.8 ± 0.2	22.8	72.9 ± 0.1	43.5	0.5
	48	36.7 ± 0.2	2.0	67.0 ± 0.1	86.7	0.02
	72	–	–	66.1 ± 0.3	91.3	–

biological dimer form of native protein, PISA[19] analysis indicate that the dimer packing is unstable in solution, with an interface area of 1256 Å^2 (*versus* 1557 Å of M^pro), calculated free energy ΔG of −13.4 kcal/mol (*versus* −14.9 kcal/mol of M^pro) for 26 potential hydrogen bonds (*versus* 33 of M^pro) and 5 potential salt bridges (*versus* 10 of M^pro). The repositioning of Ser1 by itself was responsible for the breaking of five of these hydrogen bonds, maintained by Phe140, Glu166 add Gly170. Interactions between Gly2 and Ser139, Ala7 and Val124, Ser139 and Gln299 and Phe305 with Pro122 are also not present in IMT M^pro.

While IMT M^pro DI and DII are less affected by the N-terminal insertion (r.m.s.d of 0.34 Å vs M^pro for Cα of 1–184), DIII appears to adopt a more open conformation relative to M^pro (r.m.s.d of 1.33 Å for Cα of 201–301) (Figure 2), with the interfacing residues Ala285 at a distance of 9.9 Å in the IMT M^pro (*versus* 5.5 Å in M^pro) (Figure S5). This conformation is more accentuated at chain A where the electron density of the N-terminal insertion is clearly visible in the model. For this chain, the N-terminal insertion pushes chain A helices αF and αH further away from chain B active site, opening a cleft for Phe140 rises to the surface of the molecule, leading to major conformation alterations of the chain B active site souring residues, such as Glu166, Pro168 and Gln189 (Figures 2 and S6). As DIII is known for being extremely flexible,[20] we compared the structure of IMT M^pro with the structure of M^pro in orthorhombic crystal system in order to investigate if the dislocation of DIII was being promoted by the distinct crystal packing. In fact, the structure of M^pro in the orthorhombic crystal system is much more similar to the canonical M^pro (r.m.s.d of 0.52 Å for Cα of 604 residues) rather than IMT M^pro (r.m.s.d of 0.91 Å for Cα of 604 residues), indicating that DIII dislocation is indeed caused by the extra N-terminal residues (Figure S13).

The plasticity of SARS-CoV-2 M^pro active site was already reported when apo X-ray structures collected at cryo and room temperatures were compared,[21] and its expected given the broad spectrum of endogenous substrates that M^pro is has to process. However, the IMT M^pro revealed major structural alterations in the oxyanion hole, likely affecting enzyme processing. The cascade effect of the steric hindrance caused by the N-terminal extra residues affects the position of Ser1, Phe140, Glu166 and Pro168, disrupting the shape of subsites S1, S2 and S4 (Figure 2). Within these the S1 seem to be most affected, assuming an unusual flattened configuration that seem to disrupt the cavity responsible for the recognition of glutamine side-chain, likely affecting substrate recognition (Figure 2). This not only explains the diminished activity of this construct, as well as shows the importance of full N-terminal processing for the correct folding of M^pro. Despite the significant changes of the active site, relative position of the catalytic dyad Cys145-His41 remains unchanged in this form (Figure 2).

Fragment screening of M^pro immature

Recently, a small-fragment library of more than 1,250 unique fragments were screened against SARS-CoV-2 M^pro, identifying 74 high-value fragment hits, including 23 non-covalent and 48 covalent hits in the active site, and 3 hits close to the dimerization interface.[22] In here, we applied the same technique to probe new druggable cavities in IMT M^pro. Although the difference in scale of our experiment, we were able to identify five distinguishable sites in this form of the protein (Figure 3). Site #1 is the active site of chain A, in which fragment f2xe03 was identified interacting with Glu166 N and Cys145 S, corresponding to the S1 substrate subsite. In contrast, the orthorhombic crystal form of IMT M^pro seams promising for compound soaking due to the crystal packing, especially for those binding in S3-S4 subsites. Interesting, a unique cavity marked as Site #3 was identified in our experiments, bound to fragment f2xg05 by Arg4 main chain O. That cavity lies between the interface of chains A and B, and is not present in M^pro which adopt a more closed conformation (Figure S14). This new site and fragment could serve as an anchor for development of new
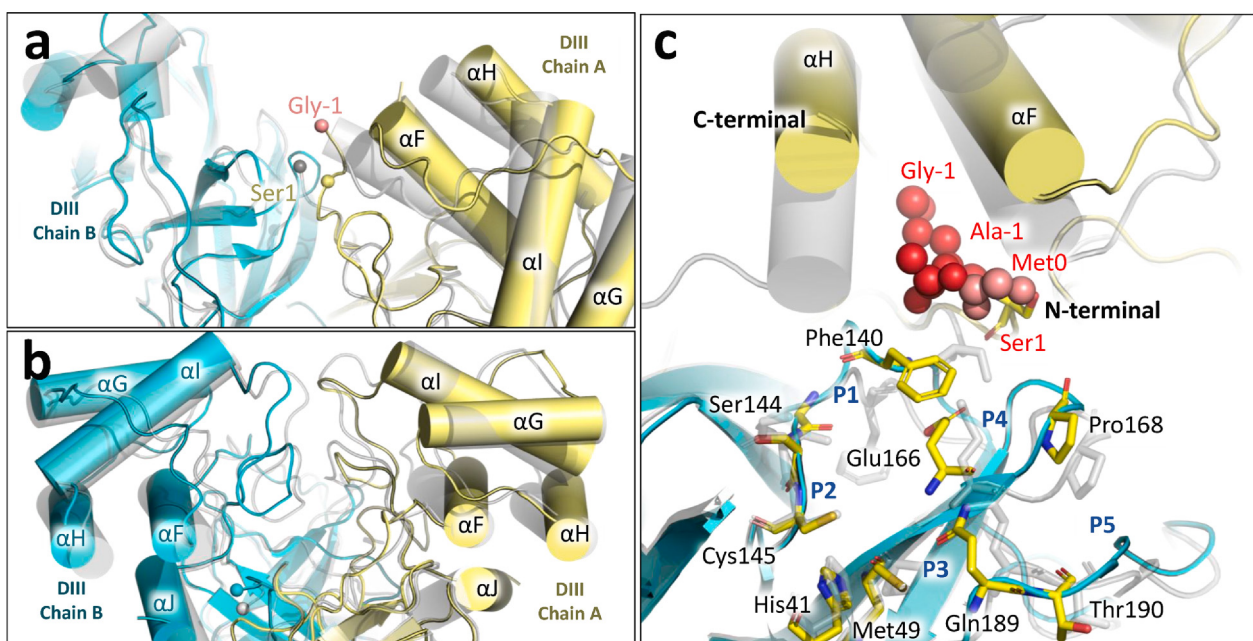
Figure 2. (a) Overview of DIII region from IMT M^{pro} (chain A yellow and B cyan) superposed with M^{pro} (grey ghost). N-terminal residues are depicted as spheres. (b) Rotated view showing IMT M^{pro} DIII from a distinct angle. (c) Active site residues of IMT M^{pro} chain B (cyan cartoon) superposed with M^{pro}. Catalytic residues are depicted as yellow sticks. N-terminal chain A residues are depicted as spheres. M^{pro} structure and residues are shown as a grey ghost.

inhibitors targeting M^{pro} dimerization process, a mode of action that was too date only theorized.[23] Details about data processing and statistics are given in Table S3.

Crystal structure of C145S M^{pro} in complex with N and C-terminal residues

The tetramer peaks of C145S M^{pro} were crystallized and its X-ray structure was determined at 2.8 Å and R_{work}/R_{free} of 0.20/0.25 (Table S2), revealing a new crystal form in which N-terminal cleaved residues (Ser-4, Ala-3, Val-2, Leu-1, Gln-0) were found trimmed in the active site of chain A, occupying subsites S1-S5 (Figure 4(a)). Despite the site directed mutagenesis of the catalytic cysteine to serine, electron density shows that Gln0 and Ser1 are non-covalently bound in the amino region, clearly indicating that the N-terminal cleavage was completed. At the S1 subsite, Gln0 NE2 interacts with Glu166 OE1 by a hydrogen bound (2.7 Å), while Gln0 form interacts with Ser145 in the position of the native oxyanion hole (Figures 4, S8 and S9). To accommodate the hydrophobic sidechain of Leu-1 at P2, Met49 and Met165 are pushed further of each other (Figure S10), leading to a more opened groove of this subsite relatively to the apo-state, explaining the ability of this subsite to accommodate a variety of hydrophobic side chain residues, such as Leu, Met, Ile, Val and Phe.[6,24] Yet, from the eleven endogenous recognition sites of coronaviruses, S2 Leu carrying sequences are the ones in which M^{pro}

display higher catalytic efficiency, highlighting the importance of this conformation for drug design. At subsites S3-S5, the interactions of Val-2, Ala-3 and Ser-4 are mainly maintained by hydrogen bounds between the polar residues of protein and peptide side chains (Figure 4(a)), which explains the ability of M^{pro} to recognize a large variety of chain sequences at those positions.

The crystal structure of C145S M^{pro} revealed another important step in the maturation process of M^{pro}, as we can see that chain B C-terminal residues (301–306) are ~180° rotated from its canonical conformation (Figure S11) and trimmed in the active site of a symmetric related chain B (Figure 4(b)), a phenomenon that was also recently observed by another group in the C145S M^{pro} mutant.[25] During this event, two C145S M^{pro} dimers appear to be linked by the interaction of the C-terminal and a respective active site, revealing details of the dimeric association in a non-closed complex (Figure 5). The electron density of this dataset indicates that chain B Ser145 OG is covalently bound to Gln306 C from crystallographic symmetric correlated chain B (distance of 1.4 Å), with the loss of one oxygen by Gln306 (Figure S9). We believe the diminished activity of the mutant as allowed the formation of these crystals after almost 20 days, from which we were able to capture this intermediary state of the maturation. We highlight here that the model deposited model does not depicted this covalent bound, as we found impossible to link two atoms outside the asymmetric unit
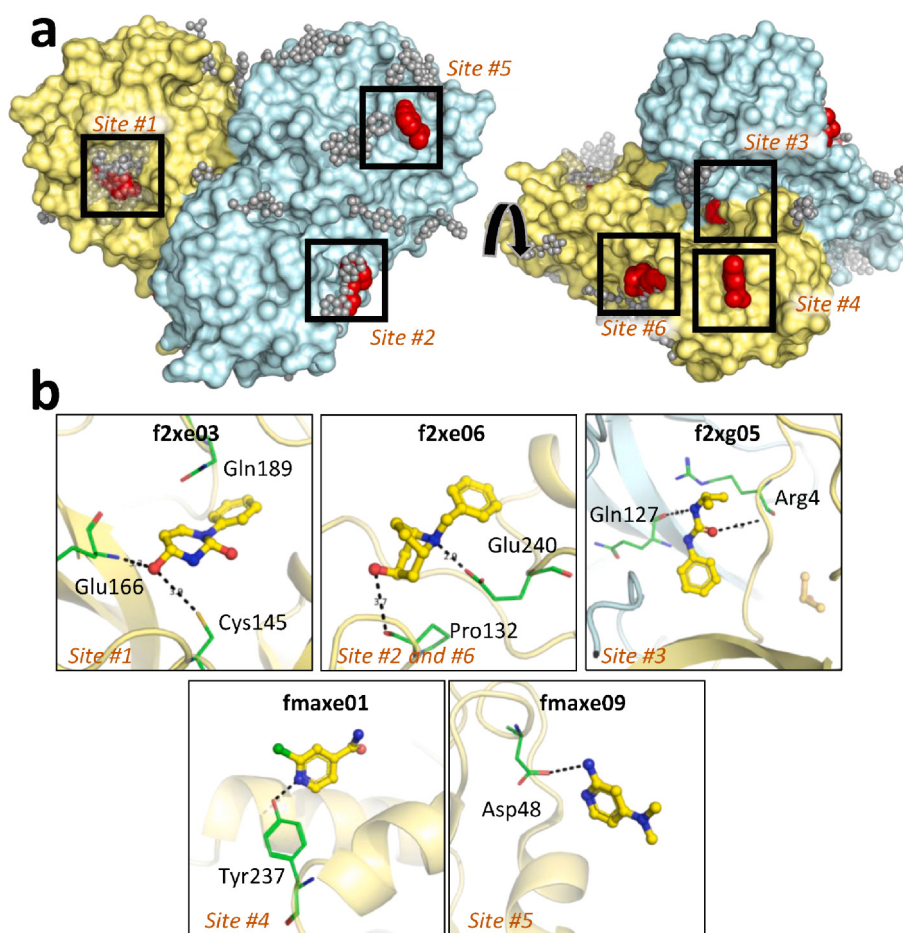
Figure 3. (a) Location of IMT M^{pro} probing fragments identified during screening. Chain A is colored as yellow surface, chain B as cyan surface. Fragments are depicted as red spheres. For comparison, fragments of previous manuscript using monoclinic M^{pro} were aligned to the structure of IMT M^{pro} and are depicted as grey spheres.[22] (b) Contact details of identified fragments bound to IMT M^{pro}. Chain A is colored as yellow cartoon and chain B as cyan cartoon. Fragments are depicted as yellow sticks. Residues forming polar contacts are depicted as green lines. Contacts are depicted as black dashes.

(even after consultant with software developers). Within the active site, Gln306 occupies the respective position of Gln0 at S1, while S2 is occupied by Phe305, increasing the distance between Met49 and Met165 relatively to chain A bound to N-terminal (Figure S10). As for the N-terminal residues, subsites S3-S5 interactions with C-terminal are mainly maintained by hydrogen bounds between main chains (Figure 4(b)).

The maturation process of M^{pro} and its impact on drug discovery

M^{pro} is firstly produced as the Nsp5 domain of the viral polyproteins before they are proteolytically processed into 15 or 16 non-structural proteins.[11] Immediately after translation, the immature form of M^{pro} would contain both N and C-terminal insertions, which requires self-processing to generate the mature form of the enzyme.[12] In M^{pro}, the cleaved N-terminals are sandwiched between the two protomers of the dimeric enzyme, being a part not only of the dimer interface but also from the respective protomer active site. In the IMT M^{pro}, the N-terminal extra amino acids seem to disrupt the active site shape at S1-S3 subsites (Figure 5), affecting its capacity of recognizing the substrate and processing. Notwithstanding, the extra N-terminal residues also seem affect the enzyme ability to form dimers by pushing the reciprocal DIII further to its native conformation.

The same process appears to occurs to C145S M^{pro} monomers with native N-terminal inserted residues, although in this case, the slow cleavage of the N-terminus results in the formation of dimers overtime (Figure 1(e)). The incubation of this samples to allow dimerization appears to significantly enhance the enzymatic residual activity of this construct, indicating that the dimeric form is important for activity even for this serine mutant (Figure 6). By monitoring the formation of dimer overtime, we saw that the monomeric
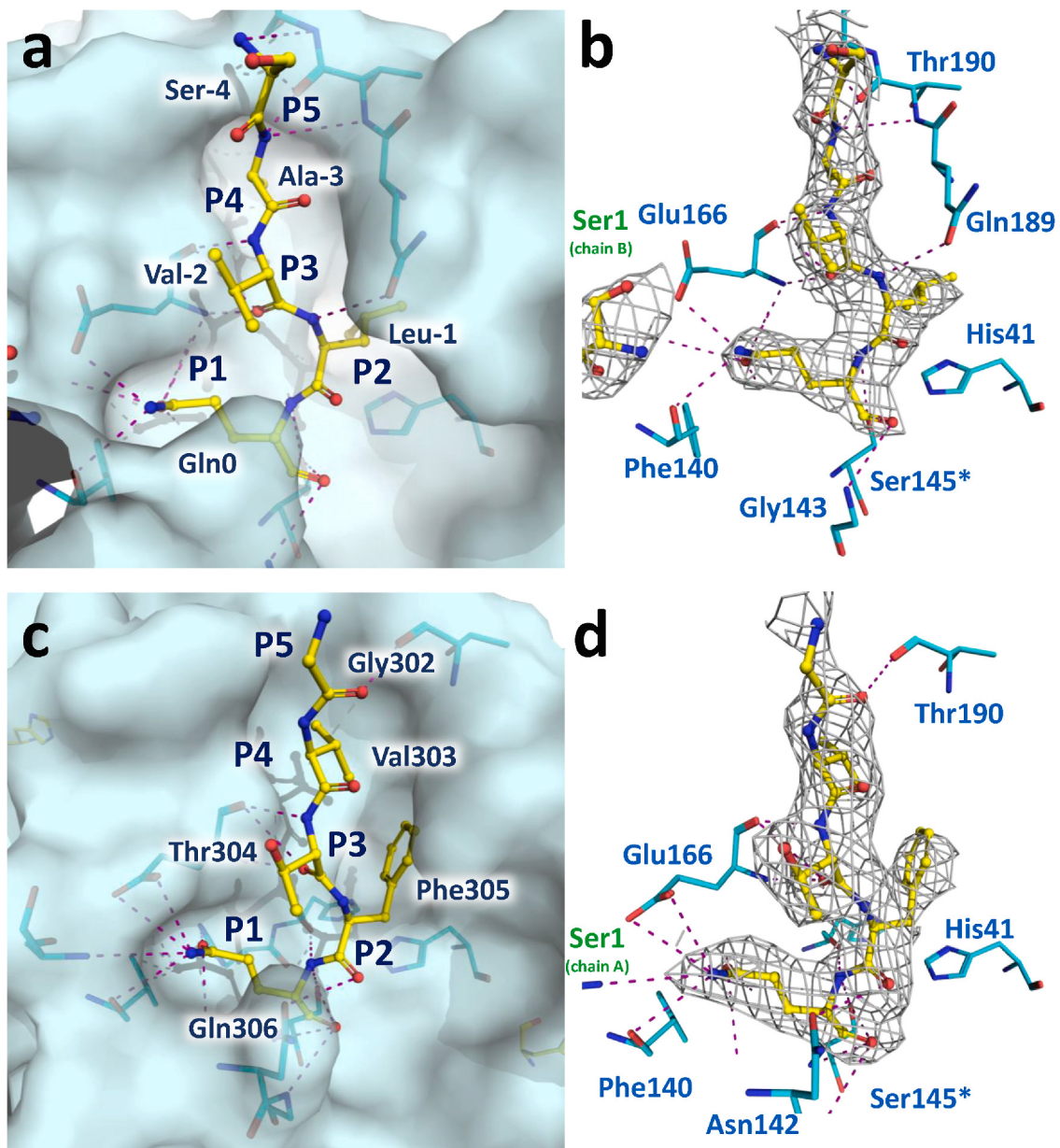
Figure 4. (a) C145S M^pro chain A active site (cyan surface) in complex with processed N-terminal residues (yellow sticks). Main interacting residues are depicted as blue lines. (b) C-terminal peptide (yellow) main interactions with C145S M^pro chain A active site residues (blue). (c) C145S M^pro chain B active site (blue surface) in complex with processed C-terminal residues (yellow sticks). Main interacting residues are depicted as blue lines. (d) C-terminal peptide (yellow) main interactions with C145S M^pro chain B active site residues (blue). For (b) and (d), the 2mFo-DFc electron density contoured at 0.8σ. Ser1 from respective dimerization partners are depicted with green letters. *Ser145 is the site-direct mutant of Cys145. Simulated annealing omit map is available in Figure S7.

enzyme is capable of processing its N-terminal, suggesting that cis-cleavage as a mechanism for the first step of the maturation process (Figure 1 (e)). However, when we compare the results from both experiments, we notice that the ratio of dimer formation seen to be far superior to the ratio of N-terminal processing (Figure 1(e)–(g)). This is in partial agreement with the model proposed by Li and colleagues (2010) in which two M^pro form a transient dimer that is stabilized by the binding the

N-terminal site of its substrate (another M^pro in polyprotein) and further cleave to free its N-terminus.[13] It is also another argument for our model of the immature form, in which a partially cleaved M^pro would result in a constrained packing of M^pro with diminished activity (Figure 2).

During our analysis we also observed that when M^pro is added to C145S M^pro, the N-terminal cleavage and dimer formation seem to be enhanced significantly (Figure 1(e)–(g)),
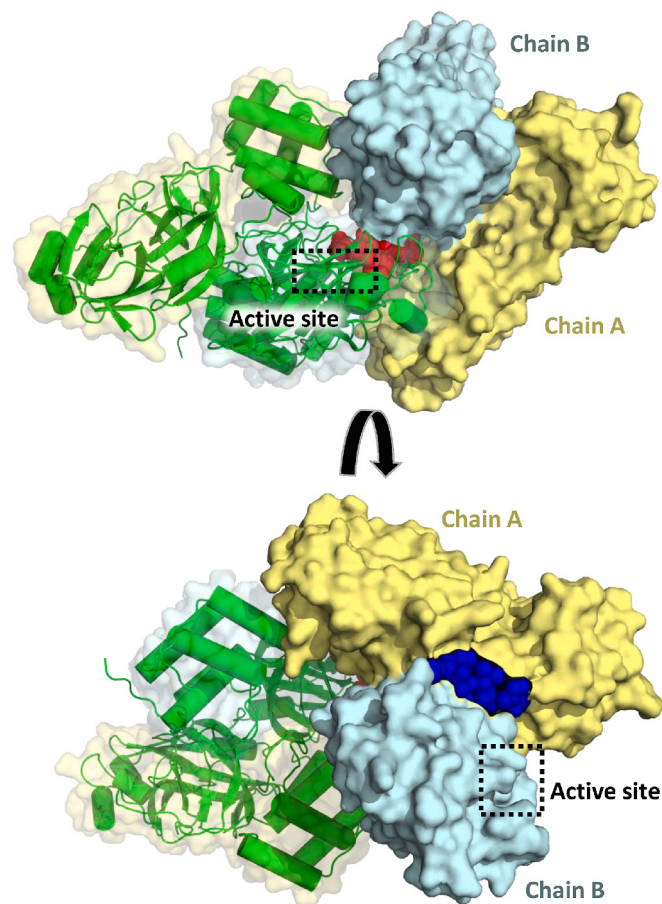
Figure 5. Overview of the dimer-dimer association intermediary formed by C145S M^{pro} tetramer during self-processing. Chain A is colored as yellow surface, chain B as cyan surface. *Trans*-cleavage M^{pro} partner is show as green cartoon. N-terminal residues are depicted as blue spheres, and C-terminal residues are depicted as red spheres.

suggesting that this initial maturation step is a mix of *cis* and *trans*-cleavage events. Despite our efforts, we were not able to obtain a complex showing the details of an intact N-terminal protomer in complex with M^{pro}. Our tentative of crystallization of monomeric C145S M^{pro} only resulted in crystals that were identical to the canonical monoclinic crystals of M^{pro}, with fully cleaved N-terminal and dimeric packing. (PDBid 7N5Z, r.m.s.d of 0.25 with M^{pro} 7KPH for 306 Cα). Still, more studies are required to understand the exact order of the events and inter-protomer interactions involved in the N-terminal cleavage.

After the active site region is matured (or even concomitantly), dimeric M^{pro} C-terminal seem to assume an unusual rotated position (Figure S11), allowing it to bound into the active site of another mature or half-mature M^{pro} dimer (Figure 4(b)). In this step, the trans-cleavage processing of the C-terminal residues would serve as an anchor for a transitory dimer-dimer association state of the protein, herein captured with the construct of the mutant C145S M^{pro} with the processing of

the N-terminal residues (Figure 5). As a result, full mature M^{pro} is produced and its ready to process other parts of the viral polyprotein. During all those maturation processes, both M^{pro} active site and surface undergo significant conformational changes, which could guide targeted drug development (Figure 7 and Supplementary Video 1). Our results not only shed light in the self-maturation process of SARS-CoV-2 M^{pro}, but also bring the perspective of developing drugs targeting intermediate states of this enzyme.

Materials and Methods

Cloning and expression of SARS-CoV-2 M^{pro} forms

The viral cDNA template (GenBank MT126808.1), kindly provided by Dr. Edison Durigon (University of São Paulo, São Paulo, Brazil), was synthetized using the SCRIPT One-Step RT-PCR kit (Cellco Biotec) and random hexamers primers. For production of IMT M^{pro},
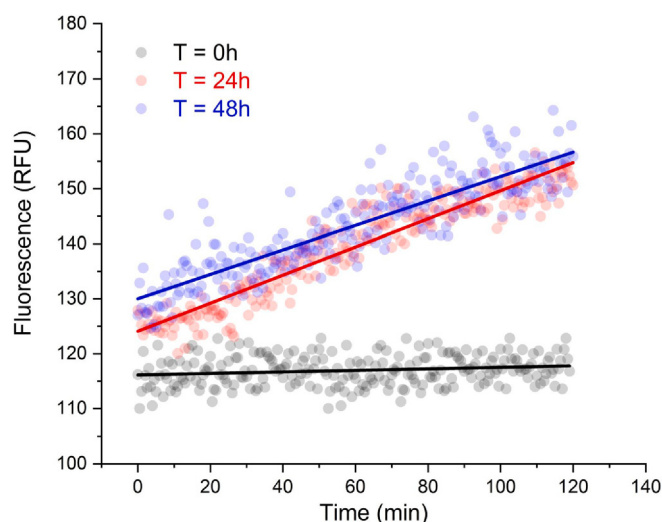
Figure 6. Time-curse reactions of C145S M^pro monomeric construct determined after different incubation periods. Activity of the construct was monitored after 0 h (black), 24 h (red) and 48 h (blue) incubation.

coding region of M^pro (residues 3264–3569) was amplified using primers: Fw 5′ CAGGGC GCCATGAGTGGTTTTAGAAAAATGGCATTC 3′ and Rv 5′ GACCCGACGCGGTTATTGGAA AGTAACACCTGAGAC 3′, and the sequence was inserted into the pET_M11 vector, which encodes an N-terminal 6xHis-tag followed by a TEV protease cleavage site (ENLYFQ↓GAM), using the LIC method,[26] forming the plasmid pET_M11-IMT-M^pro. To obtain the mature form of M^pro, native N-terminal residues (GAMSAVLQ↓SGFRK) were inserted into pET_M11-IMT-M^pro by inverse PCR using primers: Fw: 5′ GCTGCAGAG TGGTTTTAGAAAAATGGCATTC 3′ and Rv: 5′ ACGGCTGACATGGCGCCCTGAAAATA 3′. Amplified product was treated with T4 Polynucleotide Kinase (PNK, Thermo Fischer Scientific) and T4 Ligase (Cellco Biotec), forming plasmid pET_M11- M^pro. For C145S M^pro construct, pET_M11- M^pro was used as template for inverse PCR with primers Fw 5′ CCTTAATGGTT-CATCTGGTAGTG 3′ and Rv 5′ AATGAACCCT-TAATAGTGAAATTGG 3′. The PCR product was digested with DpnI (NEB), followed by treatment with PNK and T4 DNA ligase, forming the pET_M11-C145S-M^pro. All plasmids were transformed in DH5α *E. coli* competent cells. All PCRs were conducted with FastPol (Cellco Biotech). Positive clones were selected and confirmed by sequencing. Schematics of constructs are given in Figure S1.

For protein production, *E. coli* BL21 cells were transformed with respective plasmids and cultured in ZYM-5052[27] at 37 °C and 200 RPM to an OD$_{600}$ of 0.8, followed by expression at 18 °C, 200 RPM for 16 h. Cells were harvested by centrifugation at 5000g for 40 min at 4 °C, resuspended in lysis buffer (20 mM Tris pH 7.8, 150 mM NaCl, 1 mM DTT), disrupted by sonication and the lysate

was clarified by centrifugation at 12,000g for 30 min at 4 °C.

Protein purification of IMT M^pro

After expression, a large amount of IMT M^pro had its 6xHis-tag cleaved by autoproteolytic process. The small fraction of 6xHis tagged protein was removed from the lysate using Ni-NTA resin (Qiagen). The cleaved protein was purified by adding 1 M ammonium sulfate to the cell lysate followed by incubation on ice for 10 min. The precipitated protein was recovered by centrifugation at 12,000g for 30 min at 4 °C, resuspended in gel filtration buffer (20 mM Tris pH 7.8, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) and purified by size-exclusion chromatography using a HiLoad 26/100 Superdex 200 column (GE Healthcare) pre-equilibrated with gel filtration buffer. Purified fractions were aliquoted, flash-frozen and stored at −80 °C for enzymatic assays and crystallization. For crystallization, protein was concentrated to 14 mg mL^{-1} using 10 kDa MWCO centrifugal concentrators (Vivaspin, Sartorius). Protein concentrations were determined using the measured absorbances at 280 nm and the theorical extinction coefficient of 32,890 M^{-1} cm^{-1}. Protein purity was analyzed by SDS-PAGE (Figure S1).

Protein purification of M^pro

M^pro was purified similar to IMT M^pro, with an additional step of cation exclusion chromatography. After the size exclusion chromatography, the protein was buffer exchanged to 20 mM Tris pH 8.0, 1 mM DTT, and then injected into a Mono-Q 5/50 GL column (GE Healthcare). Protein was eluted using a linear
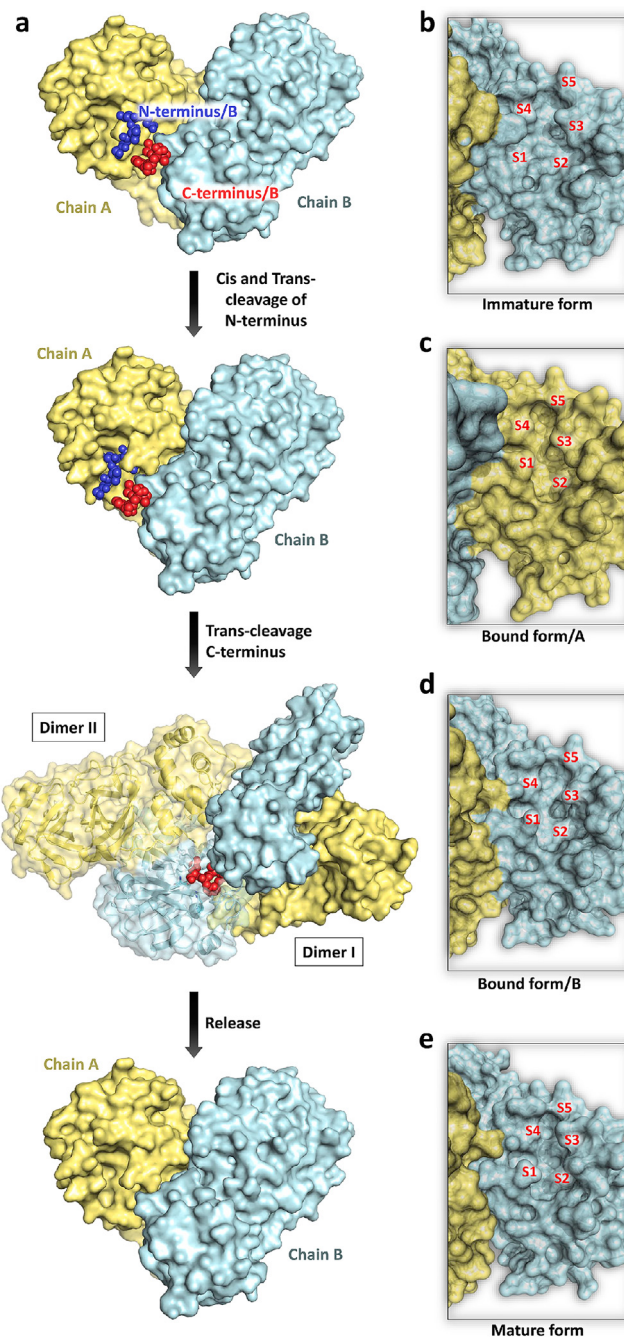
Figure 7. Scheme containing steps of SARS-CoV-2 M^pro self-maturation process. (a) At first, two protomers assembly as an immature dimer during N-terminal cis and trans-cleavage. After processing, the M^pro with the matured active site permit the correct positioning of DIII, which allows the stabilization of the dimeric form. The dimer C-terminal is them trans-cleaved by another full or at least half mature dimer, forming a transient dimer-dimer association and producing the full mature form of M^pro. (b) Surface view of chain B active site from immature form. (c) Surface view of chain A active site during N-terminal residues recognition. (d) Surface view of chain B active site during C-terminal residues recognition. (e) Surface view of full mature M^pro active site.

gradient of a buffer containing 20 mM Tris pH 8.0, 1 M NaCl and 1 mM DTT. Finally, fractions containing the purified protein were buffer exchanged to gel filtration buffer. Purified fractions were aliquoted and protein was concentrated and quantified similarly to IMT M^pro (Figure S1).

Protein purification of C145S M^pro

For C145S M^pro, protein was purified by immobilized metal chromatography (IMAC) using a 5 mL HisTrap FF column (GE Healthcare). After column washing with buffer A (20 mM Tris pH 7.8,

150 mM NaCl, 25 mM Imidazole), protein was eluted with buffer A supplemented with 250 mM imidazole. Sample was buffer exchanged using a 5 mL HiTrap desalting column (GE Healthcare) equilibrated with buffer A. To remove the 6xHis-tag, 2 mg of TEV protease and 4 mM DTT were added to the sample and incubated for 16 h at 4 °C. Next day, non-cleaved protein and TEV were removed by a second step of IMAC in buffer A. Finally, the protein was purified by size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated with gel filtration buffer. Purified fractions were aliquoted, and protein was concentrated and quantified as described for other constructions (Figure S1).

Crystallization

Crystallization screening was performed with the sitting drop vapor diffusion method in 96-well plates using a Phoenix Liquid Handling System-Gryphon LCP (Art Robbins Instruments) and commercially available kits at 20 °C. For M^{pro}, crystals appeared after 1 day in 0.1 M Bis-Tris pH 6.5, 25% PEG 3350, which were cryo-protected using the reservoir solution and 30% PEG 400. Crystals of IMT M^{pro} were observed in several conditions. After optimization, crystals grown in 0.1 M MES pH 6.5, 10% 2-propanol, 20% PEG Smear Low (BCS Screen, Molecular Dimensions) were used as seeds for the diffraction crystals, obtained in 0.1 M MES pH 6.7, 5% DMSO, 8% PEG 4000.[22] Crystals of M^{pro} in orthorhombic crystal system were obtained using seeds of IMT M^{pro} combined with M^{pro} at the condition 0.1 M MES pH 6.7, 5% DMSO, 8% PEG 4000 at 20 °C. Crystals of C145S M^{pro} tetrameric were obtained after 20 days in 0.1 M phosphate/citrate, pH 5.5, 20% v/v PEG Smear High (BCS Screen, Molecular Dimensions) at 20 °C. Crystals of C145S M^{pro} monomeric were obtained after 20 days in 0.1 M Bis-Tris pH 6.5, 25% (w/v) PEG 3350 at 20 °C.

Data collection and processing at MANACA beamline

During the initial commissioning phase (July to October 2020) the MANACA (MAcromolecular Micro and NAno CrystAllography)[28] beamline adopted an emergency commissioning plan to deliver the basic instrumentation to perform data collection of SARS-CoV-2 related samples. Thus, during this period, the beamline has operated on a fixed-energy regime (9 keV) with manual crystal mounting, single-axis goniometry, beam flux estimated to be about $1 \cdot 10^{11}$ ph/s/10 mA at 9.15 keV and adjustable beam size from about 18 (h) × 20 (v) μm^2 to 100 (h) × 80 (v) μm^2 (FWHM). This project was the first external user session at MANACA beamline and the first operating beamline at Sirius

(SIRIUS, Brazil). The focus was optimised to 61 (h) × 48 (v) μm^2 at sample position (Figure S3). Even without the full capabilities, the beamline opening was very important to SARS-CoV-2 structural biology studies.

X-ray data for apo IMT M^{pro} was collected from three isomorphous independent crystals, that were processed by XDS via autoPROC.[29,30] Data herein was used for confirm reliability of the beamline (Figure S4 and Table S1). Datasets were then scaled and merged using Aimless[31], and the resulting dataset was used for structural determination of IMT M^{pro} by molecular replacement with Phaser[32] using PDB 5RGQ as template. Model was refined with COOT[33] and BUSTER[34] at 1.6 Å and deposited under the code 7KFI.

X-ray data for mature M^{pro} and C145S M^{pro} monomeric and tetrameric were processed by XDS via autoPROC[29,30] and scaled using Aimless.[31] Mature M^{pro}, C145S M^{pro} tetrameric and C145S M^{pro} monomeric were solved by molecular replacement with Phaser[32] using template models 5RGQ, 7KFI and 5R8T respectively. Mature M^{pro} and C145S M^{pro} monomeric and tetrameric were refined with COOT[33] and phenix.refine[35], and are respectively deposited under the codes 7KPH (at 1.4 Å), 7N5Z (at 1.7 Å)) and 7N6N (at 2.8 Å). Details of data processing parameters and statistics are given in Table S2.

Fragment screening of IMT M^{pro}

For the fragment screening of IMT M^{pro}, we used the settled plates fragment libraries of FragMAXlib (Talibov *et al.*, to be published) and F2XEntry,[36,37] in a total of 192 fragments tested. In those plates, the content of each drop-well was resuspended in 1.0 μL of 0.1 M MES pH 6.7, 5% DMSO (v/v), 8% PEG 4,000 (w/v), 30% PEG 400 (w/v), and crystals were added afterwards. After 4 h soaking at room temperature, crystals were manually harvested and flash cooled for data collection.

During the commissioning phase of MANACA, 166 of those crystals were tested, leading to 77 usable datasets. To analyze the data, a simplified version of FragMAXapp was configured in our laboratory end-station computer.[38] Within Frag-MAXapp, restrictions libraries were generated by phenix.eLBOW[39] using rm1 force field for geometry optimization, and datasets were processed through autoPROC/STARANISO or DIALS via XIA2.[29,40,41] Molecular replacement and initial refinement were performed using DIMPLE[42] using PDB 7KFI as template. To highlight electron density of weak binding events, map averaging and statistical modelling were performed by PanDDA software.[43] Models were refined with COOT[33] and phenix.refine.[35] Details of data processing and refinement statistics are given in Table S3. Polder maps of fragments are available in Figure S12.[44]

Activity assays

All enzymatic assays were carried out using FRET-based substrate DABCYL-KTSAVLQ↓SGF RKM-E(EDANS)–NH$_2$ in assay buffer (20 mM Tris pH 7.3, 1 mM EDTA, 1 mM DTT). M^{pro}, IMT M^{pro} and C145S M^{pro} assays were performed at final concentration of 0.14 μM, 0.3 μM and 0.3 μM, respectively. Prior to reactions, enzymes were incubated in assay buffer at 37 °C for 10 min, followed by substrate addition. To determine the kinetics parameters (K_m, V_{max} and k_{cat}), the substrate was diluted to a range of concentrations from 100 μM to 0.78 μM. Initial velocity was derived from the slope of linear phase of each time-curse reaction, and Michaelis-Menten fitting was obtained using Origin Pro 9.5.1 Software (OriginLab). Relative efficiency of constructs was calculated by comparing the K_m/k_{cat} relative to M^{pro}. Fluorescence measures were performed in SpectraMax Gemini EM Microplate Reader with $\lambda_{exc}/\lambda_{emi}$ of 360/460 nm, every 30 s over 60 min at 37 °C. All assays were performed in triplicates.

To test C145S M^{pro} auto-cleavage activity, 6xHis-tagged C145S M^{pro} was buffer exchanged in 20 mM Hepes pH 7.3, 100 mM NaCl, 1 mM DTT. Two reactions were prepared for comparison, one containing 10 μM C145S M^{pro} and other containing 30 μM C145S M^{pro} and 5 nM of mature M^{pro}. Aliquots of each reaction were collected for the period of 44 h. Samples were analyzed by SDS-PAGE on a 12.5% SDS polyacrylamide gel.

Differential scanning fluorimetry

For differential scanning fluorimetry assays (DSF), SYPRO Orange at 5X final concentration was added to protein diluted to 25 μM protein in gel filtration buffer. Denaturation curves were obtained ranging temperature from 25 °C to 75 °C by increasing one degree per cycle and fluorescence was measured in the end of each cycle. Experiment was conducted in a qPCR system Mx3000P (Agilent). The melting temperature was obtained using the Boltzmann fitting on Origin Pro 9.5.1 Software (OriginLab).

In solution oligomeric state of M^{pro} constructs

The in solution oligomeric states of the purified samples were evaluated by size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) in running buffer composed by 20 mM Tris-HCl pH 7.8 and 100 mM NaCl. For that, 50 μL of each M^{pro} construct at concentration of 50 μM were injected in a Waters 600 HPLC system (Waters) coupled in-line with an UV detector, a miniDAWN TREOS multi-angle light scattering apparatus (Wyatt Technology), a column Superdex 200 Increase 10/300 GL (GE Healthcare), and a refractive index detector Optilab T-rEX (Wyatt Technology). The light-scattering detectors were normalized with bovine serum albumin (Sigma-Aldrich). Data were collected and analyzed with the ASTRA 7 integrated software provided by Wyatt. The flow rate used was 0.5 mL/min. Results are summarized in Table S4.

For dimer formation monitoring, in solution oligomeric states of C145S M^{pro} monomer peak were determined similarly by SEC-MALS using a Superdex 75 Increase 10/300 GL (GE Healthcare). Two reactions were conducted simultaneously one containing 60 μM C145S M^{pro} monomer and other containing 60 μM C145S M^{pro} monomer and 10 nM M^{pro} (ratio 6000:1). Both reactions were maintained at room temperature and samples of 50 μL were injected with 24 h intervals for the period of 72 h.

Accession numbers

Structure factors and atomic coordinates have been deposited with the protein data bank with accession codes PDB ID **7KFI**, **7MBG**, **7KPH**, **7N6N**, **7LFE**, **7LDX**, **7LFP**, **7KVL, 7N5Z** and **7KVR**. Other data are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

G.D. Noske: Conceptualization, Methodology, Software, Data curation, Writing - original draft. **A. M. Nakamura:** Conceptualization, Methodology, Software, Data curation, Writing - original draft. **V. O. Gawriljuk:** Conceptualization, Methodology, Software, Data curation, Writing - original draft. **R. S. Fernandes:** Conceptualization, Writing - original draft. **G.M.A. Lima:** Software. **H.V.D. Rosa:** Investigation, Methodology. **H.D. Pereira:** Investigation. **A.C.M. Zeri:** Investigation. **A.A.F.Z. Nascimento:** Investigation. **M.C.L.C. Freire:** Writing - original draft. **D. Fearon:** Investigation. **A. Douangamath:** Investigation. **F. von Delft:** Investigation. **G. Oliva:** Supervision, Funding acquisition. **A.S. Godoy:** Conceptualization, Methodology, Software, Data curation, Investigation, Writing - original draft, Writing - review & editing, Software, Validation.

Acknowledgments

Mikolajek for generously providing wt M^{pro} protein for use at Diamond Light Source.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2021.167118.

† Current address: Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, UK.

Abbreviations used:
DMSO, dimethyl sulfoxide; DSF, differential scanning fluorimetry; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; FRET, fluorescence resonance energy transfer; HPLC, high performance liquid chromatography; HRV, Human Rhinovirus; IMAC, immobilized metal affinity chromatography; LIC, ligation independent cloning; M^{pro}, main protease; MERS-CoV, Middle East Respiratory Syndrome Coronavirus; ORF, open reading frame; PNK, T4 polynucleotide kinase; r.m. s.d., root mean square deviation; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; SEC-MALS, size exclusion chromatography coupled with multi-angle light scattering; TEV, Tobacco Etch Virus; 3CL^{pro}, 3C-like protease

References

1. Zhu, N., (2020). A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.*, **382**, 727–733. https://doi.org/10.1056/NEJMoa2001017.

2. Zhou, P., Lou Yang, X., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., et al., (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, **579**, 270–273. https://doi.org/10.1038/s41586-020-2012-7.

3. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z. G., Hu, Y., Tao, Z.W., et al., (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265–269. https://doi.org/10.1038/s41586-020-2008-3.

4. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., et al., (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science*, **368**, 409–412. https://doi.org/10.1126/science.abb3405.

5. Kumar, S., Nyodu, R., Maurya, V.K. & Saxena, S.K. (2020). Morphology, Genome Organization, Replication, and Pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). In Coronavirus Dis. 2019, pp. 23–31, Nature Publishing Group, Singapore. https://doi.org/10.1007/978-981-15-4814-7_3.

6. Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., et al., (2004). Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J. Biol. Chem.*, **279**, 1637–1642. https://doi.org/10.1074/jbc.M310875200.

7. Yoshino, R., Yasuo, N., Sekijima, M., (2020). Identification of key interactions between SARS-CoV-2 main protease and inhibitor drug candidates. *Sci. Rep.*, **10**, 12493. https://doi.org/10.1038/s41598-020-69337-9.

8. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., Hilgenfeld, R., (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science*, **300**, 1763–1767. https://doi.org/10.1126/science.1085658.

9. Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., Zhu, Y., Zhu, C., et al., (2020). Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nature Struct. Mol. Biol.*, **27**, 529–532. https://doi.org/10.1038/s41594-020-0440-6.

10. Ma, C., Sacco, M.D., Hurst, B., Townsend, J.A., Hu, Y., Szeto, T., Zhang, X., Tarbet, B., et al., (2020). 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. https://doi.org/10.1038/s41422-020-0356-z.

11. Xia, B., Kang, X., (2011). Activation and maturation of SARS-CoV main protease. *Protein Cell.*, **2**, 282–290. https://doi.org/10.1007/s13238-011-1034-1.

12. Hsu, M.F., Kuo, C.J., Chang, K.T., Chang, H.C., Chou, C. C., Ko, T.P., Shr, H.L., Chang, G.G., et al., (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.*, **280**, 31257–31266. https://doi.org/10.1074/jbc.M502577200.

13. Li, C., Qi, Y., Teng, X., Yang, Z., Wei, P., Zhang, C., Tan, L., Zhou, L., et al., (2010). Maturation mechanism of severe acute respiratory syndrome (SARS) coronavirus 3C-like proteinase. *J. Biol. Chem.*, **285**, 28134–28140. https://doi.org/10.1074/jbc.M109.095851.

14. Chen, S., Jonas, F., Shen, C., Higenfeld, R., (2010). 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. https://doi.org/10.1007/s13238-010-0011-4.

15. Fu, L., Ye, F., Feng, Y., Yu, F., Wang, Q., Wu, Y., Zhao, C., Sun, H., et al., (2020). Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nature Commun.*, **11**, 1–8. https://doi.org/10.1038/s41467-020-18233-x.

16. Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., Hilgenfeld, R., (2002). Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. *EMBO J.*, **21**, 3213–3224. https://doi.org/10.1093/emboj/cdf327.

17. Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., et al., (2007). Production of authentic SARS-CoV Mpro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J. Mol. Biol.*, **366**, 965–975. https://doi.org/10.1016/j.jmb.2006.11.073.

18. Sacco, M.D., Ma, C., Lagarias, P., Gao, A., Townsend, J. A., Meng, X., Dube, P., Zhang, X., et al., (2020). Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L. *Sci. Adv.*, **6**, eabe0751. https://doi.org/10.1126/sciadv.abe0751.

19. Krissinel, E., Henrick, K., (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.

20. Tan, J., Verschueren, K.H.G., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., et al., (2005). pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J. Mol. Biol.*, **354**, 25–40. https://doi.org/10.1016/j.jmb.2005.09.012.

21. Kneller, D.W., Phillips, G., O'Neill, H.M., Jedrzejczak, R., Stols, L., Langan, P., Joachimiak, A., Coates, L., et al., (2020). Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nature Commun.*, **11**, 7–12. https://doi.org/10.1038/s41467-020-16954-7.

22. Douangamath, A., Fearon, D., Gehrtz, P., Krojer, T., Lukacik, P., Owen, C.D., Resnick, E., Strain-Damerell, C., Aimon, A., et al., (2020). Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nature Commun.*, **11**, 1–11. https://doi.org/10.1038/s41467-020-18709-w.

23. Goyal, B., Goyal, D., (2020). Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy. *ACS Comb. Sci.*, **22**, 297–305. https://doi.org/10.1021/acscombsci.0c00058.

24. Rut, W., Groborz, K., Zhang, L., Sun, X., Zmudzinski, M., Pawlik, B., Wang, X., Jochmans, D., et al., (2020). SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nature Chem. Biol.*, **17**, 222–228. https://doi.org/10.1038/s41589-020-00689-z.

25. Lee, J., Worrall, L.J., Vuckovic, M., Rosell, F.I., Gentile, F., Ton, A., Caveney, N.A., Ban, F., et al., (2020). Crystallographic structure of wild-type SARS-CoV- 2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nature Commun.*, **11** https://doi.org/10.1038/s41467-020-19662-4.

26. Aslanidis, C., de Jong, P.J., (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.*, **18**, 6069–6074. https://doi.org/10.1093/nar/18.20.6069.

27. Studier, F.W., (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

28. Liu, L., Milas, N., Mukai, A.H.C., Resende, X.R., de Sá, F. H., (2014). The Sirius project. *J. Synchrotron Radiat.*, **21**, 904–911. https://doi.org/10.1107/S1600577514011928.

29. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., Bricogne, G., (2011). Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 293–302. https://doi.org/10.1107/S0907444911007773.

30. Kabsch, W. (2010). XDS, Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132. https://doi.org/10.1107/S0907444909047337.

31. Evans, P.R., Murshudov, G.N., (2013). How good are my data and what is the resolution?. *Acta Crystallogr D Biol Crystallogr.*, **69**, 1204–1214. https://doi.org/10.1107/s0907444913000061.

32. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., (2007). Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674. https://doi.org/10.1107/S0021889807021206.

33. Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., (2010). Features and development of Coot. *Acta Crystallogr. Sect. D.*, **66**, 486–501. https://doi.org/10.1107/S0907444910007493.

34. Smart, O.S., Womack, T.O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G., (2012). Exploiting structure similarity in refinement: Automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **68**, 368–380. https://doi.org/10.1107/S0907444911056058.

35. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., et al., (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D.*, **68**, 352–367. https://doi.org/10.1107/S0907444912001308.

36. Wollenhaupt, J., Metz, A., Barthel, T., Lima, G.M.A., Heine, A., Mueller, U., Klebe, G., Weiss, M.S., (2020). F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening 694–706.e5 *Structure*, **28** https://doi.org/10.1016/j.str.2020.04.019.

37. Lima, G.M.A., Talibov, V.O., Jagudin, E., Sele, C., Nyblom, M., Knecht, W., Logan, D.T., Sjögren, T., et al., (2020). FragMAX: the fragment-screening platform at the MAX IV Laboratory. *Acta Crystallogr. Sect. D, Struct. Biol.*, **76**, 771–777. https://doi.org/10.1107/S205979832000889X.

38. Lima, G.M.A., Jagudin, E., Talibov, V.O., Benz, L.S., Marullo, C., Barthel, T., Wollenhaupt, J., Weiss, M.S., et al., (2021). *FragMAXapp*: crystallographic fragment-screening data-analysis and project-management system. *Acta Crystallogr. Sect. D Struct. Biol.*, **77** https://doi.org/10.1107/S2059798321003818.

39. Moriarty, N.W., Grosse-Kunstleve, R.W., Adams, P.D., (2009). Electronic ligand builder and optimization workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **65**, 1074–1080. https://doi.org/10.1107/S0907444909029436.

40. Winter, G., Waterman, D.G., Parkhurst, J.M., Brewster, A. S., Gildea, R.J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., et al., (2018). DIALS: Implementation and evaluation of

a new integration package. *Acta Crystallogr. Sect. D Struct. Biol.*, **74**, 85–97. https://doi.org/10.1107/S2059798317017235.

41. Vonrhein, C., Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Bricogne, G., (2018). Advances in automated data analysis and processing within autoPROC combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO a360–a360 *Acta Crystallogr. Sect. A Found. Adv.*, **74** https://doi.org/10.1107/s010876731809640x.

42. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., et al., (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**, 235–242. https://doi.org/10.1107/S0907444910045749.

43. Pearce, N.M., Krojer, T., Bradley, A.R., Collins, P., Nowak, R.P., Talon, R., Marsden, B.D., Kelm, S., et al., (2017). A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nature Commun.*, **8**, 1–8. https://doi.org/10.1038/ncomms15123.

44. Liebschner, D., Afonine, P.V., Moriarty, N.W., Poon, B.K., Sobolev, O.V., Terwilliger, T.C., Adams, P.D., (2017). Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D Struct. Biol.*, **73**, 148–157. https://doi.org/10.1107/S2059798316018210.