communications biology

Article

https://doi.org/10.1038/s42003-024-07138-w

Allostery in homodimeric SARS-CoV-2 main protease

Emanuele Fornasier[1,4], Simone Fabbian[1,2,4], Haidi Shehi[1], Janine Enderle[2], Barbara Gatto[2], Daniele Volpin[1], Barbara Biondi[3], Massimo Bellanda[1,3], Gabriele Giachin[1], Alice Sosic[2] & Roberto Battistutta[1,3] ✉

Many enzymes work as homodimers with two distant catalytic sites, but the reason for this choice is often not clear. For the main protease M^{pro} of SARS-CoV-2, dimerization is essential for function and plays a regulatory role during the coronaviral replication process. Here, to analyze a possible allosteric mechanism, we use X-ray crystallography, native mass spectrometry, isothermal titration calorimetry, and activity assays to study the interaction of M^{pro} with three peptide substrates. Crystal structures show how the plasticity of M^{pro} is exploited to face differences in the sequences of the natural substrates. Importantly, unlike in the free form, the M^{pro} dimer in complex with these peptides is asymmetric and the structures of the substrates nsp5/6 and nsp14/15 bound to a single subunit show allosteric communications between active sites. We identified arginines 4 and 298 as key elements in the transition from symmetric to asymmetric dimers. Kinetic data allowed the identification of positive cooperativity based on the increase in the processing efficiency (kinetic allostery) and not on the better binding of the substrates (thermodynamic allostery). At the physiological level, this allosteric behavior may be justified by the need to regulate the processing of viral polyproteins in time and space.

The SARS-CoV-2 main protease M^{pro}, also known as 3C-like protease $3CL^{pro}$ or nsp5, is a cysteine peptidase crucial for the coronavirus replication cycle[1]. Although dimerization has been recognized as essential and has been proposed to play a direct regulatory role for M^{pro} activity during the coronaviral replication process[2], it is not clear what the advantage is for the enzyme to work as a dimer instead of a monomer (like most of the related proteases); as well as whether the two active sites, not in physical direct contact, can communicate, with some sort of allostery. This is a general issue, valid for many homodimeric enzymes with two distant catalytic sites[3].

The interest in SARS-CoV-2 M^{pro} is based on the fact that it is responsible for the proteolytic processing of the two overlapping polyproteins pp1a and pp1ab during replication, with the final formation of individual mature non-structural proteins nsp, from nsp4 to nsp16[4], and, as such, is a validated antiviral drug target[5–10]. In addition to cleaving viral polyproteins, it has been shown that M^{pro} can interfere with the normal functions of the human immune system. For example, it can attenuate immune defense of antiviral effectors encoded by interferon-stimulated genes by cleaving host histone deacetylases and the mRNA-decapping enzyme 1a (DCP1A)[11]. In 2021, the first M^{pro} inhibitor with antiviral activity in a transgenic mouse model was reported[12]. Currently, there is a SARS-CoV-2 M^{pro} inhibitor on the market with antiviral activity, the orally administered PF-07321332 or Nirmatrelvir[13].

SARS-CoV-2 M^{pro}, a 306 amino acid long polypeptide, exists in equilibrium between the monomer and the homodimer, with an apparent K_d estimated at 2.5 μM and 1.32 μM[2,14,15] by analytical ultracentrifugation. Each M^{pro} protomer is composed of three structural domains: chymotrypsin- and 3C protease-like β-barrel domains I (residues 1–99) and II (residues 100–182), which directly control the catalytic event, and the extra α-helical domain III (residues 198–306), absent in other 3C-like proteases of RNA viruses, which plays a key role in enzyme dimerization and activity regulation[2,16]. Instead of the typical catalytic triad of chymotrypsin-like proteases, M^{pro} has a catalytic dyad composed of residues His41 and Cys145. A structural element essential for the catalytic event is the "oxyanion loop"[2,17–19], very mobile, which can also adopt a stable inactive conformation[20].

To generate non-structural proteins, M^{pro} cleaves the polyproteins pp1a and pp1ab at 11 different sites. The binding site is located between domains I and II and comprises several subsites, from S1 to S6 and from S1′ to S4′, corresponding to the amino acid positions P1-P6 and P1'-P4' of the substrates, according to the convention P6-P5-P4-P3-P2-P1↓P1'-P2'-P3'-P4',

[1]Department of Chemical Sciences, University of Padova, via F. Marzolo 1, 35131 Padova, Italy. [2]Department of Pharmaceutical and Pharmacological Sciences, University of Padova, via F. Marzolo 5, 35131 Padova, Italy. [3]Institute of Biomolecular Chemistry of CNR, Padova Unit, via F. Marzolo 1, 35131 Padova, Italy. [4]These authors contributed equally: Emanuele Fornasier, Simone Fabbian. ✉e-mail: roberto.battistutta@unipd.it

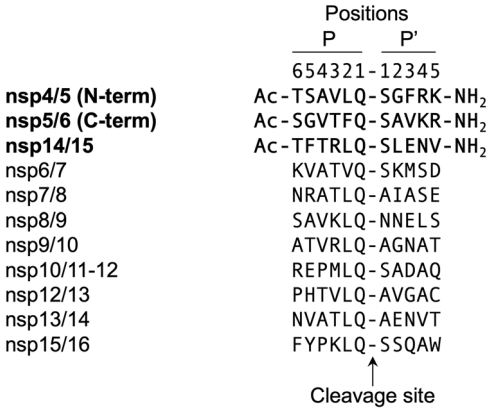

Fig. 1 | M^{pro} substrate peptides. Sequences of the peptides synthesized and studied in this work are shown in bold. The other sequences processed by M^{pro} are indicated for comparison[22]. The conventional terminology for the positions of the residues in the substrate peptides (P and P' positions) is indicated.

with ↓ indicating the hydrolyzed peptide bond[21]. The cleavage of the substrate occurs at the C-terminal end of a conserved glutamine (P1), and the most common consensus sequence is Leu-Gln↓(Ser/Ala). Interestingly, the recognition sites at the C-terminus of P1' are not conserved[22].

Some aspects of the structural bases of substrates recognition by M^{pro} have already been analyzed[22–26]. Here, we integrate and extent such pieces of information by reporting the structural, by X-ray crystallography, biophysical, by native mass spectrometry (nMS) and isothermal titration calorimetry (ITC), and kinetic characterization of the interaction of M^{pro} with 11-mer peptides mimicking selected substrates, namely nsp4/5, nsp5/6 and nsp14/15. We show that the initial binding to one active site of the dimeric enzyme allosterically perturbs the second one and generates an asymmetric dimer. The mobilization of the second site has the role to favor the first binding by entropy compensation. A thermodynamic characterization shows the presence of a moderate negative or positive cooperativity in the binding between the two active sites, depending on the substrate. However, activity measurements reveal the presence of an overall significant positive cooperativity with kinetic basis, i.e. due to the kinetic constants rather than the affinity constants.

Results

Inactive M^{pro} and substrate peptides

To study the interaction of M^{pro} with peptides mimicking different substrates ("substrate peptides"), we produced an inactive form of the enzyme mutating the two catalytic residues His41 and Cys145 with alanine, generating an enzymatically inactive double mutant, from now on simply called M^{pro} (or "inactive mutant of M^{pro}" when relevant). To avoid ambiguities, the term "wtM^{pro}" is sometimes used to refer to the wild type protein.

As substrates, we choose those corresponding to the cleavage sites at the extremities of M^{pro}, nsp4/5 and nsp5/6, and the cleavage site between non-structural proteins 14 and 15, nsp14/15, one of the most dissimilar in sequence when compared to the other substrates, in particular in positions P5, P4, P3, P2', P3', P4', and P5' (Fig. 1). We synthetized the three substrate peptides by Solid Phase Peptide Synthesis (Supplementary Note 1). Peptides are made of 11 residues (11-mer), from position P6 to position P5', to encompass the entire binding region (Fig. 1). They were capped at the extremities, by acetylation at the N-terminus and amidation at the C-terminus, to mimic the peptide bonds and make the peptides appear more like native protein.

Structure of the symmetric free form dimer of M^{pro}

The structure of the ligand free (free form) inactive mutant of M^{pro} studied here and the structure of the free form of the wild-type enzyme (we take the PDB code 6Y2E as reference[2]) show the same crystal packing, that is, the same space group (the monoclinic C2), and very similar geometric axis and angles of the unit cell. Statistics on data collection and model refinement are reported in Table 1. The $2F_o$-F_c electron density map of the binding site of the M^{pro} inactive mutant His41 Ala and Cys145Ala is shown in Supplementary Fig. 1A; the lack of the electron density corresponding to the side chains demonstrates the presence of both mutations. With one monomer per asymmetric unit, the dimer is formed by two identical subunits connected by a crystallographic two-fold axis, therefore it is perfectly symmetrical, like the vast majority of the more than one thousand M^{pro} structures in the PDB in the C2 space group. The comparison between the two structures shows that the two mutations Cys145Ala and His41Ala do not introduce significant alterations at the substrate binding site, included the oxyanion loop, and, in general, in the entire protein structure (total root-mean-square deviation, rmsd, 0.310 Å, Fig. 2). The "connecting region" (residues 190–198, known in fact as a flexible region), linking the second and third domains of M^{pro}, shows the highest variations, with deviations of Cα atoms between 0.4 and 1.7 Å. The water molecule bound to His41 in wtM^{pro}, believed to be involved in catalysis, is conserved in inactive M^{pro}.

Structure of the asymmetric dimer of M^{pro} in complex with nsp4/5 (by cocrystallization)

First, we examined the complex between M^{pro} and the 11-mer peptide nsp4/5, representing the N-terminal autocleavage site, the first hydrolyzed in the maturation process of the enzyme. Peptides, of variable lengths, used in enzymatic essays in literature are commonly based on such substrate sequence. In a cocrystallisation experiment, the M^{pro}-nsp4/5 complex crystallizes in a different space group, the triclinic P1, with two molecules (that is, the full dimer) in the asymmetric unit. Statistics on data collection and model refinement are reported in Table 1. The Polder F_o-F_c omit map of this peptide substrate bound to the active site, as well as those of nsp5/6 and nsp14/15, are reported in Supplementary Fig. 1B. Both active sites are occupied by the peptide, which binds with small but significant differences, especially at the extremities. Binding of the substrate peptide nsp4/5 to M^{pro} is illustrated in Figs. 3 and 4, and the details of the interactions are illustrated in the Supplementary Note 2 and in Supplementary Figs. 2 and 3. Despite substantial conservation in the binding mode of the substrate to the two subunits, the dimer is now asymmetric, with clear differences between the two chains (Fig. 3). First, in mobility, as shown by analysis of B-factors, especially in domain I facing the binding site; second, in some structural details, mainly in residues 44–50 (forming a short helix in free M^{pro}), which in subunit B is more shifted away from the peptide, with the Cα of Asp48 1.45 A apart (Fig. 3, panels E and F). Another relevant difference between the two subunits is at the level of the C-terminal tail, from residue 300 on. In one subunit (B), visible until 303, residues 301–303 are placed between the two subunits, as in free wtM^{pro}; in the other subunit (A), visible until 302, the last two residues point outward, toward the solvent (Fig. 3, panels A, B, E, and F, and Fig. 4). Note that these regions are not in direct contact with the substrate.

In subunit A (Supplementary Fig. 2), considering the hydrogen bonds, the substrate backbone makes 10 direct interactions (3 only for Gln P1) and 3 water-mediated ones (in P4, P5, and P6), and only the side chains of Gln P1 and Arg P4' are strongly interacting with the protein, with three H-bonds each. Regarding the hydrophobic residues, Leu P2 and Ala P4 fit hydrophobic cavities, while Val P3 and Phe P3' are exposed (and mobile). Overall, residues that contribute more strongly to the binding are those at positions P1, P2, P3, and P4 and, at the C-terminus of Gln P1, P2', and P4'. The specificity of the binding, determined by the nature of the side chains, depends mainly on positions P1, P2, P4 and P4'. In subunit B (Supplementary Fig. 3), the substrate is bound in a similar way, making essentially the same interactions with the protein. Differences are visible principally at the extremities, at the level of Thr P6, slightly shifted, Arg P4' and Lys P5', now not visible in the electron density. The analysis reported above clearly indicates that in the crystal structure of M^{pro} in complex with the substrate peptide nsp4/5, the absence of a two-fold crystallographic axis connecting the dimer (as seen

Table 1 | Data collection and refinement statistics

PDB code	9EX8	9EXU	9EYA	9EZ4	9EZ6
Data collection					
X-ray source	ESRF ID23-1				
Wavelength (Å)	0.8856				
Sample	Free form	nsp4/5 cocryst.	nsp4/5 soak	nsp5/6 cocryst.	nsp14/15 cocryst.
Space group	C2	P1	C2	$P2_12_12_1$	$P2_12_12_1$
N° in ASU	1	2	1	2	2
Cell dimensions					
a, b, c (Å)	114.35, 53.24, 44.70	53.56, 61.36, 67.90	113.69, 52.19, 45.15	67.75, 99.04, 101.99	67.80, 99.03, 100.84
α, β, γ (°)	90.00, 102.91, 90.00	92.22, 109.10, 108.38	90.00, 103.31, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	48.04–1.80 (1.84–1.80)	46.88–1.78 (1.81–1.78)	47.20–1.70 (1.73–1.70)	49.52–1.80 (1.84–1.80)	55.95–1.87 (1.91–1.87)
R_{merge}	0.076 (1.502)	0.038 (0.993)	0.068 (1.103)	0.077 (1.461)	0.057 (0.907)
R_{meas}	0.083 (1.625)	0.045 (1.195)	0.075 (1.278)	0.081 (1.726)	0.067 (1.064)
R_{pim}	0.031 (0.609)	0.024 (0.652)	0.031 (0.633)	0.025 (0.603)	0.034 (0.547)
N° observations	170902 (8756)	240323 (7094)	146742 (3491)	668786 (25605)	213122 (12099)
N° unique	24239 (1303)	68784 (2308)	26488 (911)	64002 (3564)	55875 (3408)
Mean(I)/σ(I)	11.3 (1.1)	13.8 (0.9)	11.9 (1.1)	16.1 (1.0)	10.5 (1.2)
$CC_{1/2}$	0.999 (0.495)	0.999 (0.665)	0.998 (0.487)	0.998 (0.485)	0.998 (0.536)
Completeness (%)	99.4 (91.4)	93.1 (55.0)	93.3 (60.1)	99.5 (95.2)	98.4 (95.2)
Multiplicity	7.1 (6.7)	3.5 (3.1)	5.5 (3.8)	10.4 (7.2)	3.8 (3.6)
B Wilson ($Å^2$)	36.0	38.47	28.91	32.30	37.98
Refinement					
Resolution (Å)	43.57–1.80	46.88–1.78	47.20–1.70	44.54–1.80	55.95–1.87
R_{work}/R_{free} No. reflections	23022/1217	65384/3400	25219/1269	60830/3172	53133/2742
R_{work}/R_{free} (%)	19.3/21.1	17.9/20.4	15.9/18.6	18.3/21.5	19.5/22.8
Number of atoms					
Protein	2365	4651	2376	4675	4710
Water	163	331	302	325	185
Average B, all atoms ($Å^2$)	39.9	50.7	30.3	40.4	51.2
R.m.s. deviations					
Bond lengths (Å)	0.007	0.018	0.004	0.013	0.007
Bond angles (°)	0.695	1.481	0.721	1.295	0.732
Ramachandran statistics					
Favored (%)	97.69	98.06	97.44	97.82	96.31
Allowed (%)	1.98	1.94	2.56	2.01	3.52
Outliers (%)	0.33	0	0	0.17	0.17

Values in parentheses are for highest-resolution shell.

in free wtM^{pro}) is due to the differences between subunits A and B induced by substrate binding, with the consequent formation of an "asymmetric dimer".

Focusing on the binding site (Fig. 4), comparison with the structure of M^{pro} in the free form reveals that the binding of the substrate peptide induces variations in the following regions: β-turn 22–26 due to the interaction with the P' positions (mainly P2' and P4') of the substrate; connecting region 187–199, residues 166–172 (part of the β-hairpin loop), and residues 44–50 due to the binding of the N-terminal part of the substrate (P positions). The most relevant is the shift of residues 44–50, mainly caused by the insertion of Leu P2 into the subpocket S2 that causes a shift of Met49 and, consequently, of the entire region residues, with deviations of Cα atoms as high as 2.5 Å compared to M^{pro} (Fig. 4, panel B). Instead, the area of interaction of Gln P1, including the oxyanion loop, is well conserved between the two structures.

Structure of the symmetric dimer of M^{pro} in complex with nsp4/5 (by soaking)

Many known complexes of M^{pro} with peptides substrates were obtained by soaking the free enzyme crystallized in the C2 space group, with one molecule in the ASU. Clearly, in these cases, the two subunits of the dimeric M^{pro} complex are perfectly identical, being related by a crystallographic two-fold axis. Therefore, the structure of the complex of the substrate peptide nsp4/5 was also obtained by soaking M^{pro} previously crystallized in free form in the common C2 space group. Statistics on data collection and model refinement are reported in Table 1. The Polder F_o-F_c omit map of the peptide substrate bound to the active site is reported in Supplementary Fig. 1B.

Both the substrate peptide and the protein show differences compared to the cocrystallized structure (Fig. 5). Regarding the peptide, although the hydrogen bonds involving the backbone are conserved, only Gln P1 and Ser

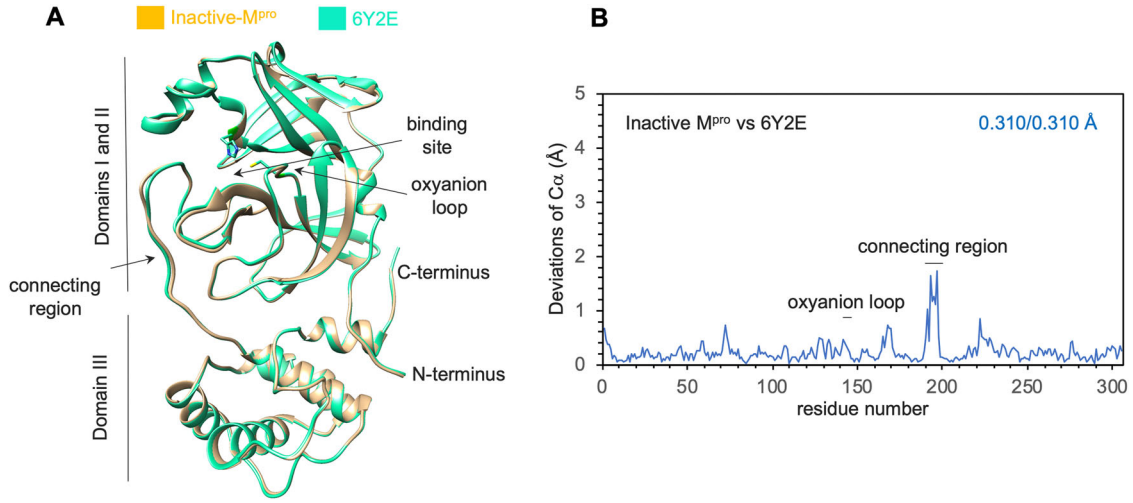

Fig. 2 | Structure of the free form of the inactive mutant of M^{pro} (symmetric dimer). A The superposition of our inactive mutant of M^{pro} and wtM^{pro} (6Y2E) shows how the two mutations Cys145Ala and His41Ala did not introduce significant alterations in the structure of the enzyme. His41 and Cys145 in the active site of 6Y2E wtM^{pro} are shown in stick. **B** Plot of the deviations of the Cα atoms versus the residue number for inactive M^{pro} (this study) superposed to active wtM^{pro} (PDB 6Y2E); total rmsd values (in Å) are indicated for the best aligned and total atom pairs, respectively. The region connecting the second and third domains of M^{pro}, residues 190-198 ("connecting region"), shows the highest variations. The structure of the oxyanion loop and other relevant areas surrounding the active site are conserved.

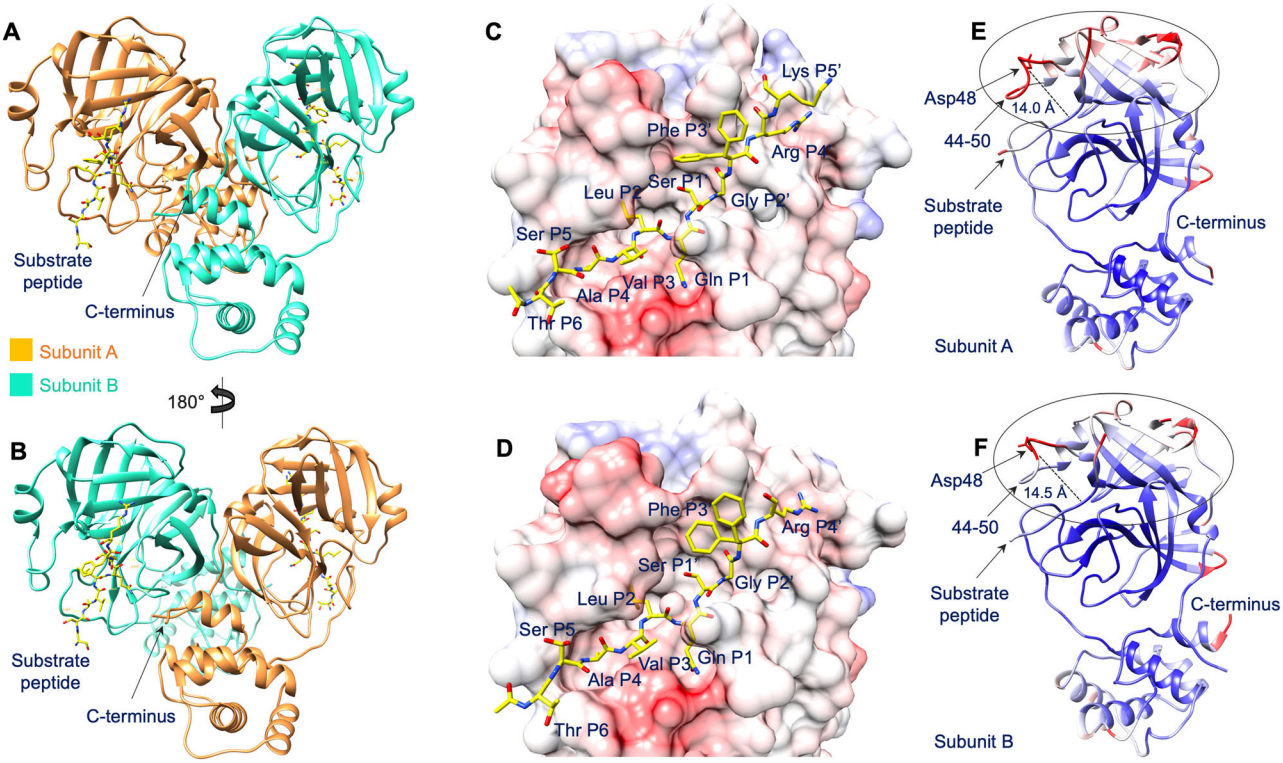

Fig. 3 | Binding of the substrate peptide nsp4/5 induces the formation of an asymmetric dimer. A, **B** The substrate peptide nsp4/5 (in sticks) bound to the two active sites of dimeric M^{pro}. **C**, **D** Close-up views of the substrate peptide nsp4/5 at the active site of subunits A and B, respectively. The surface is colored according to the electrostatic potential. Note the differences in the residues at the extremities, in particular P6 and P5' (missed in nsp4/5 bound to chain B). **E**, **F** Subunits A and B, respectively, colored according to B-factors (from 25 Å^2 blue to 75 Å^2 red; average B for all atoms 50.7 Å^2). The major differences are at domain I (circled) facing the binding site, in particular at residues 44–50. Asp48 (in stick) and its distance to Gln P1 (dashed line, between Cα atoms) are indicated. In (**A**, **B**, **E**, **F**) the C-terminal ends are indicated to underline their differences in the two subunits.

P1' are well superposed, with identical interactions also for the side chains (Fig. 5, panel B). Leu P2 has an identical Cα position but the side chain rotamer is different. For the P positions, the backbone from P2 to P6 is shifted, particularly for Thr P6. The P' positions are progressively differentiating, remarkably, with Arg P4' in a completely different position, with the side chain now anchored to the CO of Gly23 instead of the side chains of Thr24 and Gln69. The result is that in this structure, the backbone conformation of the ten residues is more bent. These differences in the

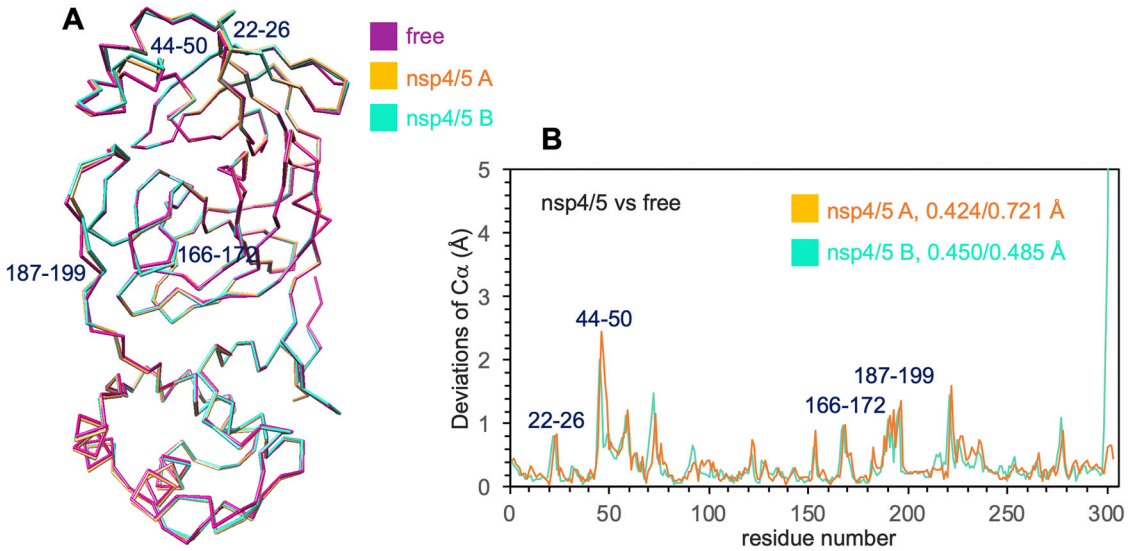

Fig. 4 | Binding of the substrate peptide nsp4/5 induces structural modifications of the active site. A Superposition of chains A and B of the nsp4/5-M^{pro} complex to the free form of the enzyme. The highest variations induced by binding are near the active site, and locate at β-turn 22–26, residues 166–172 (part of the β-hairpin loop), connecting region 187–199 and, mainly, residues 44–50. **B** Plot of the deviations of the Cα atoms versus the residue number of the two subunits of the nsp4/5-M^{pro} complex superposed to free M^{pro}; total rmsd values (in Å) for the best aligned and total atom pairs, respectively, are indicated.

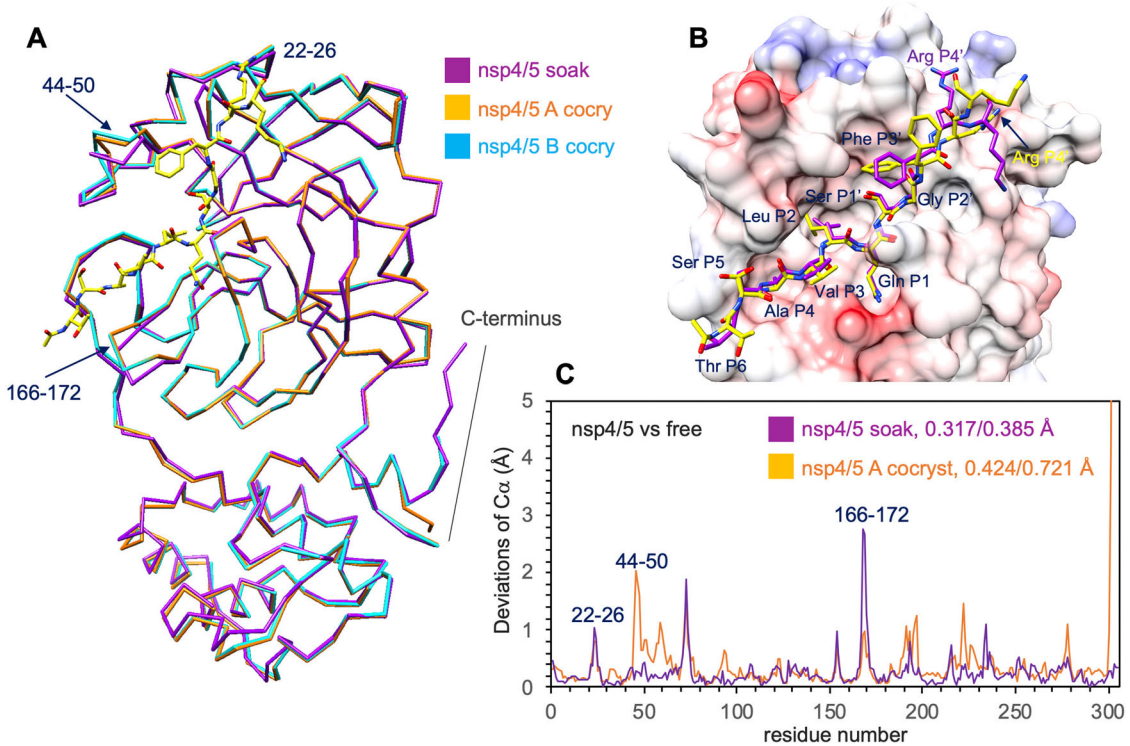

Fig. 5 | Comparison between soaked and cocrystallized complexes of substrate peptide nsp4/5. A Superposition of the peptide substrate nsp4/5 complexes for the soaked and cocrystallized structures. The C-terminal tail of the nsp4/5 soaked structure is similar to the free enzyme and to subunit B in the cocrystallized structure, but different from subunit A in the cocrystallized structure. **B** Superposition of nsp4/5 in the soaked and cocrystallized structures; surface colored according to electrostatics corresponds to the soaked structure. Note that Arg P4' is in very different positions. **C** Plot of the deviations of the Cα atoms versus the residue number of cocrystallized and soaked nsp4/5-M^{pro} complexes, superposed to free M^{pro}; total rmsd values (in Å) are indicated for the best aligned and total atom pairs, respectively.

conformation of the substrate correspond to a binding site more similar to the free form of the enzyme, indicating that, in this case, the crystallographic constraints present in the soaking procedure contrast the structural variations seen in the cocrystallized structure. Although differences may occur between structures obtained by cocrystallization or soaking experiments, this is not obvious and is not always the case; it can depend on many factors, the nature and size of the ligand, the nature of the interaction, the type of protein and its flexibility, the crystal packing, and so on. For the protein matrix, again there are movements in regions 22–25, 187–199, and 166–172 (Fig. 5, panel C) but they are different. Remarkably, region 44–50 is now

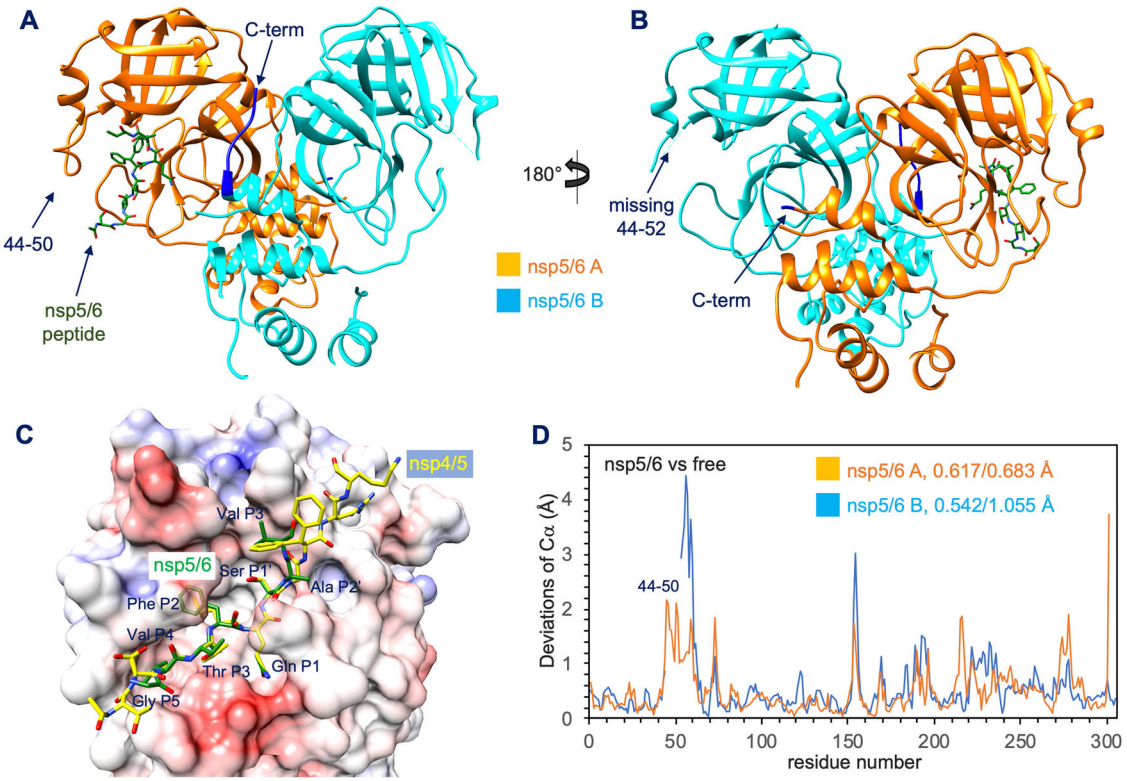

Fig. 6 | The substrate peptide nsp5/6 binds to only one subunit. A, **B** The substrate peptide nsp5/6 (green carbon atoms, in sticks) bound to only one subunit (**A**) of dimeric M^{pro}; the C-terminal tail of subunit A is disordered in solution. In subunit B, the active site is empty, residues 44–52 are missing, and the C-terminal tail is in the canonical position in the inter-subunit space, near the binding site of A. **C** superposition of substrate peptides nsp4/5 and nsp5/6; the backbone is similar from position P3 to position P2'. Lys P4' and Arg P5' are flexible and were not modeled in nsp5/6. Surface colored according to electrostatics corresponds to the nsp5/6 complex. **D** Plot of the deviations of the Cα atoms versus the residue number of the two subunits of the nsp5/6-M^{pro} complex superposed on free M^{pro}; total rmsd values (in Å) are indicated for best aligned and total atom pairs, respectively.

similar to the free structure, hence different compared to the cocrystallized structure. Furthermore, the NH of Leu P2 is not directly interacting with the Gln189 side chain, which shows a different conformation; rather, this interaction is now mediated by a water molecule. It appears that, in the soaking experiment, Leu P2 is unable to induce the reshaping of subsite S2 and, in particular, to alter the region 42–52 through interactions with Met49 due to opposing crystalline constraints. Analyzing other parts of the protein structure, the C-terminal tail of the enzyme, visible until residue 306, shows a conformation similar to that of the free M^{pro}, without the movements appreciated in the cocrystallized structure (Fig. 5, panel A). Again, in the soaking experiment, the crystallographic restraints prevented these structural modifications, as was seen for the cocrystallized complex.

Structure of the M^{pro} dimer with nsp5/6 in a single binding site

Given the possible artefacts of structures determined by soaking of peptide substrates on M^{pro} crystals in free form (typically in the space group C2), with the other substrate peptides under study we decided to perform only cocrystallization experiments. M^{pro} in complex with the substrate peptide nsp5/6 crystallizes in a third space group, orthorhombic $P2_12_12_1$, with two molecules in the asymmetric unit, as in the case of the substrate peptide nsp4/5. Statistics on data collection and model refinement are reported in Table 1. In this case, significantly, the active site of only one subunit (subunit A) is now occupied by the substrate, whereas the other active site (subunit B) is empty (Fig. 6A, B). The Polder F_o-F_c omit map of the peptide substrate bound to the active site is reported in Supplementary Fig. 1B.

The details of the interactions are illustrated in the Supplementary Note 3 and in Supplementary Fig. 4. The way in which the substrate peptide nsp5/6 binds to M^{pro} shows significant differences compared to nsp4/5 (panel C of Fig. 6 and Supplementary Figs. 2, 3 and 4). The only well superposed residues are Gln P1 and Ser P1', whose peptide bond is the scissile bond, which conserve the interactions with the protein. In general, the binding appears weaker, with the extremities poorly or without interacting, and the loss of two (out of 10) hydrogen bonds involving the backbone in position P4 and P4', as well as those involving the guanidine of Arg P4', substituted by a lysine. Substitutions of alanine by valine at position P4 and arginine by lysine at position P4' seem the main determinant of the decrease in affinity, as showed by nMS and ITC (see below).

In subunit A, binding of the substrate peptide nsp5/6 causes relevant modifications in region 44–50 (Fig. 6, panel D). Notably, there are no relevant perturbations in the region 22–26, in accordance with the lack of interaction with the P4' position of the substrate.

Surprisingly, higher perturbations are observed in subunit B, whose active site is empty, than in subunit A, where binding occurs. In subunit B, the electron density of residues 44–52 is not visible, indicating very high mobility, and the rmsd of the subsequent region up to position 60 is high, up to 4.5 Å (Fig. 6, panel D). This is not the only region that is different between the two subunits. In fact, the rmsd values between them are 0.713 and 1.152 Å (for the best 281 over a total of 294 atom pairs, respectively). This indicates that binding to a single subunit is capable of generating structural differences between the two subunits of the dimer to an extent that has not been observed before. Similarly to free M^{pro} and the cocrystallized nsp4/5 complex, now the C-terminal tail in subunit B, facing the active site occupied of subunit A, places between the two subunits (Fig. 6, panel A). Instead, the C-terminal tail of subunit A, facing the empty active site of subunit B, points outward toward the solvent (Fig. 6, panel B).

Overall, the binding to only one subunit (A in this case) perturbs the other active site (in subunit B), in particular, inducing a higher mobility not only in residues of B facing its active site but also in the C-terminus of subunit A nearby the active site of subunit B.

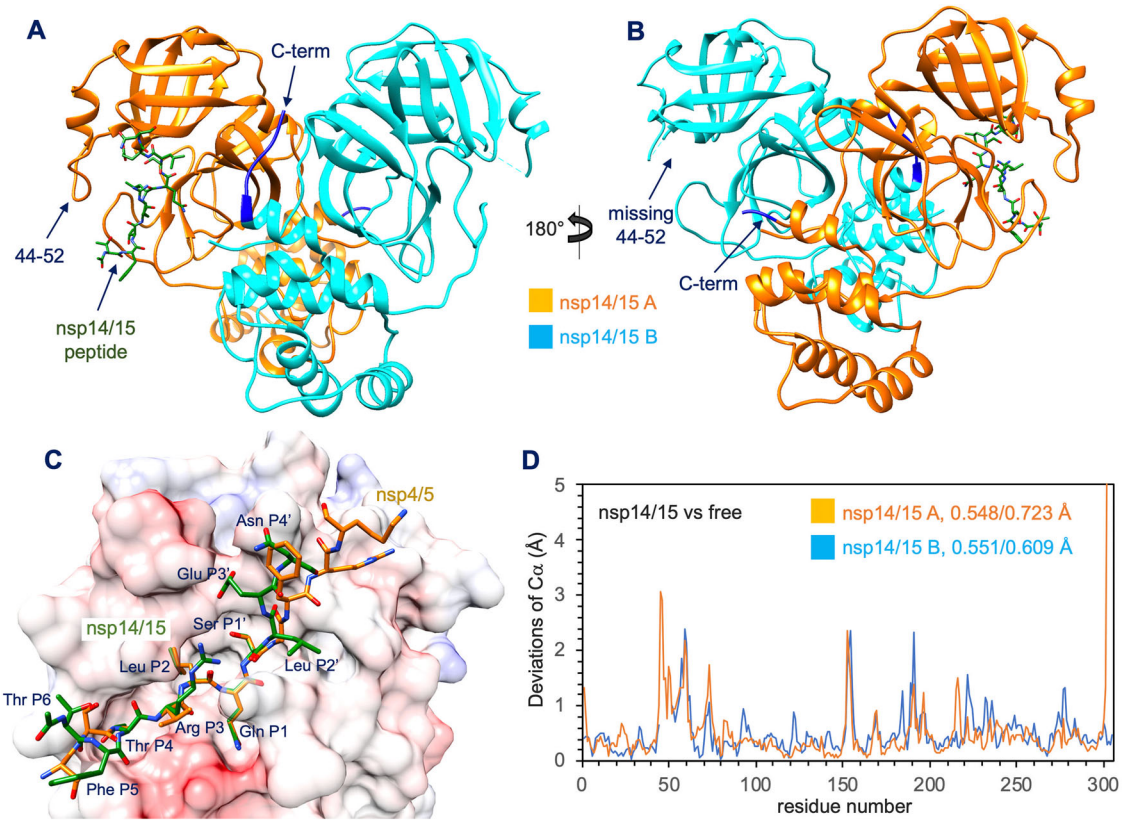

Fig. 7 | As nsp5/6, the substrate peptide nsp14/15 binds to only one subunit. **A**, **B** The substrate peptide nsp14/15 (green carbon atoms, in sticks) binds to only one subunit (A) of the dimeric M^{pro}, similarly to nsp5/6. The C-terminal tail of subunit A is disordered in solution. In subunit B, the active site is empty, residues 44–52 are missing and the C-terminal tail is in the canonical position in the inter-subunit space, near the binding site of (**A**). **C** Superposition of nsp14/15 and nsp4/5; the backbone is similar from position P4 to P2', while positions P5, P6, P3' and P4' are different. The surface of the nsp14/15 complex colored according to electrostatics is shown. **D** Plot of the deviations of the Cα atoms versus the residue number of the two subunits of the nsp14/15-M^{pro} complex, superposed on free M^{pro}; total rmsd values (in Å) are indicated for best aligned and total atom pairs, respectively.

Structure of the M^{pro} dimer with nsp14/15 in a single binding site

As in the case of the substrate peptide nsp5/6, the nsp14/15/M^{pro} complex crystallizes in orthorhombic SG $P2_12_12_1$ (with the same crystal cell dimensions), with two molecules in the asymmetric unit and, remarkably, with the active site of only one subunit (subunit A) occupied by the substrate (Fig. 7, panels A and B). The binding mode is analogous to that of nsp5/6, with the backbone similar from position P3 to P3'. Statistics on data collection and model refinement are reported in Table 1. The Polder F_o-F_c omit map of the peptide substrate bound to the active site is reported in Supplementary Fig. 1B. The details of the interactions are illustrated in the Supplementary Note 4 and in Supplementary Fig. 5.

Among the three substrates studied, this is the only one with a clear interaction between position P6 and the enzyme; it appears to be due to the presence of threonines P4 and P6. Compared to nsp4/5, Thr P4 pushes apart the peptide from the protein, as in nsp5/6, but, unlike the last, Thr P6 (instead of a serine) can fold back the chain due to the side chain interaction (Fig. 7, panel C). The position P4' is important for the anchoring (or not) of the substrate to the enzyme; the anchoring is possible for nsp4/5 (via an arginine), and for nsp14/15 (via a glutamine) but not for nsp5/6 (with a flexible lysine not visible in the electron density). However, the mode of interaction of this region in nsp4/5 and nsp14/15 is different, due to the different properties of the residues involved (arginine and glutamine), but also due to the nature of the residue at position P3', a mobile phenylalanine in nsp4/5 and a fixed glutamate in nsp14/15.

The perturbation of the protein matrix by nsp14/15 is similar to that induced by nsp5/6, with the binding to one subunit able to induce greater mobility of the active site of the other (Fig. 7, panel D). Important structural modifications are seen in region 44–52, especially in subunit B with the empty active site, where the electron density is not visible, and the inter-subunit deviations of the Cα atoms of the subsequent region up to position 60 is as high as 2.5 Å. Again, the C-terminal tail of subunit B, facing the occupied active site of subunit A, is between the two subunits; instead, the C-terminal tail of subunit A, facing the empty active site of subunit B, is pointing outward toward the solvent (Fig. 7, panels A and B).

For peptide substrates nsp5/6 and nsp14/15 only the binding site of an M^{pro} subunit is occupied because of their low solubility (compared to nsp4/5) and their low affinity for M^{pro} (as shown by nMS measurements).

Arginines 4 and 298 are key elements in the transition from symmetric to asymmetric dimer

All crystal structures presented here were obtained from crystals growth with the same protocol (see experimental section); therefore, structural differences discussed in the text are to be considered genuine responses to the binding of different substrate peptides rather than to different experimental crystallization conditions. Our crystal structures show a transition from a symmetric architecture of the free M^{pro} dimer to an asymmetric one upon substrate binding, particularly evident when only one binding site is occupied. The subunit interface is made up of symmetrical residues related by the two-fold axis. The interface has an important contribution from the electrostatic forces, with many positive and negative residues in close proximity, making intricate networks of electrostatic interactions (Supplementary Fig. 6). The importance of electrostatic interactions in the activity of dimeric M^{pro} is supported by the observation that enzymatic activity decreases with increasing salt concentrations[27].

In the symmetric dimer, these networks are perfectly symmetric around the two-fold axis. In the main cluster Arg4 from one subunit inserts

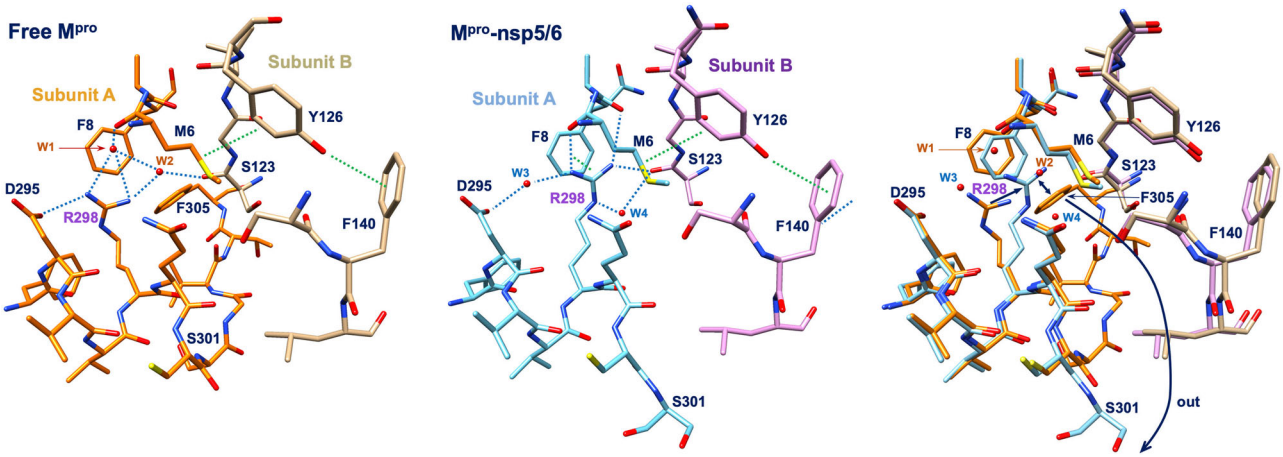

Fig. 8 | Arg298 movement connected to the first binding. In the free form of M^{pro}, Arg298 interacts mainly with Asp295 from the same subunit, as shown in the left panel. As a consequence of the first binding to subunit A (nsp5/6 in the central panel), Arg298-A moves away from Asp295-A towards Met6-A and Ser123-B, making a stacking interaction with Phe8-A. The stacking interaction of aromatic residues Phe8-A, Met6-A, Tyr126-B and Phe140-B, connecting Arg298-A to the active site of subunit B, are indicated with green dotted lines. Main hydrogen bonds are indicated with blue dotted lines. The superposition of the two structures with the underlined main movements around Arg298-A are shown in the right panel. Electron density maps are reported in Supplementary Fig. 7.

into a network of charged residues of the other subunit formed by Lys5, Arg131, Lys137, Asp197, Glu288, Asp289 and Glu290, four a total of 4 positive and 4 negative charges (Supplementary Fig. 6A). Intriguingly, positive residues belong to the first two domains, the so-called chymotrypsin fold, while negative residues belong to domain III, which is responsible for the dimerization. Other important salt bridges involve Lys12/Glu14 and Asp295/Arg298 in the same subunit (Supplementary Fig. 6B). Most of these residues have been found to be important for the catalytic event and/or the stability of the dimer[2,14,28,29].

The symmetric arrangement of these electrostatic clusters is perturbed by substrate binding, leading to the formation of asymmetric dimers. In particular, the nsp5/6 and nsp14/15 complexes (with only the binding site of subunit A occupied) show that upon the first binding there is a perturbation of the C-terminal tail of subunit A, with an associated rotation in the side chain of Arg298 which can now occupy the space freed by the exit of the subsequent residues, in particular that of Phe305 (Fig. 8).

Arg298-A moves away from Asp295-A toward Met6-A and Ser123-B, making a stacking interaction with Phe8-A. This movement can be transmitted to the other active site because Met6-A is connected through Tyr126-B to Phe140-B, which is an important element of the oxyanion loop. The electron density of the CG and CD atoms of Arg298 is weak, and this is also found in other structures such as 6Y2E and 7T70; this indicates a certain degree of flexibility of the side chain of this residue, and the ability to adopt different conformations, most likely related to the global conformational changes that M^{pro} is capable of. The conformation of Arg298-A found in the nsp5/6 and nsp14/15 complexes is maintained in the structure of the nsp4/5 complex (importantly, in a different space group), when both binding sites are occupied. In this last structure, another movement is seen, that of Arg4 of subunit A that is shifting away from Glu290-B to interact with Gln127-B and Lys137-B (Fig. 9). This shift is not seen for Arg4-B. This perturbation can be attributed to binding to the second active site because it is present only in the structure of the doubly bound nsp4/5 complex.

To our knowledge, Arg4 and Arg298 are not mutated in any known variant of SARS-CoV-2 M^{pro} and are not indicated as "variants of concern" (VOC). Furthermore, they are not (yet) designated as "resistance mutations". The two residues are conserved in SARS-CoV but not in MERS M^{pro}, where substitutions Arg4Val and Arg298Met are present, suggesting that the two enzymes can operate with partially different mechanisms. Accordingly, biochemical and AUC studies indicated that MERS M^{pro} has almost the same proteolytic activity as SARS-CoV M^{pro} although it has a higher dissociation constant in aqueous buffer[30].

Thermodynamics of the interaction between substrate peptides and M^{pro}

The crystal structures of the complexes show that binding to the first active site increases the mobility of the second one, indicating allosteric communication between them. Crystal structures with single-ligand occupancy of dimeric M^{pro} were recently obtained in the presence of 1:1 ligand/protomer stoichiometry with covalent inhibitors. Molecular dynamics simulations suggested that the empty binding site exhibits a catalytically inactive geometry, and then it was proposed that M^{pro} is regulated by negative cooperativity[31]. To shed light on this important point, we first investigated the thermodynamics of the binding to the two active sites by determining the relative dissociation constants using two different techniques, nMS and ITC. Importantly, for data consistency, we used the same 11-mer peptides utilized in the crystallographic experiments.

Native mass spectrometry (nMS) measurements

We employed electrospray ionization mass spectrometry (ESI-MS) under native conditions to obtain information about the composition and stoichiometry of non-covalent substrate-enzyme complexes in solution. Initial determinations were made in the absence of substrate in an aqueous buffer containing 150 mM ammonium acetate (pH = 7.4). In particular, nMS experiments were performed at a monomer concentration of 8.5 μM, higher than the dimer dissociation constant. Consistently, we detected only the charge state distribution corresponding to the free M^{pro} dimer (indicated with D). In fact, as shown in panel A of Fig. 10, the high abundance of signals centered on the 18^+ charge state corresponds to an experimental mass of 67396.84 Da, consistent with the expected sequence mass of dimeric doubly mutated M^{pro} (67397.04 Da). These data confirm the large prevalence of the dimeric form over the monomeric one under these experimental conditions (in the detection limit of the technique). Representative spectra obtained from mixtures of M^{pro} with nsp4/5 at different molar ration are reported in panel B of Fig. 10, which displays the region containing only the 17^+ charge state for the sake of clarity. The signal provided by free M^{pro} was readily distinguished from those corresponding to its 1:1 (indicated with DL) and 1:2 (indicated with DL_2) complexes. The species with experimental mass 68630.90 Da and 69864.96 Da displayed an increase over the mass of the free M^{pro} dimer of 1234.06 Da and 2468.12 Da, consistent with the formation of the non-covalent M^{pro} complexes containing 1 and 2 equivalents of the peptide nsp4/5, respectively. The spectra obtained from samples of M^{pro} with increasing concentrations of nsp5/6 and nsp14/15 also contained signals corresponding to free unbound dimeric M^{pro}, as well as stable complexes with stoichiometries up to 2:1 (Supplementary Fig. 9). This validated the

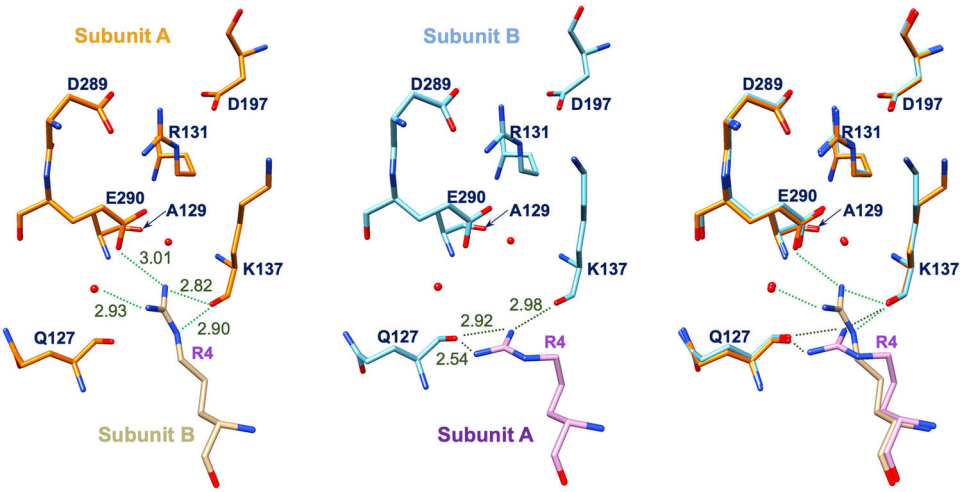

Fig. 9 | Arg4 movement connected to the second binding. In the nsp4/5 complex, Arg4-B is inserted into subunit A (left panel) and interacts strongly with Glu290-A. The same interaction is seen in symmetric free-M^{pro} where it is repeated for the symmetrically related Arg-A and Glu290-B. Instead, in the asymmetric nsp4/5 complex, Arg4-A moves away from Glu290-B toward Gln127-B and Lys137-B (central panel). In the right panel, superposition of the two conformations is shown, with light green and dark green dotted lines indicating the different interactions of Arg4-B and Arg4-A, respectively (as reported in the left and central panels). Electron density maps are reported in Supplementary Fig. 8.

possibility of distinguishing and quantifying the single species (D, DL, DL_2) with a titration experiment. The histograms in panel C of Fig. 10 report the percentage of bound M^{pro} (that is, M^{pro} with at least one binding site occupied by the ligand) observed at the 1:10 molar ratio for each peptide. These percentages provide a first qualitative clue of the relative affinities of the substrate peptides for M^{pro}, indicating that nsp5/6 and nsp14/15 have comparable affinities, which are lower than that of nsp4/5 under the same experimental conditions.

We then proceeded to quantify the dissociation constants of the three peptides by analyzing the results of titration experiments at increasing concentrations of each peptide, in particular with molar ratios 1:0.5, 1:1, 1:2, 1:3.5, 1:5 and 1:10 (protein:substrate ratios). Characteristic binding curves are exemplified in panels D, E, and F of Fig. 10. Using a sequential binding model (see experimental for details), we could determine the dissociation constants of peptides nsp4/5, nsp5/6 and nsp14/15 for the two binding sites. The experimentally derived apparent dissociation constants K_1' and K_2' were converted into intrinsic dissociation constants K_1 and K_2 using the statistical relationship $K_1 = 2K_1$' and $K_2 = 0.5K_2$' (Table 2). The condition for positive cooperativity is $K_2\backslash K_1 < 1$ (i.e. $K_2'/K_1' < 4$), for negative cooperativity is $K_2\backslash K_1 > 1$ (i.e., $K_2'/K_1' > 4$); for $K_2\backslash K_1 = 1$ (i.e., $K_2'/K_1' = 4$) there is no cooperativity. The results confirm a general weaker affinity for peptides nsp5/6 and nsp14/15 compared to peptide nsp4/5, with dissociation constants approximately one order of magnitude higher. For nsp4/5, nMS shows a lower affinity for binding to the second active site, with apparent constants $K_1' = 3.5$ μM and $K_2' = 45.6$ μM, corresponding to intrinsic constants $K_1 = 6.9$ and $K_2 = 22.8$ μM, indicating moderate negative cooperativity. Similarly, the second binding is weaker for nsp5/6, with apparent constants $K_1' = 41.8$ μM and $K_2' = 273$ μM, corresponding to intrinsic constants $K_1 = 83.6$ μM and $K_2 = 137$ μM. Instead, nsp14/15 shows a slightly higher affinity in the second binding, with apparent constants $K_1' = 65.4$ μM and $K_2' = 116$ μM, corresponding to intrinsic constants $K_1 = 131$ μM and $K_2 = 58.2$ μM, indicative of moderate positive cooperativity.

Isothermal titration calorimetry (ITC) measurements

For peptide nsp4/5, we also analyzed binding to M^{pro} by ITC. Representative ITC measurements of M^{pro} titrated with substrate peptide nsp4/5 are shown in Fig. 11. Raw titration data (panel A) show a considerable release of heat for each injection. The resulting binding isotherm (panel B) was interpolated with a sequential binding model, which returned the apparent constants $K_1' = 4.4$ μM and $K_2' = 27.2$ μM, corresponding to the intrinsic dissociation constants $K_1 = 8.8$ μM and $K_2 = 13.6$ μM. These constants are in line with those obtained by nMS (see Table 2), which confirms moderate negative cooperativity. As shown in insert panels C and D, the entropic contributions to the binding are different, with a more favorable entropy for the first event. This is in accordance with the concept of an entropy compensation effect for the first binding due to the mobilization (that is, an increase of entropy) of the region 42–50 in the second binding site that compensates for the entropic penalty of the first binding (see the discussion section for a further analysis of this point). Interestingly, in a monomeric form of M^{pro} (mutant Glu290Ala and Arg298Ala), the value of K_d for a substrate based on nsp4/5 is 193 μM[14], 24 times higher compared to the averaged K_1 derived here by nMS and ITC (7.9 μM). This higher affinity of the dimer can be partially attributed to the entropy compensation effect.

We tried to determine dissociation constants for the other two substrates but were unsuccessful. Unlike nMS experiments, which allowed us to work at a low peptide concentration (4.5–80.5 μM) and therefore to determine dissociation constants for the three peptides, ITC measurements, which require higher working concentrations, were limited by the modest solubility of the peptides nsp5/6 and nsp14/15 in the aqueous environment coupled with a lower affinity (as shown by nMS). Titration with the peptide nsp5/6 shows a partially incomplete ITC thermogram, even under the best experimental conditions, while titration with the peptide nsp14/15 (poorly soluble as nsp5/6) show a very low heat change, hampering the characterization of its binding to M^{pro}.

ITC experiments with an nsp4/5 peptide in which Gln P1 is replaced by an asparagine, which conserves the amide function but with a shorter side chain, show that binding is abolished (Supplementary Fig. 10). This confirms the stringent geometric requirements of having a glutamine in that position for binding, and the key role of the interactions involving the side chain amide function, with the Nε2 of His163, CO of Phe140, carboxylate of Glu166 and with Asn142 side chain via a water-mediated interaction (Supplementary Fig. 2).

Overall, in solution nMS and ITC measurements show that the structural perturbation of the second binding site followed by the binding to the first one, revealed by the crystal structures, does cause a change in the affinity constants, even if moderate, which is dependent on the type of substrate. For nsp4/5 and nsp5/6, intrinsic constant K_2 is higher than K_1 (moderate negative cooperativity), while for nsp14/15 is lower (moderate positive cooperativity).

Kinetic analysis

Allosteric communications can influence the thermodynamics of binding, as well as the kinetics of the reaction at two connected active sites. Therefore, we assessed the activity of the enzymatically active wtM^{pro} on peptide substrates employing a fluorescence resonance energy transfer (FRET)-based cleavage assay in vitro. To be consistent with the structural and thermodynamics data, we ensured that the substrate peptides with fluorescent probes used for the enzymatic assay fully comprise the 11-mer sequence of those used in the previous analysis.

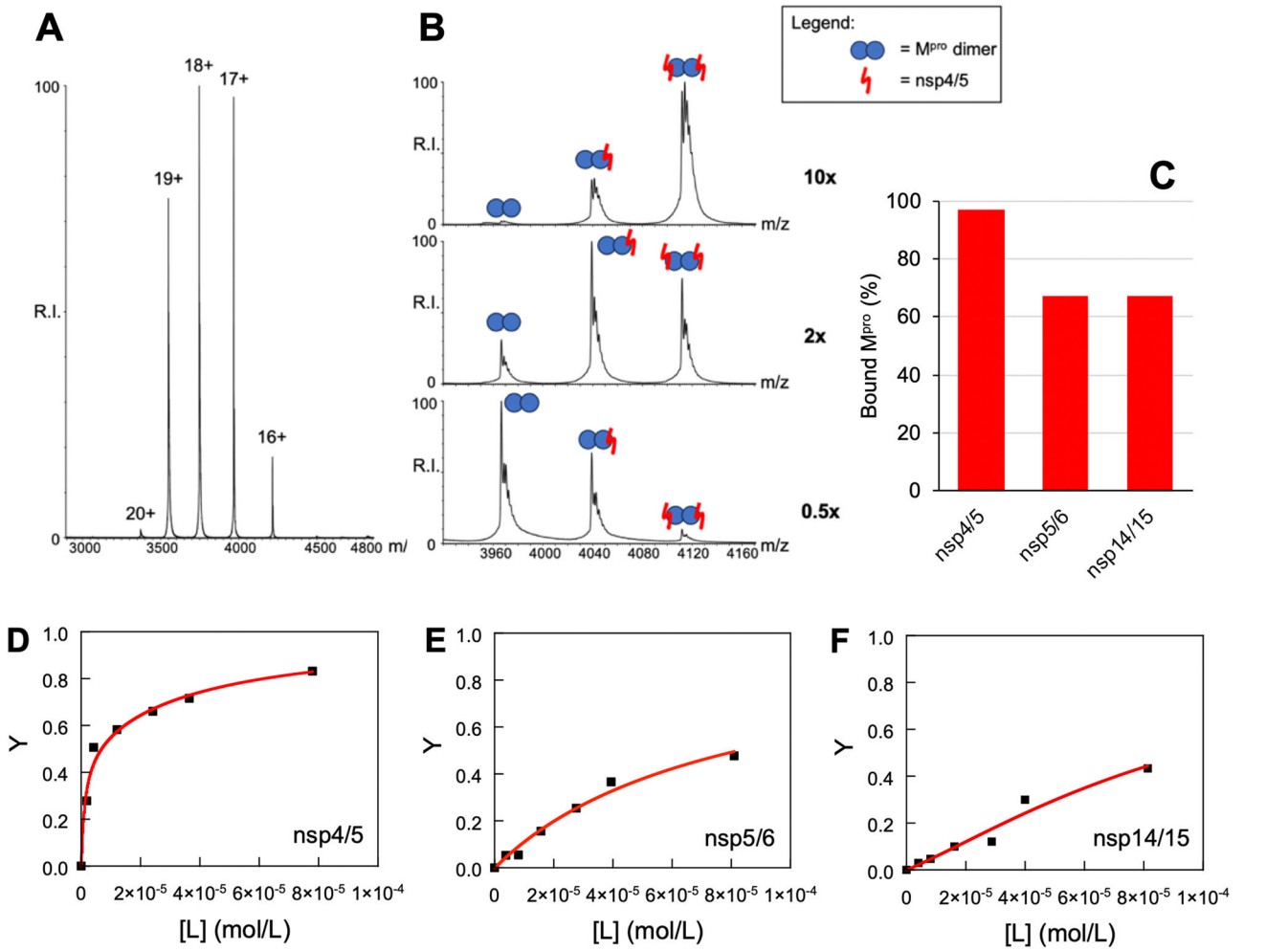

Fig. 10 | Analysis of the M^{pro} interactions with substrate peptides by nMS. Representative ESI-MS spectrum of free M^{pro} at a concentration of 8.5 µM (monomer) (**A**) and of samples containing a 1:0.5, 1:2, and 1:10 molar ratio of M^{pro} and nsp4/5 (**B**) in 150 mM ammonium acetate. **C** Histograms displaying the percentage of M^{pro} molecules with at least one active site occupied by the ligand (bound M^{pro}) observed at the 1:10 molar ratio between M^{pro} and each peptide. **D**, **E**, **F** Representative saturation curves of M^{pro} active sites (Y is the fraction of active sites occupied by the ligand) by increasing amounts of substrate peptides nsp4/5, nsp5/6 and nsp14/15, respectively. In the abscissa, [L] is the concentration of the free ligand (i.e. substrate peptides). The red curves refer to the fitting by the sequential binding model (see experimental, Eq. (3)); fitting $R^2 = 0.985$ for nsp4/5 (**D**), 0.982 for nsp5/6 (**E**) and 0.958 for nsp14/15. The derived apparent and intrinsic dissociation constants reported in Table 2 are the mean of the values obtained by the fittings of two independent titration experiments. Reported errors are the half-difference between the maximum and the minimum values.

The curve of the reaction rate versus the substrate concentration shows a sigmoidal trend rather than the hyperbolic shape of the classical Michaelis-Menten model, suggesting the presence of positive cooperativity. This was confirmed by the fitting with the Hill equation (see the experimental section) that gave a Hill coefficient of 1.59 (Fig. 12). This is in agreement with data reported in the literature[32]. We then fitted these data with a model representing an allosteric homodimer enzyme with two apparent dissociation constants K_1' and K_2' and two kinetic constants k_{c1} e k_{c2} (see the experimental section) (Fig. 12). The parameters derived from the fittings are reported in Table 2. In the hypothesis assumed in our model and illustrated in the experimental section, the constants K_1' and K_2' determined by kinetic measurements correspond to the thermodynamic dissociation constants measured by ITC and nMS. The values of the apparent constants $K_1' = 9.7$ µM and $K_2' = 32.4$ µM are close to those derived by the thermodynamic characterization of binding by nMS and ITC, indicating moderate negative cooperativity. Remarkably, a large difference is found between the two kinetic constants, with $k_{c2} = 1.20\ s^{-1}$ much higher than $k_{c1} = 0.014\ s^{-1}$. Notably, our k_{c1} is close to the k_{cat} of $0.016\ s^{-1}$ derived for the monomeric mutant Glu290Ala and Arg298Ala of M^{pro} [14]. It can be concluded that, overall, a positive allostery is present and that it has kinetic rather than thermodynamic bases.

We tried to also kinetically characterized the peptide substrates nsp5/6 and 14/15, but the results were not reliable due to the tendency to aggregate of these peptides, most probably due to the presence of fluorescent (hydrophobic) probes at the extremities (not present in the nMS experiments, as well as in the crystallographic data).

Discussion

The crystal structures reported here show that different substrate peptides interact with M^{pro} in different ways. Substrate peptide bindings are primarily mediated by polar interactions, mainly hydrogen bonds, involving both the backbone and the side chains (Supplementary Figs. 2 to 5). Apolar interactions are restricted to residues that interact with subsites P4 and, mainly, P2, where a leucine is preferred (9 out of 11 situations, with Phe and Val as alternatives). Variations in substrate peptide sequences correspond to differences in the mode of binding for both the ligand and the enzyme, which undergoes modifications mainly, but not exclusively, at the active site, showing how M^{pro} plasticity is exploited to face differences in the sequences of the natural substrates.

The most conserved interaction is that of the invariant Gln P1 and Ser P1', whose peptide bond is cleaved by protease. In all three peptides, Gln P1 strongly interacts with the S1 subsite, forming analogous interactions. Substitution of this glutamine with an asparagine abolishes binding (Supplementary Fig. 10). The backbone of the oxyanion residues 143-145 is wrapped around the peptide bond to be cleaved, in the classical active conformation needed to ensure the proper stabilization of the oxyanion intermediate. Instead, clear differences are visible at both ends of the peptide substrates, depending on the chemical nature of certain residues, particularly at positions P4, P6, and P4'. Relatively bulky residues at position P4 (valine and threonine in nsp5/6 and nsp14/15, respectively, instead of the alanine in nsp4/5) push away from the protein the peptide backbone. For nsp14/15, Thr P6 (instead of a serine in nsp5/6) is capable of folding the chain towards the protein due to interactions of the side chain hydroxyl group. In nsp4/5 and nsp14/15, position P4' is important for the anchoring of the substrate to the enzyme through an arginine and a glutamine, respectively (although in a different manner). In nsp5/6 this region is essentially mobile, with no interaction with the protein matrix.

Table 2 | Thermodynamic and kinetic parameters

	nsp4/5			nsp5/6	nsp14/15
	Kinetics*	ITC	nMS	nMS	nMS
K_1' (μM)	9.7 ± 3.5	4.5 ± 1.0	3.5 ± 2.3	41.8 ± 2.2	65.4 ± 25.5
K_2' (μM)	32.4 ± 12.4	27.1 ± 3.9	45.6 ± 6.6	273 ± 102	116 ± 14
K_1 (μM)	19.4 ± 7.0	8.9 ± 2.0	6.9 ± 4.6	83.6 ± 4.4	131 ± 51
K_2 (μM)	16.2 ± 6.2	13.5 ± 2.0	22.8 ± 3.3	137 ± 51	58.2 ± 6.9
k_{c1} (s^{-1})	0.014 ± 0.005				
k_{c2} (s^{-1})	1.20 ± 0.15				
Hill coefficient	1.59 ± 0.27				

Experimentally determined "apparent" (K_1' and K_2') and derived "intrinsic" (K_1 and K_2) dissociation constants for peptide substrates nsp4/5, nsp5/6 and nsp14/15 as determined by nMS and ITC. "Kinetics" K_1', K_2', K_1, K_2, k_{c1} and k_{c2} for the peptide nsp4/5 were derived from fitting the plot of the initial rates of the catalyzed reaction versus the substrate concentrations with the equation corresponding to the allosteric model described in the experimental section. See also Figs. 10, 11, and 12. Reported errors are the half-difference between the maximum and the minimum values obtained by independent experiments (see Methods for more details). *In our hypothesis, illustrated in the experimental section, the constants K_1' and K_2', determined by kinetic measurements, correspond to the thermodynamic dissociation constants measured by ITC and nMS. As the constants K_1' and K_2' obtained by kinetic measurements are close to those obtained by nMS and ITC, the assumption is acceptable.

Binding of substrates perturbs not only the area around the active site (Fig. 4) but also distant regions (Figs. 3, 6, and 7), to such an extent that these rearrangements are not compatible with certain crystal packings, C2, for instance, as demonstrated by differences in our M^{pro} complexes with the substrate peptide nsp4/5 determined by cocrystallization and soaking (Figs. 4 and 5). Comparison with other structures of M^{pro} complexes with substrate peptides is provided in the Supplementary Note 5. In particular, structures 7T70, 7N89, 7MGS, 7DVP, 7T8M, 7DVW, 7DW0 are examined for their strengths and weaknesses. In general, this analysis indicates that the full structural effects of the substrate binding can be seen if at least an 11-mer peptide is used (with no apparent advantage to use a longer one), from position P6 to P5', and using a co-crystallization experiment (rather than a soaking one), in which rearrangements and perturbations of the two subunits are enabled.

Substrate binding causes a global rearrangement of the M^{pro} architecture, from a symmetric to an asymmetric dimer. The relative orientation of the two subunits is modified, as well as that of the single domains in each subunit; this is illustrated in Supplementary Movies 1 and 2. The movements of arginine 4 and 298 side chains in a single subunit play a key role in the symmetry breaking observed in substrate peptide complexes (Supplementary Figs. 6, 8 and 9). For substrate peptides nsp5/6 and nsp14/15, only the binding site of one subunit is occupied (Figs. 6, 7), but, interestingly, the binding site in the free subunit is also significantly affected. This indicates that the two binding sites are structurally connected and can allosterically communicate with each other. In particular, binding at one site increases the mobility of the other, mostly in region 44–60 which contributes to the hydrophobic subsite S2.

Based on the determination of the binding constants by nMS (Fig. 10), the affinity of the peptides nsp5/6 and 14/15 is one order of magnitude lower than that of nsp4/5 (Table 2). For nsp5/6 this is due to the lower number of interactions with the protein, as illustrated in Supplementary Fig. 4, with greater mobility of the two extremities. For nsp14/15 (Supplementary Fig. 5) the number of interactions, in particular H-bonds, as well as the hydrophobic contribution are similar to those of nsp4/5 (Supplementary Fig. 2). What is different is a larger distortion in the area around residues 23–24 and, mainly, 43–51, principally due to the presence of glutamic acid in P3' in the ligand, with a minor contribution of asparagine in P4'. These two residues are inserted between regions 23–24 and 43–51, strongly interacting with Thr24 (see Supplementary Fig. 5) and causing a steric shift of the adjacent region 43–51. Another negative entropic contribution arises from the exposure to the aqueous environment of two hydrophobic residues present only in

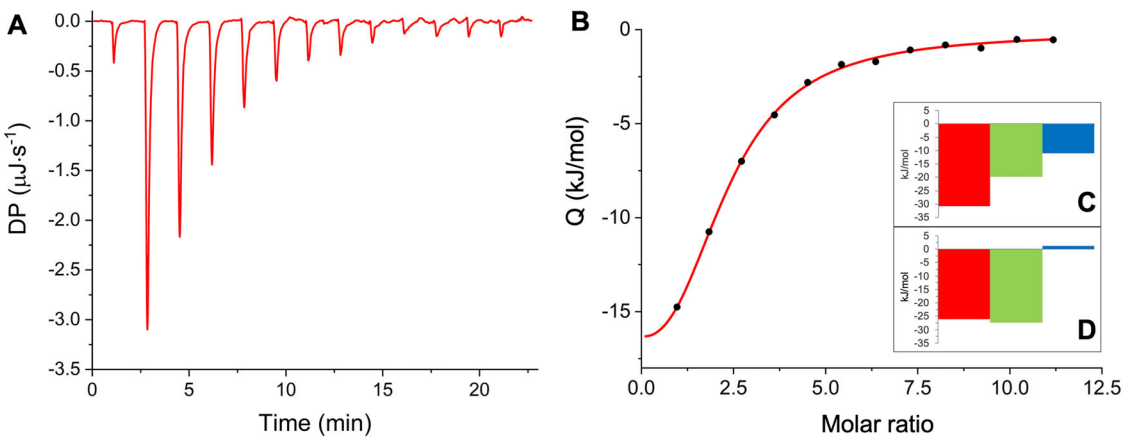

Fig. 11 | ITC substrate binding measurements. In (**A**, **B**), representative raw ITC data and final binding thermogram of the M^{pro} titration by the peptide substrate nsp4/5, respectively. In (**B**), the red curve refers to the sequential binding model fitting as implemented in the Microcal PEAQ-ITC analysis software (reduced $\chi^2 = 0.027$). The derived dissociation constants reported in Table 2 are the mean of three independent experiments. Errors are the half-difference between the maximum and the minimum values. In insert panels C and D, the energetics of the first and second bindings are reported, respectively. ΔG in red, ΔH in green and ΔS in blue.

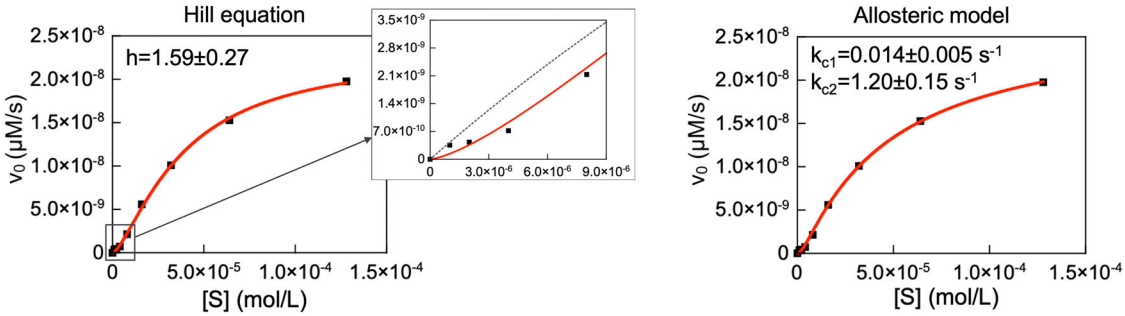

Fig. 12 | wtM^{pro} kinetics with the substrate peptide nsp4/5. A representative plot (one out of three experiments) of the initial velocities v_0 versus increasing concentrations of substrate peptide nsp4/5. The experimental points (black squares) were fitted with the Hill equation (left, $R^2 = 0.999$) (see experimental, Eq. (4)). The inserted panel, which represents the zoomed-in data at very low substrate concentrations, shows the sigmoidal trend typical of a positive cooperativity; for comparison, the hyperbolic fitting with the classical Michaelis-Menten model is shown with a grey dotted line. The same plot was then fitted with an allosteric model (Eq. (7)) described in the experimental section (right, $R^2 = 0.999$). R^2 is the coefficient of determination of the fittings. The final Hill coefficient h, the apparent dissociation constants K1' and K2', as well as the kinetic constants k_{c1} and k_{c2} are reported in Table 2. The reported values of h, k_{c1} and k_{c2} are the mean of three independent experiments, and the errors are the half-difference between the maximum and the minimum values.

nsp14/15, namely Leu in P2' (instead of Gly and Ala) and Phe in P5 (instead of Ser and Gly).

This unique feature of nsp14/15 due to glutamic acid and glutamine in P3'-P4' is also at the bases of the positive cooperativity of binding, instead of the negative one seen for nsp4/5 and nsp5/6 (see Table 2). The area of higher mobility in the second binding site (42–50) is precisely the area that is most disturbed by the binding of nsp14/15 (Fig. 7). It is conceivable that the second binding of nsp14/15 is entropically less unfavorable than the first one as the site is more suited to accept the substrate. Therefore, the positive allostery is connected to the presence in nsp14/15 of Asn-P4' and, above all, of Glu-P3'. This pair of polar residues is not present in other M^{pro} substrates (Fig. 1).

Kinetic analysis shows marked positive cooperativity in the processing of the substrate nsp4/5, as demonstrated by the Hill coefficient of 1.59 (Fig. 12). In general, cooperativity can have not only thermodynamic (i.e. differences in dissociation constants) but also kinetic bases (in terms of catalytic efficiency). Analysis of the kinetic data by an allosteric model gives dissociation constants close to those derived by nMS and ITC (see Table 2). Remarkably, a significant difference is observed between the values of k_{c1} (0.014 s^{-1}) and k_{c2} (1.20 s^{-1}) for the substrate peptide nsp4/5, indicating that the positive cooperativity is to be attributed to the kinetics rather than the thermodynamics (of binding).

The model for a positive cooperativity mechanism of catalysis for M^{pro} derived by this analysis of the structural and electrostatic potential modifications described here is shown in Fig. 13. The model has three principal states and is described below.

a. In the absence of substrate binding, both subunits of the dimeric enzyme are potentially active; the free enzyme is symmetric (state 1) as well as the electrostatic clusters organized around the two-fold axis. This is verified by our structure of free-M^{pro} (Fig. 2), as well as by the known structures of free-wtMpro available in the PDB. Before any binding event, the two binding sites can be considered perfectly equivalent.
b. Upon binding of the substrate to one subunit (a), the enzyme becomes asymmetric (state 2): This is clearly observed for our M^{pro} structurers in complex with nsp5/6 and nsp14/15 (Figs. 6, 7); the Arg298 side chain of subunit a moves into the space freed by exposure of the C-terminal tail. The free subunit (b) can have a regulatory role by means of an entropy compensation mechanism: the higher dynamic in the subunit b compensates for the entropy penalty associated with the binding to subunit a (see ITC data and Fig. 11). This mechanism is well documented, for example, in the case of the homodimeric fluoroacetate dehalogenase FAcD, for which substrate binding to one protomer is allosterically communicated to the empty protomer with an entropy compensation effect. In the singly occupied dimeric M^{pro}, the reaction can proceed with k_{c1} to reconstitute the free symmetric enzyme and the release of the products.
c. A second substrate molecule can bind to the free subunit of state (2) with moderate negative (for nsp4/5 and nsp5/6) or positive (for nsp14/15) cooperativity, depending on the type of substrate (Table 2), generating state (3). The double occupied enzyme, state (3), is still asymmetric, as shown by our structure of the nsp4/5 complex (Fig. 3). Arg4 of subunit a moves away from Glu290 towards Gln127 and Lys137 of the subunit b. The reaction can proceed in one of the two sites with the reconstitution of state (2), with a k_{c2} around 80 times higher than k_{c1} (analogous to the monomeric M^{pro} k_{cat}), and this is the basis of the overall positive cooperativity (Fig. 12). It can be deduced that allostery between active sites is working properly only in state (3) when both active sites are occupied, and not in the singly occupied state (2).

On the basis of this allosteric model, the advantage of the dimeric form of M^{pro} compared to the monomeric one lies not only in a better binding of the substrates, with the second subunit participating in the formation and stabilization of the binding site of the first (thermodynamic effect), but also in a significant increase in processing efficiency (kinetic effect).

Our results and the proposed positive cooperativity model are supported by other studies. The presence of allosteric communication networks between subunits in M^{pro} was effectively verified by dynamical non-equilibrium molecular dynamics (D-NEMD) simulations, performed on a dimer with only one occupied binding site[33]. It was shown that in response to substrate binding to a single site, there are local responses (of residues in direct contact with the substrate) and also remote ones, on more distant residues in areas not in direct contact with the substrate (allosteric responses to active site binding).

The major local responses shown by D-NEMED are in the oxyanion loop and in residues 165–166, 189–192, and 23–26 (of subunit a), according to our structural analysis. These local alterations in subunit a, in particular of the oxyanion loop and Glu166, are transmitted to the second subunit (b) following a path that connects residues from the binding site of subunit a to the N-terminal region of subunit b (1–2), to the adjacent region 214–215 (of subunit b), and finally to the C-terminal region of subunit b, around residues 300–302. This is a pathway (pathway 2 in Fig. 11) that connects the binding to subunit a to the perturbation of the C-terminal end of subunit B and therefore of the dimeric interface. Regarding the mobilization of region 42–50 in subunit b, it was recognized that this region is involved in allosteric responses to the binding, but the connection to the active site of subunit a (pathway 1 in Fig. 11) is not clear. A discussion of other studies supporting our results and the proposed positive cooperativity model is reported in the Supplementary Note 6.

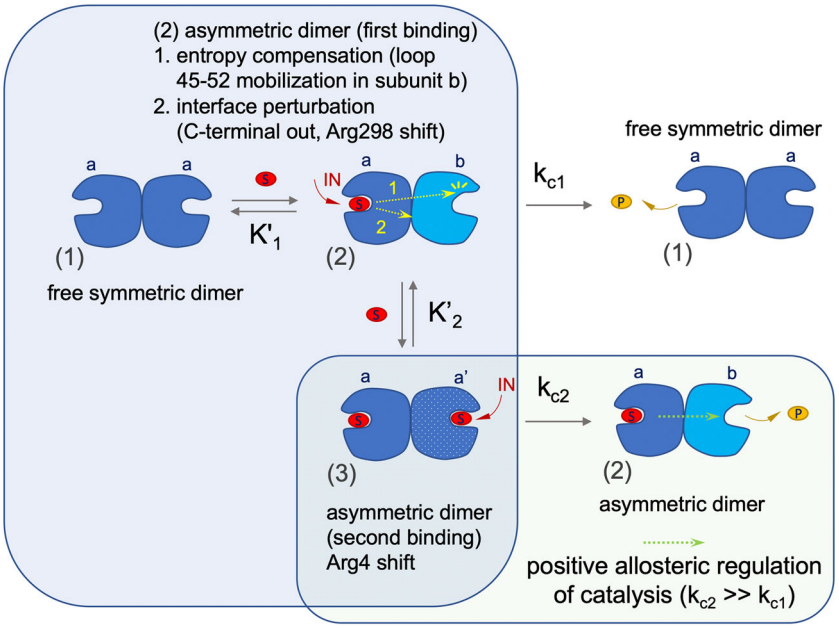

Fig. 13 | Proposed model of the M^{pro} catalytic cycle regulated by positive cooperativity. State (1) corresponds to the structure of the symmetric free enzyme, state (2) to the complexes with nsp5/6 or 14/15, and state (3) to the complex with nsp4/5. For nsp4/5, $K_2' > K_1'$, and K_2 approximately twice K_1 (moderate negative cooperativity for binding), while k_{c2} is 1.20 s^{-1} and k_{c1} is 0.014 s^{-1}, determining the overall positive cooperativity of the catalysis. Further details are discussed in the main text.

The allosteric model based on positive cooperativity through asymmetric dimers proposed here finds justification at the physiological level in the need to fine regulate the processing of viral polyproteins in time and space to precisely control the stepwise polyprotein processing mechanism that leads to the production of functional units (nsp proteins), with the binding and hydrolysis of many different substrate sequences in a defined order. Our data with three different substrate peptides suggest different M^{pro} properties depending on the substrate to process. For example, the higher affinity of substrate nsp4/5 is probably related to the fact that it is the first to be hydrolyzed in the maturation process of the M^{pro} dimer. It is also possible that a dimer of M^{pro} binds and processes two different substrates at the same time or in a strictly regulated sequential process, with the first binding modulating the second event.

Methods

Substrate peptide synthesis

The peptides were synthesized using the SPPS methodology according to the Fmoc/tBu protocol. Details of peptide synthesis, purification, and characterization are reported in the Supplementary Note 1. We also synthesized an analogue of nsp4/5 modified by adding a Lys(Dabcyl) probe at the N-terminus and a Glu(Edans) probe at the C-terminus with the aim of confirming the cleavage site by FRET experiments. A further modification of the native sequence of nsp4/5 has been performed introducing an additional Lys residue at position P6' to increase the solubility in buffer solution.

Cloning, expression and purification of M^{pro}

The sequence encoding wild-type SARS-CoV-2 M^{pro} was inserted in a pET-SUMO expression vector with the restriction-free cloning method. Subsequently, for the production of the catalytic inactive mutant, the resulting plasmid was double-mutated to H41A and C145A with a single site-directed mutagenesis reaction with primers 3′-AGAGGTGCA-GATGAC**AGC**ACGAGGGCAATAGAC-5′ and 5′-TTCCTTAATGG-CAGC**GCT**GGTTCGGTGGGCTTT-3′ (mutation codons indicated in bold). The resulting plasmid was sequenced to assess the success of both the insertion and mutation procedures. Production and purification of wtM^{pro} (used for kinetic assays) and inactive mutant of M^{pro} (used for crystallography, nMS and ITC measurements) were similar and are detailed in the following. The plasmid was transformed into *Escherichia coli* strain BL21(DE3) and precoltured in LB broth at 37 °C overnight in the presence of kanamycin (50 µg/mL). The following day, the precolture was used to inoculate fresh LB medium supplemented with the antibiotic and the cells were allowed to grow at 37 °C to an OD^{600} of 0.6–0.8 before induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The induction phase was allowed to proceed overnight at 20 °C. The following day, cells were harvested by centrifugation (5000 g, 6 °C, 20 min) and pellets were resuspended in buffer A (20 mM Tris, 150 mM NaCl, pH 7.8) supplemented with DNase I and $MgCl_2$ for lysis. The lysate was clarified by centrifugation at 12,000 g at 6 °C for 0.5 h and loaded onto a HisTrap HP column (GE Healthcare) equilibrated with 98% buffer A and 2% buffer B (20 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.8). The column was washed with 98% buffer A and 2% buffer B, and His-tagged M^{pro} was then eluted with a linear imidazole gradient from 10 to 500 mM. The pooled fractions containing the target protein were subjected to buffer exchange with buffer A using a HiPrep 26/10 desalting column (GE Healthcare). Next, ULP-1 protease was added (~ 2 µg of ULP-1 per mL of M^{pro} solution) to cleave the N-terminal HisTag-SUMO moiety and allowed to incubate at 12 °C overnight. The following day, the protein solution was loaded onto a HisTrap HP column equilibrated in buffer A to separate the cleaved M^{pro} from the His-tagged ULP-1 protease, the HisTag-SUMO moiety, and the uncleaved protein. M^{pro} was finally purified using a Superdex 200 prep grade 16/600 SEC column (GE Healthcare) equilibrated with buffer C (20 mM Tris, 150 mM NaCl, pH 7.8). The fractions containing the high purity target protein were pooled, concentrated at 12 mg/ml and flash-frozen in liquid nitrogen for storage in small aliquots at −80 °C.

Protein crystallization and X-ray crystallography

For crystallization, frozen M^{pro} aliquots at 12 mg/mL (0.36 mM) were thawed in ice and clarified by centrifugation at 17000 g. The peptides were dissolved in buffer C (20 mM Tris, 150 mM NaCl, pH 7.8) at a concentration of 100 mM. The peptides nsp5/6 and nsp14/15 did not completely dissolve and thus the corresponding suspensions were briefly agitated with a vortex immediately prior to their withdrawal. M^{pro} was crystallized both in the free form and in the presence of peptides by co-crystallization. Crystals for all structures were obtained with the same protocol: by microseeding small crystals of the free enzyme, and using the hanging-drop vapor diffusion method at 18 °C, mixing 1.0 µl M^{pro} solution with 1.0 µl precipitant solution [0.1 M MMT (dl-malic acid, MES and Tris

base in a 1:2:2 molar ratio) pH 7.0, 25% PEG 1500] and 0.2 µl seed stock (diluted 1:100, 1:1000 or 1:2000 with precipitant solution) and equilibrating with a 300 µl reservoir of precipitant solution. The crystals appeared overnight and completed growth within 48 h after the crystallization drops had been prepared. For co-crystallization, M^{pro} was incubated for 20 h at 4 °C with a 14-fold molar excess of peptide (95:5 ratio between the M^{pro} solution and the peptide stock solution, for a theoretical final concentration of peptide of 5 mM). In the case of the nps5/6 and nsp14/15 peptides, the stock solution suspension of partially solubilized peptides was added directly after mixing. For soaking experiments, 0.2 µL of the nsp4/5 stock solution was added to the crystals of free M^{pro} and allowed to incubate for 24–48 h before fishing. For data collection, crystals were harvested from the drops and flash-cooled in liquid nitrogen. Data collections were performed at ESRF on the ID23-1 beamline. Integration and scaling of the diffraction data were performed with XDS[34], data reduction and analysis with Aimless[35]. The structures were solved by Molecular Replacement (MR) with Phaser[36] from Phenix[37], using structures 6Y2E and 7NIJ as search models. The final models were obtained by alternating cycles of manual refinement with Coot[38] and automatic refinement with phenix.refine[39]. Statistics on data collection and refinement are reported in Table 1. Interface analysis was performed using PISA[40].

Native mass spectrometry (nMS)

The inactive mutant of M^{pro} was buffer exchanged in 150 mM ammonium acetate (pH 7.5) by using Micro Bio-Spin Chromatography Columns (Bio-Rad) to minimize the presence of salts that can adversely interfere with the performance of ESI. Samples for binding studies were prepared by mixing appropriate volumes of M^{pro} (8.5 µM final monomer concentration) with each substrate peptide, i.e. nsp4/5, nsp5/6 and nsp14/15, in 150 mM ammonium acetate (pH 7.5). The final mixtures contained 1:0.5, 1:1, 1:2, 1:3.5, 1:5 and 1:10 M^{pro}:peptide molar ratio. Control experiments were carried out on 8.5 µM solutions of M^{pro} alone in 150 mM ammonium acetate. All samples were analyzed by direct infusion electrospray ionization (ESI) on a Q-Tof Xevo G2S spectrometer (Waters). The analyses were performed in nanoflow mode by using quartz emitters produced in-house by using a Sutter Instruments Co. (Novato, CA, USA) P2000 laser pipette puller. Up to 5.0 µL samples were typically loaded onto each emitter by using a gel-loader pipette tip. A stainless-steel wire was inserted into the back end of the emitter to supply an ionizing voltage in the range of 1–1.5 kV. The source temperature was set at 30 °C and the desolvation voltage was 40 V. All experiments were performed in positive ion mode. Data were processed using Mass Lynx (v 4.2, SCN781; Waters) software. The abundances of the different species in each experiment were calculated aided by in-house developed MATLAB (R2016b) scripts, expressed as a percentage to calculate their molar fraction, which were then used to calculate the concentration of each species in solution. Solution concentrations obtained from the respective molar fractions were used to obtain the binding curves. The fraction of occupied binding sites, Y, and the concentration of free ligand in solution, [L], were derived from the quantification of the free dimer, D, the dimer with one occupied binding site, DL, and the dimer with two occupied binding sites, DL_2, considering the total concentration of the dimer (4.25 µM) and the total concentration of the added ligand (1:1, 2:1, 4:1, 7:1, 10:1 and 20:1 as ligand:dimer ratios). The data points of Y vs [L] were fitted with the equation corresponding to the sequential binding model (Eq. (3)) with apparent dissociation constants K_1' and K_2' (Eqs. (1) and (2)).

$$D + L \leftrightharpoons DL \quad K'_1 = \frac{[D][L]}{[DL]} \quad K_1 = 2K'_1 \tag{1}$$

$$DL + L \leftrightharpoons DL_2 \quad K'_2 = \frac{[DL][L]}{[DL_2]} \quad K_2 = 0.5K'_2 \tag{2}$$

$$Y = \frac{\frac{[L]}{K'_1} + 2\frac{[L]^2}{K'_1K'_2}}{2\left(1 + \frac{[L]}{K'_1} + \frac{[L]^2}{K'_1K'_2}\right)} \tag{3}$$

The experimentally derived apparent dissociation constants K_1' and K_2' were converted into intrinsic dissociation constants K_1 and K_2 using the statistical relationship $K_1 = 2K_1$' and $K_2 = 0.5K_2$' (Eqs. (1) and (2)). Statistic factors 2 and 0.5 derive from the consideration that the species D has two equivalent free active sites, while DL has only one. That is, for the first site (to generate DL), there is only one possibility to dissociate and two for associate; for the second site (to generate DL_2) the opposite, two possibilities to dissociate and one for associate. The condition for positive cooperativity is $K_2 < K_1$ (i.e. $K_2'/K_1' < 4$), for negative cooperativity is $K_2 > K_1$ (i.e. $K_2'/K_1' > 4$); for $K_2\backslash K_1 = 1$ (i.e. $K_2'/K_1' = 4$) there is no cooperativity.

The values for substrate peptides nsp4/5, nsp5/6 and nsp14/15 reported in Table 2 are the mean of the values derived by the fittings of two independent titration experiments. Reported errors are the half-difference between the maximum and the minimum values.

Isothermal titration calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) experiments were performed using a Malvern PEAQ-ITC microcalorimeter, at 25 °C, with a 270 µL sample cell and a computer controlled microsyringe for titrant injections. The inactive mutant of M^{pro} (30–80 µM) and peptide substrate nsp4/5 (0.9–1.2 mM) were in 20 mM Tris, 150 mM NaCl, pH 7.8. After baseline stabilization, a further delay of 60 s was used before the first injection. In each of the three titrations, a starting injection of nsp4/5 of 0.4 µL in 0.8 s was followed by others of different volume: 12 injections of 3 µL with a duration of 6 s each for the first titration and 24 injections of 1.5 µL with a duration of 3 s each for the second and third ones. A delay of 100 s was applied between each injection of nsp4/5. The raw data was then integrated to give the Wiseman plot, automatically obtained by software analysis (excluding the heat referred to the first 0.4 µL injection). The Wiseman plot was finally fitted by the theoretical binding curve using the sequential binding model implemented in the Microcal PEAK-ITC analysis software. The experimentally derived apparent dissociation constants K_1' and K_2' were converted into intrinsic dissociation constants K_1 and K_2 using the statistical relationship $K_1 = 2K_1$' and $K_2 = 1/2K_2$'. The K_1' and K_2' values reported in Table 2 are the mean of three independent experiments. Reported errors are the half-difference between the maximum and the minimum values.

Titration with the peptide nsp5/6 shows a partially incomplete ITC thermogram, even under the best experimental conditions, due to lower solubility coupled with lower affinity. ITC titration experiments with the peptide nsp14/15 (poorly soluble as nsp5/6) show a very low heat change, hampering the characterization of its binding to M^{pro}.

Enzyme kinetics

To characterize the enzymatic activity of our recombinant wtM^{pro}, we adopted a FRET-based assay using the substrate (Dabcyl)K-TSAVLQSGFRKK-E(Edans), synthesized as reported in the Supplementary Note 1. The assay was carried out by mixing wtM^{pro} with various concentrations (from 1 to 128 µM) of substrate (diluted from a stock solution in 100% DMSO) in a buffer composed of 20 mM Tris, 100 mM NaCl, 1 mM EDTA, 1 mM DTT pH 7.3, for a constant final M^{pro} concentration of 0.050 µM. Maximum final concentration of DMSO was 3.75%. The fluorescence intensity (excitation at 355 nm and emission at 460 nm) was monitored at 37 °C with a VictorIII microplate reader (Perkin Elmer). A calibration curve was created by measuring multiple concentrations (from 0.001 to 5 µM) of free Edans in a final volume of 100 µL reaction buffer. The initial velocities were determined from the linear section of the curve, and the corresponding relative fluorescence units per time unit (ΔRFU s^{-1}) were converted to the amount of cleaved substrate per time unit (µM s^{-1}) by fitting to the calibration curve of free Edans.

The Hill coefficient "h" was derived by the non-linear fitting of the sigmoidal plot of the initial velocities of substrate hydrolysis versus the substrate concentration with the following Hill equation (Eq. (4), a modification of the Michaelis-Menten equation), with "v" initial velocities, and [S] the concentration of the substrate peptide:

$$v = \frac{V_{\max}[S]^h}{K_M{}^h + [S]^h} \quad V_{\max} = [E]_{tot}k_{cat} \tag{4}$$

As the Hill coefficient was higher than 1, indicating positive cooperativity, the same sigmoidal plot of the initial velocities of substrate hydrolysis versus the substrate concentration was then fitted by an allosteric model which includes two thermodynamic constants K_1' and K_2' and two kinetic constants k_{c1} and k_{c2}. The model is illustrated below (equations 5, 6, and 7) and assumes that the velocity of dissociation of the complexes (k_1 for DL and k_2 for DL_2) are much higher than the velocity of the catalytic event (k_{c1} and k_{c2}). In this hypothesis, K_1' and K_2' of this model correspond to the thermodynamic apparent dissociation constants defined in Eqs. (1) and (2). This assumption is supported by the results of the analysis of the kinetic data, which gives dissociation constants close to those derived by nMS and ITC, as reported in Table 2. Values of kinetic data reported in Table 2 are the mean of three independent experiments. Reported errors are the half-difference between the maximum and the minimum values.

$$D + L \underset{k_{-1}}{\overset{k_1}{\leftrightharpoons}} DL \longrightarrow^{k_{c1}} D + P \quad K'_1 = \frac{k_1}{k_{-1}} \tag{5}$$

$$DL + L \underset{k_{-2}}{\overset{k_2}{\leftrightharpoons}} DL_2 \longrightarrow^{k_{c2}} DL + P \quad K'_2 = \frac{k_2}{k_{-2}} \tag{6}$$

$$v = [DL]K_{c1} + [DL_2]K_{c2} = [D]_{tot}\frac{([L]K'_2k_{c1} = 2[L]^2k_{c2})}{2(K'_1K'_2 + [L]K'_2 + [L]^2)} \tag{7}$$

Statistics and Reproducibility

XRD data processing and model refinement statistics are reported in Table 1. For nMS, ITC, and kinetics experiments, the best operative conditions and reproducibility were tested with preliminary experiments. The relative data shown in Table 2 were obtained by final experiments as reported in the following. For nMS experiments, the final data are the mean values of the non-linear fittings of two independent titration experiments. For ITC and kinetic measurements, the final data are the mean of three independent experiments. The coefficient of determination R^2 for non-linear fittings is reported. For all three types of data, reported errors are the half-difference between the maximum and minimum values.

Reporting summary

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

Data availability

The coordinates and structure factors were deposited in the Protein Data Bank with the following accession codes: 9EX8, free M^{pro} in monoclinic C2; 9EYA, M^{pro} in complex with nsp4/5 in monoclinic C2 (by soaking); 9EXU, M^{pro} in complex with nsp4/5 in triclinic P1 (by cocrystallization); 9EZ4, M^{pro} in complex with nsp5/6 in orthorhombic $P2_12_12_1$ (by cocrystallization); 9EZ6 M^{pro} in complex with nsp14/15 in orthorhombic $P2_12_12_1$ (by cocrystallization). Other source data files are provided in Supplementary Data 1–4. The global rearrangement of the M^{pro} architecture upon substrate binding is shown in Supplementary Movie 1 (front view) and Supplementary Movie 2 (side view).

Received: 25 July 2024; Accepted: 24 October 2024;
Published online: 04 November 2024

References

1. Xia, B. & Kang, X. Activation and maturation of SARS-CoV main protease. *Protein Cell* **2**, 282–290 (2011).
2. Zhang, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **368**, 409–412 (2020).
3. Kim, T. H. et al. The role of dimer asymmetry and protomer dynamics in enzyme catalysis. *Science* **355**, eaag2355 (2017).
4. Snijder, E. J., Decroly, E. & Ziebuhr, J. Chapter Three - The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. in *Advances in Virus Research* (ed. Ziebuhr, J.) vol. 96 59–126 (Academic Press, 2016).
5. Dai, W. et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* **368**, 1331–1335 (2020).
6. Günther, S. et al. X-ray screening identifies active site and allosteric inhibitors of SARS-CoV-2 main protease. *Science* **372**, 642–646 (2021).
7. Ullrich, S. & Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg. Medicinal Chem. Lett.* **30**, 127377 (2020).
8. Douangamath, A. et al. Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nat. Commun.* **11**, 5047 (2020).
9. Jin, Z. et al. Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
10. Jin, Z. et al. Structural basis for the inhibition of SARS-CoV-2 main protease by antineoplastic drug carmofur. *Nat. Struct. Mol. Biol.* **27**, 529–532 (2020).
11. Song, L. et al. The main protease of SARS-CoV-2 cleaves histone deacetylases and DCP1A, attenuating the immune defense of the interferon-stimulated genes. *J. Biol. Chem.* **299**, 102990 (2023).
12. Qiao, J. et al. SARS-CoV-2 Mpro inhibitors with antiviral activity in a transgenic mouse model. *Science* **371**, 1374–1378 (2021).
13. Owen, D. R. et al. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. *Science* **374**, 1586–1593 (2021).
14. 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).
15. Kovalevsky, A. et al. Visualizing the Active Site Oxyanion Loop Transition Upon Ensitrelvir Binding and Transient Dimerization of SARS-CoV-2 Main Protease. *J. Mol. Biol.* **436**, 168616 (2024).
16. El-Baba, T. J. et al. Allosteric Inhibition of the SARS-CoV-2 Main Protease: Insights from Mass Spectrometry Based Assays. *Angew. Chem. Int. Ed. Engl.* **59**, 23544–23548 (2020).
17. Anand, K. et al. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* **21**, 3213–3224 (2002).
18. 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).
19. Frey, P. A. & Hegeman, A. D. *Enzymatic Reaction Mechanisms*. (Oxford University Press, Oxford, New York, 2007).
20. Fornasier, E. et al. A new inactive conformation of SARS-CoV-2 main protease. *Acta Cryst. D.* **78**, 363–378 (2022).
21. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763–1767 (2003).

22. Shaqra, A. M. et al. Defining the substrate envelope of SARS-CoV-2 main protease to predict and avoid drug resistance. *Nat. Commun.* **13**, 3556 (2022).
23. Kneller, D. W., Zhang, Q., Coates, L., Louis, J. M. & Kovalevsky, A. Michaelis-like complex of SARS-CoV-2 main protease visualized by room-temperature X-ray crystallography. *IUCrJ* **8**, 973–979 (2021).
24. MacDonald, E. A. et al. Recognition of Divergent Viral Substrates by the SARS-CoV-2 Main Protease. *ACS Infect. Dis.* **7**, 2591–2595 (2021).
25. Zhao, Y. et al. Structural basis for replicase polyprotein cleavage and substrate specificity of main protease from SARS-CoV-2. *Proc. Natl Acad. Sci. USA* **119**, e2117142119 (2022).
26. 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).
27. Wralstad, E. C., Sayers, J. & Raines, R. T. Bayesian Inference Elucidates the Catalytic Competency of the SARS-CoV-2 Main Protease 3CLpro. *Anal. Chem.* **95**, 14981–14989 (2023).
28. Shi, J. & Song, J. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *FEBS J.* **273**, 1035–1045 (2006).
29. Ferreira, J. C., Fadl, S. & Rabeh, W. M. Key dimer interface residues impact the catalytic activity of 3CLpro, the main protease of SARS-CoV-2. *J. Biol. Chem.* **298**, 102023 (2022).
30. Ho, B.-L. et al. Critical Assessment of the Important Residues Involved in the Dimerization and Catalysis of MERS Coronavirus Main Protease. *PLoS One* **10**, e0144865 (2015).
31. Albani, S. et al. Unexpected Single-Ligand Occupancy and Negative Cooperativity in the SARS-CoV-2 Main Protease. *J. Chem. Inf. Model.* **64**, 892–904 (2024).
32. Vuong, W. et al. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **11**, 4282 (2020).
33. Chan, H. T. H., Oliveira, A. S. F., Schofield, C. J., Mulholland, A. J. & Duarte, F. Dynamical Nonequilibrium Molecular Dynamics Simulations Identify Allosteric Sites and Positions Associated with Drug Resistance in the SARS-CoV-2 Main Protease. *JACS Au* **3**, 1767–1774 (2023).
34. Kabsch, W. XDS. *Acta Cryst. D.* **66**, 125–132 (2010).
35. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Cryst. D.* **69**, 1204–1214 (2013).
36. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
37. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Cryst. D.* **75**, 861–877 (2019).
38. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Cryst. D.* **60**, 2126–2132 (2004).
39. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix refine. *Acta Cryst. D.* **68**, 352–367 (2012).
40. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

Acknowledgements

We thank the staff of the ESRF beamline ID23-1 for help and assistance with data collection. This work was supported by funding from the CARIPARO Foundation ("Progetti di ricerca sul Covid-19" N. 55812), MUR PRIN bando 2017 (code: 2017BZEREZ_003), MUR PRIN bando 2022 (code: 2022Y5C7KL) and Department of Chemical Sciences at UNIPD (P-DiSC # BIRD2023-UNIPD).

Author contributions

D.V. and B.B. performed substrate peptide synthesis. E.F. performed cloning and crystallization of the inactive mutant of M^{pro} and wtM^{pro}. E.F. and S.F. performed expression and purification of the inactive mutant of M^{pro} and wtM^{pro}. E.F. and G.G. performed x-ray data collection. H.S. performed structure analysis. S.F. and A.S. performed native mass spectrometry experiments. S.F. and M.B. performed ITC experiments. J.E., S.F., and A.S. performed kinetic experiments. G.G., M.B., and R.B. supervised and analyzed the production and biophysical characterization of the inactive mutant of M^{pro} and wtM^{pro}. A.S. and B.B. supervised and analyzed nMS and kinetic measurement. A.S., B.G., and R.B. provided the funding. R.B. refined, analyzed and deposited the crystal structures, implemented the allosteric model, prepared the figures, wrote the manuscript, conceived and coordinated the project. All authors provided critical feedback and helped in the conduction of the research, in the analysis of the data, and provided the critical reading of the manuscript.

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/s42003-024-07138-w.

Correspondence and requests for materials should be addressed to Roberto Battistutta.

Peer review information *Communications Biology* thanks Andrey Kovalevsky, Jian Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Lei Zheng and Laura Rodríguez Perez. A peer review file is available.