



UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO SCIENZE DELLA VITA E DELL'AMBIENTE

Corso di Laurea Magistrale in Biologia Molecolare e Applicata

**ESPRESSIONE RICOMBINANTE, PRODUZIONE SU LARGA SCALA E ATTIVITÀ  
ENZIMATICA DELLA PROTEASI PRINCIPALE DI SARS-CoV-2**

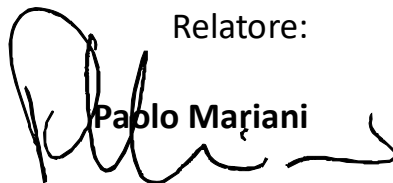
RECOMBINANT EXPRESSION, HIGH-SCALE PRODUCTION AND ENZYMATIC  
ACTIVITY OF SARS-CoV-2 MAIN PROTEASE

Candidata:

**Norhan Belhaj**

*Norhan Belhaj*

Relatore:

  
**Pablo Mariani**

Correlatore:

**Lucia Silvestrini**

*Lucia Silvestrini*

Sessione: Autunnale

Anno Accademico: 2020/2021



## Indice

<b>1. Scopo della tesi.....</b>	<b>6</b>
<b>2. Introduzione.....</b>	<b>8</b>
2.1 Il virus SARS-CoV-2 e le principali caratteristiche.....	9
2.2 La proteasi principale (M <sup>pro</sup> ) di SARS-CoV-2.....	12
2.3 M <sup>pro</sup> come target di inibitori.....	18
<b>3. Materiali e Metodi.....</b>	<b>21</b>
3.1 Terreni di coltura.....	21
3.2 Plasmide per l'espressione ricombinante di M <sup>pro</sup> .....	23
3.3 Ceppi microbici utilizzati.....	25
3.4 Trasformazione di Escherichia coli.....	26
3.5 Colture Batteriche.....	28
3.6 Estrazione di proteine totali.....	29

3.7 Cromatografia di affinità.....	31
3.7.1 GST-Glutatione.....	35
3.8 Dialisi.....	37
3.9 Quantificazione Proteine.....	38
3.10 SDS- PAGE.....	40
3.11 Western Blotting.....	45
<b>4. Risultati.....</b>	<b>48</b>
4.1 Trasformazione batterica del ceppo BL21 DE3 pLys con il plasmide pGEX-6p1- M <sup>pro</sup> .....	48
4.2 Espressione ricombinante di M <sup>pro</sup> .....	49
4.2.1 Terreni e Induzione.....	49
4.2.2 Valutazione dell'espressione di M <sup>pro</sup> in BL21(DE3)pLys e Rosetta <sup>TM</sup> (DE3).....	51
4.2.2.1 Estrazione M <sup>pro</sup> in ceppi Rosetta <sup>TM</sup> (DE3) e BL21(DE3)pLys...	51
4.2.3 Purificazione M <sup>pro</sup> .....	51

4.2.4 Dialisi.....	54
4.2.5 Western Blotting.....	55
4.3 Quantificazione Proteica.....	58
4.4 Rimozione His tag.....	59
4.5 Attività enzimatica M <sup>pro</sup> .....	61
4.6 Inibizione dell'attività enzimatica.....	63
<b>5. Discussione.....</b>	<b>66</b>
<b>6. Conclusione.....</b>	<b>68</b>
<b>7. Bibliografia.....</b>	<b>69</b>
<b>Appendici.....</b>	<b>73</b>

## 1. Scopo della tesi

Questa tesi si è svolta nell'ambito di un'attività di ricerca iniziata durante la recente pandemia da Covid-19 e svolta in collaborazione con le Università di Perugia e di Palermo e un gruppo del CNR di Perugia. Il tema di questa collaborazione era la ricerca di piccoli inibitori della proteasi principale di SARS-CoV-2 ( $M^{\text{pro}}$ ) che potessero bloccare l'attività e l'analisi del loro meccanismo di inibizione (in particolare, blocco della formazione del dimero di  $M^{\text{pro}}$ , che è stato riconosciuto come forma attiva della proteasi). La ricerca prevedeva la scelta mediante tecniche di *molecular docking* dei piccoli inibitori, lo studio dell'inibizione da loro prodotta sull'attività di  $M^{\text{pro}}$  e l'analisi mediante tecniche biofisiche dell'equilibrio monomero/dimero di  $M^{\text{pro}}$  in presenza e in assenza degli inibitori. Come si può immaginare, a monte di queste attività c'è la necessità di disporre di quantità relativamente grandi di  $M^{\text{pro}}$ , opportunamente purificata e attiva.

Questa tesi riguarda il lavoro da me svolto nell'ambito di questa collaborazione, e cioè lo sviluppo e la messa a punto di un protocollo di espressione proteica ricombinante della proteasi principale  $M^{\text{pro}}$  di SARS-CoV-2.

A parte essere riusciti a produrre le quantità necessarie di  $M^{\text{pro}}$  utili per poter effettuare l'analisi strutturale mediante tecniche SAXS (ovvero diffusione a

piccoli angoli dei raggi X) e per misurare l'attività enzimatica della proteina ricombinante, investigando il potenziale di inibizione delle piccole molecole selezionate, vari sono i risultati ottenuti. Prima di tutto, la piattaforma microbica utilizzata per la produzione di M<sup>pro</sup> ricombinante è stata *Escherichia coli*, ma sono stati testati due ceppi, BL21(DE3)pLys e Rosetta<sup>TM</sup>(DE3), arrivando a concludere che quantità consistenti di M<sup>pro</sup> in soluzione sono ottenute utilizzando in particolare il ceppo BL21(DE3)pLys. Per seconda cosa, sono stati applicati diversi protocolli per la purificazione della proteasi principale del SARS-CoV-2 mediante cromatografia di affinità, arrivando a selezionare le migliori condizioni che garantiscono stabilità, purezza e attività. Come terzo risultato si può osservare che la scelta dei migliori protocolli per la produzione e la purificazione di M<sup>pro</sup> su ampia scala ha portato ad una resa ottimale di 3-4 mg di proteina per litro di coltura batterica.

## **1. Introduzione**

Nel dicembre 2019 a Wuhan, in Cina, è stato isolato un nuovo virus, SARS-CoV-2, l'agente eziologico della pandemia da COVID-19 (Corona, Virus, Disease 19, l'anno di identificazione del virus) **(1)**.

SARS-CoV-2 condivide circa il 79,5% di omologia genomica con SARS-CoV mentre solo circa il 50% di somiglianza con MERS-CoV **(2)**.

La sequenza virale di questo nuovo Coronavirus ha un'omologia di circa il 76% rispetto al virus che causò la pandemia di SARS nel 2002/2003, dunque i due virus sono molto simili.

Il meccanismo principale di trasmissione del SARS-CoV-2 avviene per via aerea dopo il contatto con i casi sia sintomatici che asintomatici. In quest'ultimo caso rimane possibile, seppur molto meno frequente, la possibilità di trasmissione virale.

Il periodo di incubazione può raggiungere le due settimane e i sintomi sono inizialmente simili a quelli di una sindrome influenzale, che, in casi gravi, può essere responsabile di polmoniti acute e severe**(3)**.



Questo virus appartiene all'ordine dei Nidovirus a cui appartiene la famiglia dei coronavirus, suddivisa in quattro generi: Alfa, Beta, Gamma e Delta. Il SARS-CoV-2 appartiene al genere Beta.

E' stato identificato come serbatoio iniziale del SARS-CoV-2 il *Rhinolophus affinis* (4). Da qui il virus, attraverso un salto di specie, è arrivato all'uomo. Ad oggi invece non è ancora chiaro quale sia esattamente l'animale che ha fatto da intermediario della trasmissione di SARS-CoV-2 dal pipistrello all'uomo.

### ***2.1 Il virus SARS-CoV-2 e le principali caratteristiche***

Le particelle virali di SARS-CoV-2 hanno un diametro di ca. 60-140 nanometri e contengono un genoma costituito di RNA a singolo filamento a polarità positiva di 29.9 kb. L'RNA può essere direttamente tradotto nella cellula ospite una volta infettata (5).

L'RNA genomico di ogni particella virale di SARS-CoV-2 forma un complesso con la proteina del nucleocapside (N) che interagisce con la membrana virale durante l'assemblaggio del virione, giocando un ruolo chiave nella replicazione del virus (6.7).

Il genoma virale è composto da frammenti genici codificanti sia per proteine strutturali che non strutturali. Il filamento di RNA che costituisce il genoma

virale dirige la sintesi di due grandi poliproteine costituite da unità proteiche non ancora funzionali. Per la maturazione del virus e per l'assemblaggio di nuove particelle virali all'interno delle cellule infette, queste unità proteiche devono essere liberate e rese funzionali. Le proteine strutturali sono rappresentate dalle proteine S (Spike), M (Membrane), E (Envelope), N (Nucleocapsid) **(8)**.

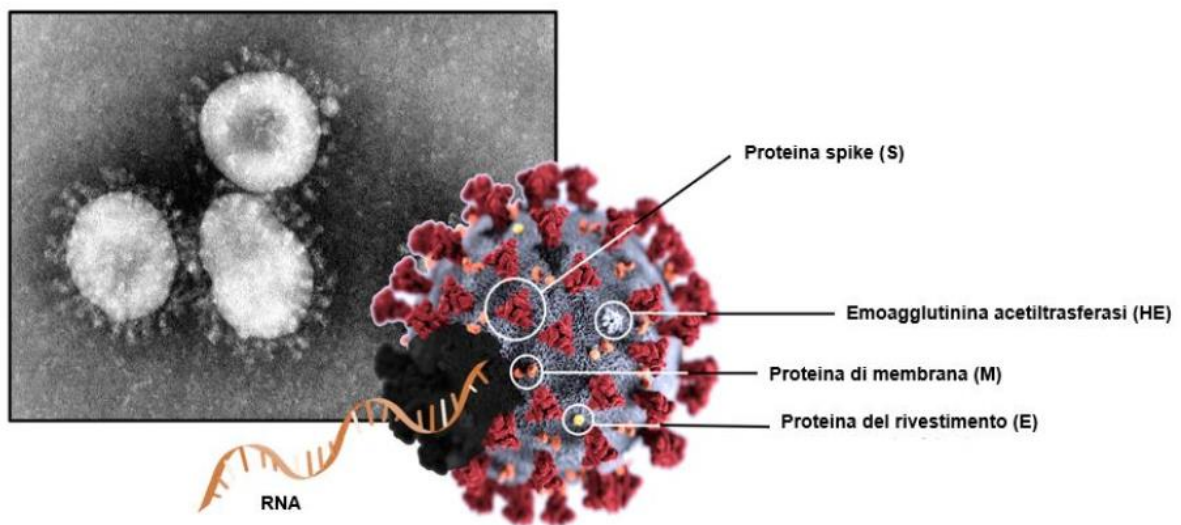
La proteina Spike (S) costituisce le protrusioni che caratterizzano la corona da cui deriva il nome di questi virus (coronavirus). La proteina S svolge un ruolo essenziale per l'attacco del virus al recettore umano ACE2, localizzato sulla cellula da infettare **(9)**.

All'interno della membrana virale sono localizzate altre proteine, tra cui la proteina E, che costituisce l'envelope virale, e la proteina M, fondamentale per poter legare il complesso ribonucleoproteico formato dalla nucleoproteina con l'RNA virale. Oltre a queste, è presente l'emoagglutinina esterasi (HE), una proteina del rivestimento che svolge una funzione importante durante la fase di rilascio del virus all'interno della cellula ospite **(10; Fig. 1)**. Oltre alle proteine strutturali, le poliproteine sono costituite da proteine non strutturali (NSP, Non-Structural Proteins) sono in totale 16 elementi. Esse derivano da due grandi proteine, chiamate poliproteina 1a (pp1a)

e poliproteina 1ab (pp1ab), le quali a loro volta sono codificate, rispettivamente, dai geni virali noti come replicasi 1a e replicasi 1ab.

Il processo di formazione delle proteine non-strutturali a partire dalle due poliproteine vede il coinvolgimento di due specifici enzimi virali, ossia la Papain-like (PLpro) e la Proteasi Principale 3CL<sub>pro</sub> (M<sup>pro</sup>, NSP5), prodotti precocemente dal virus che tagliato le poliproteine per dare origine alle singole Proteine non- strutturali; prima del taglio da parte delle due proteasi, le proteine ancora incluse nelle poliproteine sono inattive e non funzionali; queste diventano funzionali solo dopo l'intervento delle proteasi **(10)**.

La funzione principale delle proteine non-strutturali di SARS-CoV-2 riguarda la trascrizione e la replicazione dell'RNA virale.

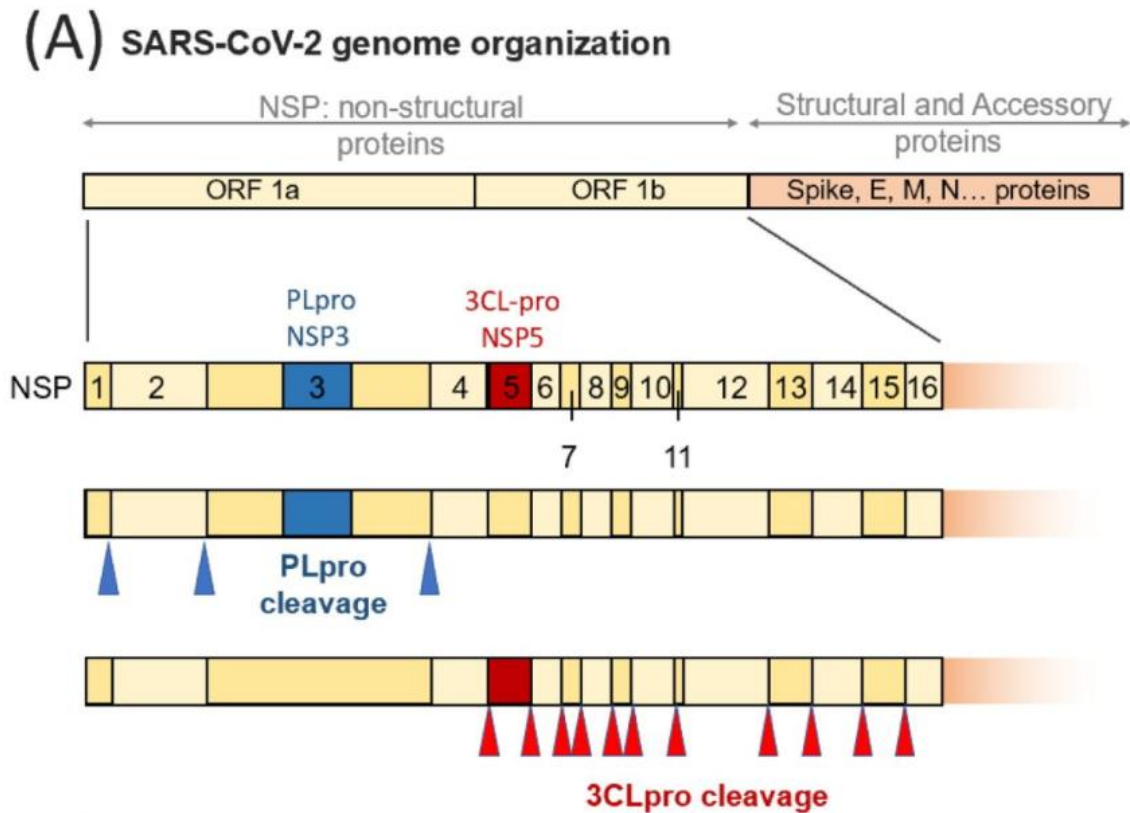


**Figura 1.** Immagine al microscopio elettronico a trasmissione, sono visibili le proteine Spike che “decorano” la superficie e che conferiscono al virione la caratteristica forma a corona. Il capside virale (in grigio), che contiene l’RNA a singolo filamento associato alle proteine del nucleocapside, è costituito da un doppio strato lipidico in cui sono inserite le proteine di superficie. (Immagine in alto a destra: [CDC/Dr. Fred Murphy](#); Sylvia Whitfield; disegno in centro adattato da: [CDC/Alissa Eckert, MSMI, Dan Higgins, MAMS](#)) (8).

## 2.2 La proteasi principale ( $M^{pro}$ ) di SARS-CoV-2

Durante l’infezione della cellula ospite, le unità proteiche vengono separate dalle poliproteine mediante l’attività proteolitica delle proteasi. In particolare, la proteasi principale  $M^{pro}$  taglia la poliproteina in 11 siti specifici. Le unità proteiche funzionali così ottenute sono necessarie alla maturazione del virus e all’assemblaggio di nuove particelle virali all’interno delle cellule infette **(10)**. PLpro di NSP3 è responsabile della scissione proteolitica di NSP tra 1 e 4,

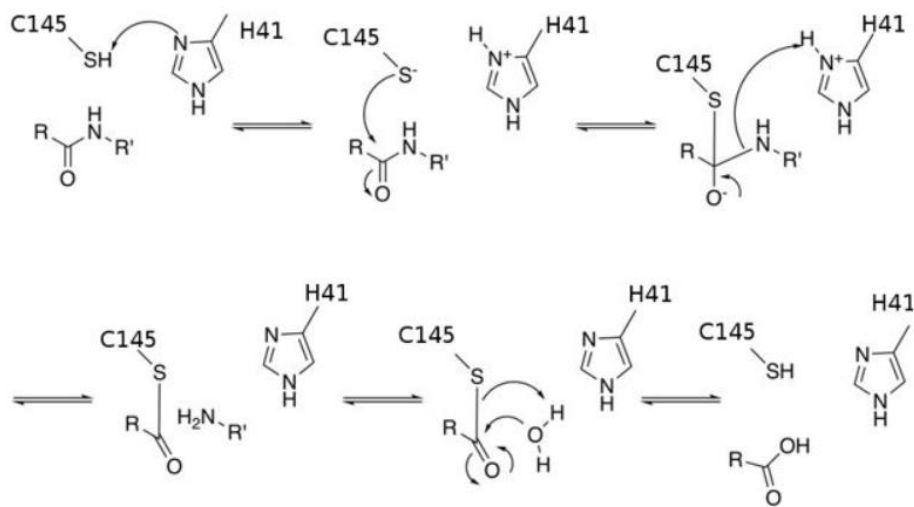
mentre la 3CLpro è responsabile del taglio proteolitico dei siti di clivaggio tra NSP 5 e 16 (11; Fig.2).



**Figura 2.** Schema dell'organizzazione del genoma di SARS-CoV-2; in ORF1a sono codificate due proteasi: NSP3 o proteasi simile alla papaina (PLpro) e NSP5 o proteasi simile a 3C (3CLpro). PLpro è responsabile di tre scissioni proteolitiche, mentre 3CLpro taglia la grande poliproteina in undici siti diversi (11).

Queste due proteasi giocano quindi un ruolo essenziale nel ciclo vitale di SARS-CoV-2.  $M^{pro}$ , che effettua il maggior numero di tagli sulla poliproteina, ha una taglia di 33,8 kDa ed è fondamentale per la replicazione virale. Per questa ragione  $M^{pro}$  rappresenta un buon bersaglio per lo sviluppo di nuovi

farmaci antivirali: il blocco delle sue funzioni sarebbe infatti letale per il virus, ma sicuro per gli esseri umani.  $M^{pro}$  è una cisteina proteasi, caratterizzata da una diade catalitica identificata dai residui His41 e Cys145. Il meccanismo di azione di  $M^{pro}$  è simile a quello di altre proteasi in cui un residuo accettore (di solito un'istidina, His) che rimuove protoni da un gruppo ossidrilico della catena laterale di un secondo residuo (di solito una serina, Ser, o una cisteina, Cys) che agisce come potente nucleofilo, ossia un potente donatore di elettroni (12;Fig.3).

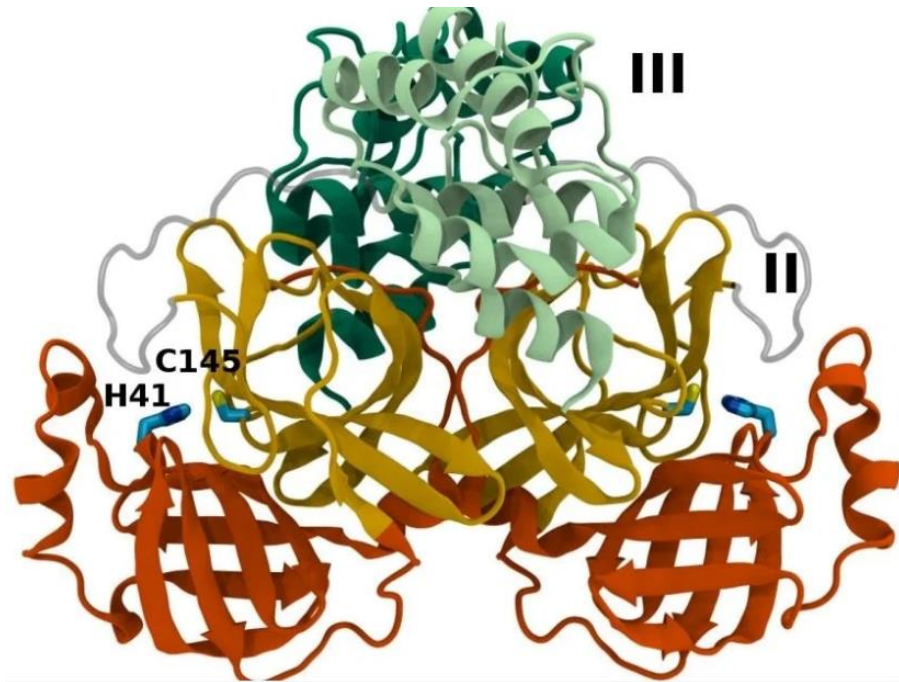


**Figura 3.** La catalisi inizia con la deprotonazione del tiolo della Cys145 da parte dell'His41, che è seguita dall'attacco nucleofilo della cisteina de-protonata al carbonio del legame peptidico della proteina substrato. Viene quindi rilasciata una proteina virale più piccola con terminale amminico libero, il residuo di istidina, His41, della proteasi recupera la sua forma de-protonata e si forma un intermedio tioestere che collega il nuovo carbossi-terminale del substrato al tiolo Cys145. Successivamente, si forma un tioestere

intermedio che lega il nuovo terminale carbossilico al tiolo della Cys145. Il tioestere è quindi idrolizzato per generare un terminale carbossilico sulla poliproteina rimanente, rigenerando l'enzima libero. I residui Cys145 (C) e His41 (H) coinvolti nella reazione catalitica. La posizione P1 corrisponde al residuo immediatamente prima del sito di taglio, seguito da P2, P3, P4, P5, fino alla parte *N*-terminale del sito di taglio (**12**).

L'analisi della struttura mostra che la proteina appare come dimero, formato da due subunità identiche di 306 aminoacidi (A e B). Ogni monomero è formato da tre differenti domini (**Fig.4**):

- dominio I (residui 1–101); chymotrypsin-like
- dominio II (residui 102–184) picornavirus 3C protease-like
- dominio III (residui 201–306) regulation of dimerization



**Figura 4.** Homodimer di  $M^{\text{pro}}$  con i tre domini illustrati ed il colore codificato come segue: Monomero verde chiaro/verde scuro del dominio I (arancione scuro), del dominio II (oro) e del dominio III (A/B) con i residui di elemento bivalente, il His41 e il Cys145 catalitici (13).

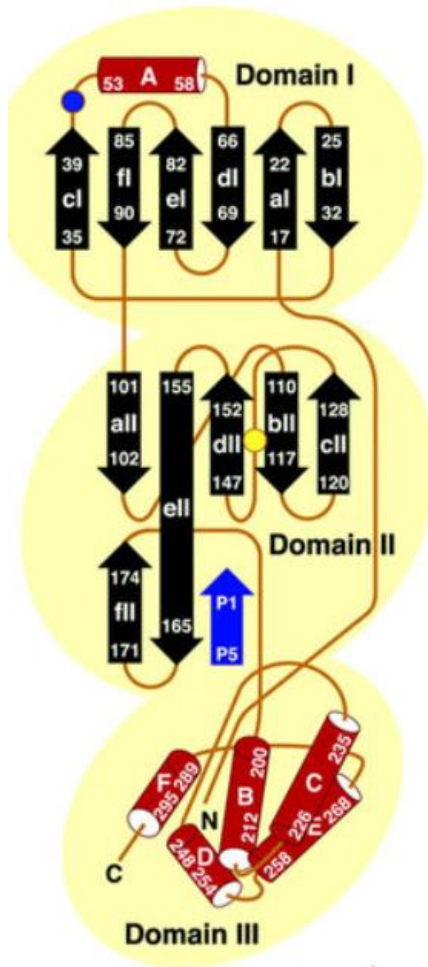
Le due molecole di  $M^{\text{pro}}$  formano un omodimero attivo funzionale coinvolti nella dimerizzazione. In tale omodimero, i due monomeri sono disposti quasi perpendicolarmente l'uno all'altro e ciascun monomero comprende la diade catalitica His41-Cys145 e il sito di legame al substrato situato in una fessura tra i domini I e II. (13).

I domini I e II hanno una tipica struttura a barile  $\beta$  in cui i filamenti  $\beta$  si dispongono in modo antiparallelo, mentre il dominio III è formato da cinque  $\alpha$ -eliche ed è direttamente coinvolto nel controllo della dimerizzazione



di M<sup>pro</sup> principalmente attraverso un ponte salino tra Glu290 di un monomero e Arg4 dell'altro. Il dominio III è unito al dominio II da un lungo ripiegamento (residui 185–200). *Codice di accesso in PDB: 6lu7*.

I singoli monomeri sono enzimaticamente inattivi mentre la forma attiva è quella dimerica. La ragione strutturale alla base della funzionalità del dimero è probabilmente dovuta all'interazione delle dita di zinco di ciascuno dei due monomeri con Glu166 dell'altro monomero, che stabilisce la forma della cosiddetta tasca S1 del sito di legame del substrato. Per avvicinarsi a questo sito di interazione, i residui di amminoacidi N-terminali vengono schiacciati tra i domini II e III del monomero e il dominio II dell'altro monomero.

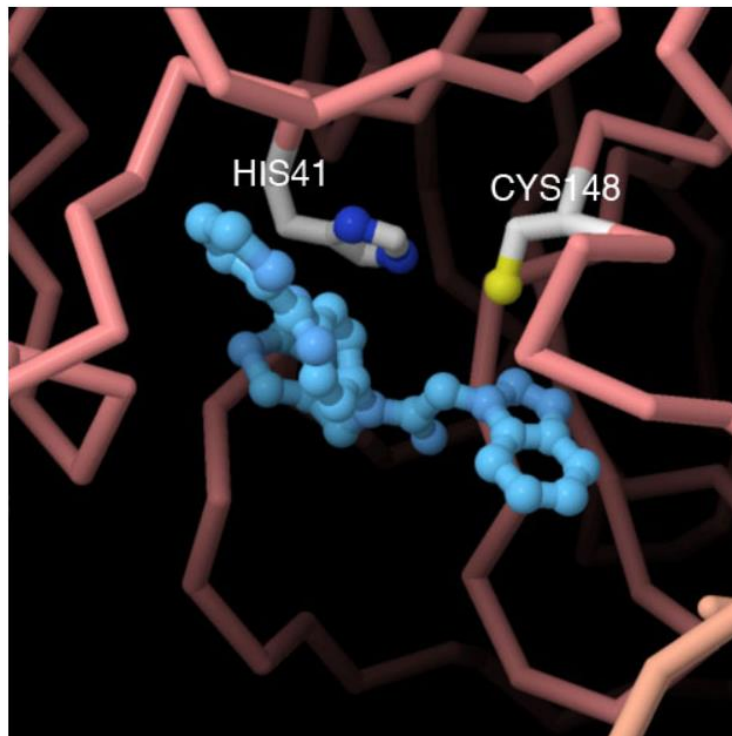


**Figura 5.** Rappresentazione topologica degli elementi della struttura secondaria di un monomero  $M^{pro}$ .  $\alpha$ -eliche e  $\beta$ -filamenti sono rappresentati rispettivamente come cilindri e frecce. I numeri indicano i residui N- e C-terminali degli elementi della struttura secondaria. Cys144 (giallo) e His41 (blu) sono indicati da cerchi. Sono indicate le posizioni dei terminali N e C. Inoltre, viene mostrata la presunta localizzazione della regione P5-P1 di un substrato modello (come riportato in blu).

### **2.3 M<sup>pro</sup> come target di inibitori**

La M<sup>pro</sup> di SARS-CoV, taglia le poliproteine pp1a e pp1ab in specifiche posizioni amminoacidiche, identificando i siti di taglio grazie a particolari “sequenze di base” nelle poliproteine **(14)**.

M<sup>pro</sup> è attiva nella forma dimerica. La funzionalità del dimero è probabilmente dovuta all'interazione del dito di zinco di ciascuno dei due monomeri con Glu166 dell'altro monomero, che stabilisce la forma della cosiddetta tasca S1 del sito di legame del substrato. Per avvicinarsi a questo sito di interazione, i residui amminoacidici N-terminali sono schiacciati tra i domini II e III del monomero e dominio II dell'altro monomero. Questo suggerisce che un potente inibitore di CoV-2 potrebbe essere un composto ad ampio spettro contro tutte le infezioni da coronavirus. Tale composto lega fortemente la cavità dell'M<sup>pro</sup> che normalmente alloggia il substrato, formando un legame covalente con la Cys145 **(15-17)**.

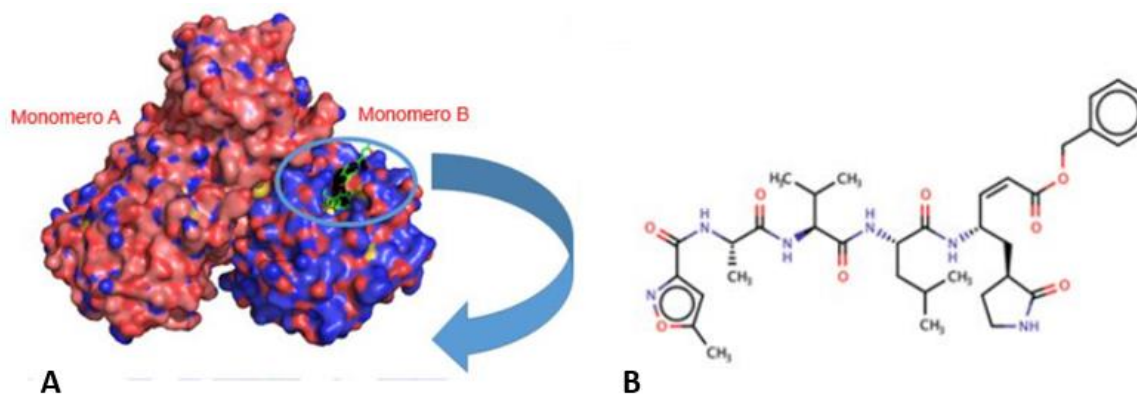


**Figura 6.** Il sito attivo della proteasi principale del coronavirus di pipistrello (HKU4). I residui amminoacidici cisteina e istidina sono mostrati mentre iniziano ad interagire con l'inibitore in azzurro) (18).

In uno studio hanno ottenuto la struttura cristallografica ad alta risoluzione della proteina M<sup>pro</sup> in complesso con N3 (Fig. 7, A), una molecola nota per legare e inibire la proteasi principale di altri coronavirus, come SARS-CoV e MERS-CoV (codice di accesso in PDB: [6lu7](#)). L'N3 è un inibitore simile a un oligopeptide (14.Fig.7,B).

L'analisi della struttura ha consentito di capire che l'inibitore N3 lega fortemente la cavità dell'M<sup>pro</sup> che normalmente alloggia il substrato, formando un legame covalente con la Cys145. L'analisi dell'interazione tra

N3 e i residui che rivestono la cavità hanno fornito importanti informazioni strutturali per progettare inibitori potenti e reversibili dell'M<sup>pro</sup>.



**Figura 7.** (A) Rappresentazione di superficie dell'omodimero di M<sup>pro</sup>. Gli atomi di carbonio dei due monomeri A e B sono colorati rispettivamente in rosa salmone e blu chiaro. Gli atomi di azoto sono colorati in blu, quelli di ossigeno in rosso e quelli di zolfo in giallo. L'inibitore N3 è rappresentato con la struttura a bastoncini ed è colorato in verde. (B) Formula strutturale dell'inibitore N3 (N-[(5-Metilisossazol-3-yl) carbonil]alanil-L-valil-N~1~((1R,2Z)-4-(Benzilossi)-4-osso-1-{[(3R)-2-ossopirrolidin-3-il]metil}But-2-enil)-L-Leucidammide (**14**).

### 3. Materiali e Metodi

#### 3.1 Terreni di coltura

Il terreno di Lysogeny broth (LB), è un terreno ricco per la crescita di batteri.

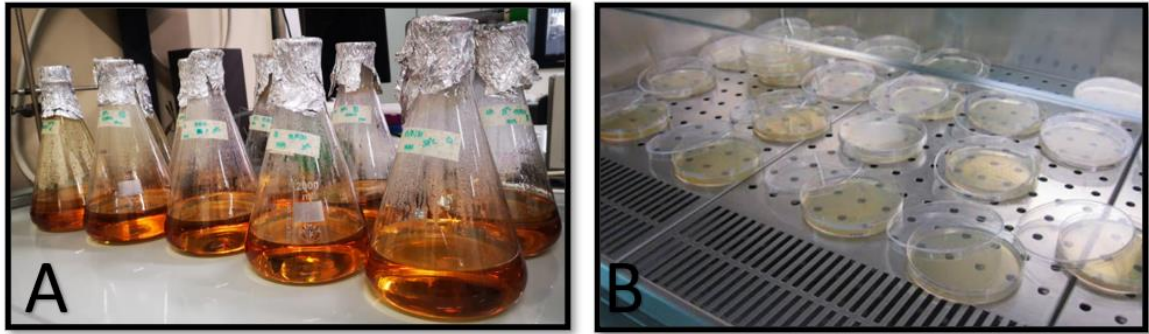
Il terreno può essere solido o liquido. Il terreno liquido è composto da:

Reagente	Quantità
Tryptone (LLG-Labware)	10g/L
Estratto di lievito (LLG-Labware)	5g/L
Cloruro di Sodio (LLG-Labware)	10g/L

**Tabella 1.** componenti Lysogeny Broth

Per il terreno solido viene aggiunto Agar 20g/L.

I terreni pesati vengono diluiti in un 1L di H<sub>2</sub>O distillata o deionizzata e si mescola con un **Agitatore magnetico con riscaldamento LLG-uniSTIRRER** (Fig. 9) fino ad ottenere una soluzione omogenea di colore giallo arancio.



**Figura 8.** Terreno LB per la crescita di *E. coli*. (A). LB liquido in beute da 2 L. (B). LB solido in piastre Petri durante l'asciugatura sotto cappa sterile a flusso laminare.



**Figura 9.** Agitatore Magnetico con riscaldamento LLG-uniSTIRRER 3 (A), Autoclave FVG (B).

I terreni vengono trasferiti in beute da 2L e autoclavati per circa 2 ore. Una volta raffreddato a temperatura ambiente si aggiungono gli antibiotici

specifici per il ceppo da coltivare. Gli stocks di antibiotici e induttore sono stati preparati alla concentrazione indicata in tabella 2.

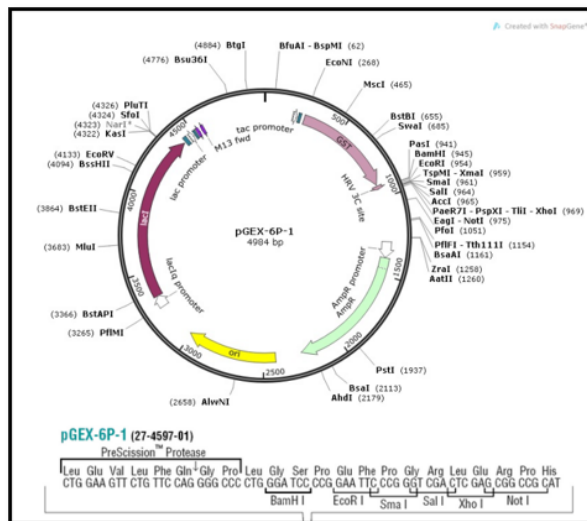
**Tabella 2.**

<b>Antibiotici e Induttore</b>	<b>Concentrazione stock 1000x</b>
Ampicillina	100 mg/ml
Cloramfenicolo	34 mg/ml
IPTG	1 M

### ***3.2 Plasmide per l'espressione ricombinante di M<sup>pro</sup>***

Il plasmide pGEX-6p1- M<sup>pro</sup> è stato acquistato da GenScript. Il plasmide è stato replicato nel ceppo di *E. coli* DH5a sia per tenere uno stock a -80°C sia per averne disponibile per continui utilizzi. Per l'espressione ricombinante di SARS-CoV-2 M<sup>pro</sup> il ceppo BL21DE3pLysS T1R è stato trasformato con il plasmide sopra citato.





Nome	pGEX6P1
Tipologia di plasmide	Batterico
Dimensioni	4984 bp
Tag	GST (N-term), Precision
Promotore	tac
Resistenza	Ampicillina e cloramfenicolo
Sito di clonaggio gene per Mpro	BamHI-XhoI

**Figura 10.** Mappa del plasmide pGEX-6p-1.

Il plasmide pGEX6p-1 si basa sul plasmide pBR322 comunemente utilizzato in *E. coli* (BL21(DE3)pLys), le cui caratteristiche principali sono:

- un'origine di replicazione
- un sito di clonaggio multiplo
- un marcatore selezionabile

SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDM  
 LNPNYEDLLIRKSNHNFVLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTP  
 KYK FVRIQPGQTF SVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGF  
 NIDYDCVSFCYMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTI  
 TVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDIL  
 GPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCS  
 GVTFQGP**HHHHHH**

**Figura 11.** Sequenza proteica della M<sup>pro</sup>, con il tag istidinico scritto in rosso presente al C-terminale.

### 3.3 Ceppi microbici utilizzati

Per l'overespressione di M<sup>pro</sup> illustrata in questo lavoro di tesi, *E. coli* è stato utilizzato come piattaforma microbica. In particolare, sono stati utilizzati due ceppi di *E. coli* comunemente impiegati per l'espressione di proteine ricombinanti e indicati in tabella 3.

**Tabella 3.**

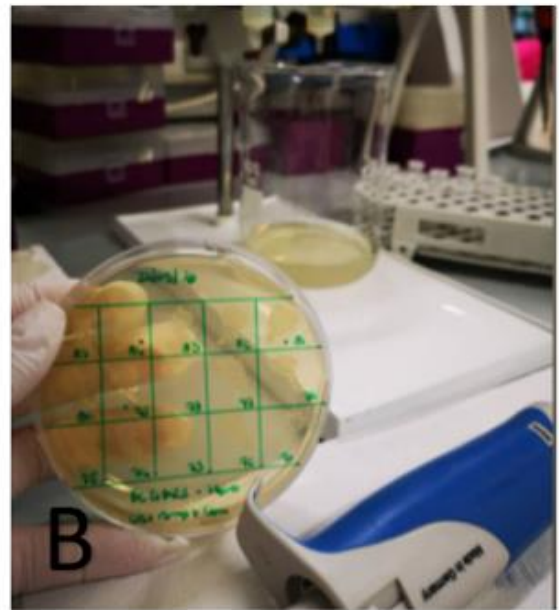
<b>Ceppo</b>	<b>Genotype</b>
<b>BL21 (DE3) pLys T1R (Promega)</b>	F <sup>-</sup> , <i>ompT</i> , <i>hsdS<sub>B</sub></i> ( <i>r<sub>B</sub></i> <sup>-</sup> , <i>m<sub>B</sub></i> <sup>-</sup> ), <i>dcm</i> , <i>gal</i> , $\lambda$ (DE3), pLysS, Cm <sup>r</sup>
<b>Rosetta<sup>TM</sup>(DE3) (Novagen)</b>	F <sup>-</sup> <i>ompT</i> <i>hsdS<sub>B</sub></i> ( <i>r<sub>B</sub></i> <sup>-</sup> <i>m<sub>B</sub></i> <sup>-</sup> ) <i>gal</i> <i>dcm</i> (DE3) pRARE (Cam <sup>R</sup> )

BL21(DE3) pLys-T1R competente *Escherichia coli* è un ceppo di espressione adatto per l'induzione di alto livello e l'espressione di geni da qualsiasi vettore di espressione basato sul promotore T7. La designazione DE3 indica che il ceppo è lisogenico per un profago lambda contenente una RNA polimerasi T7 inducibile, che è sotto il controllo del promotore lacUV5. L'espressione della T7 RNA polimerasi è indotta dall'aggiunta di IPTG alla coltura.

Le cellule hanno un'efficienza di trasformazione maggiore di  $1 \times 10^8$  cfu/ $\mu\text{g}$ , e il ceppo viene conservato a  $-80^\circ\text{C}$ .

### ***3.4 Trasformazione di Escherichia coli***

La trasformazione batterica avviene in condizioni sterili, sotto cappa a flusso laminare; una aliquota di cellule competenti BL21DE3pLys T1R e Rosetta<sup>TM</sup>(DE3) è stata scongelata in ghiaccio e sono stati aggiunti 1  $\mu\text{l}$  (ca. 40ng) di plasmide pGEX6p1- M<sup>pro</sup> (2.4). La sospensione è stata sottoposta a shock termico ad una temperatura di  $42^\circ\text{C}$  per 45 secondi (Thermoblock SBH130D, Fig.12C) e mantenute in ghiaccio per qualche minuto. Il terreno LB è stato poi aggiunto e la soluzione è stata incubata per un'ora a 225 rpm, (SL SHEL LAB; SI Sheldon Mfg, **Fig. 12A**).



**Figura 12.** Incubatore Shel Lab (A), Piastra con colonie replicate su piastre master (B), (C)Thermoblock usato per lo shock termico.

Una volta trascorso il tempo previsto, le cellule sono state piastrate su terreno LB solido contenente ampicillina e cloramfenicolo per la selezione

plasmidica. Le piastre sono state incubate a 37°C per 18h. Le colonie ottenute sono state testate per l'espressione di SARS-CoV-2 M<sup>pro</sup> su piccola scala.

Le colonie ottenute vengono poi replicate su piastre master (Figura 9B).

### ***3.5 Colture batteriche***

Le colture batteriche dei ceppi per l'espressione di M<sup>pro</sup> sono state condotte su terreno LB liquido. La biomassa inoculata e le colture di espressione sono state lasciate crescere ad una temperatura di 37°C in agitazione negli incubatori **Innova 44, Incubator Shaker Series (Fig. 13)**.

Dopo circa 3 h di crescita in agitazione a 37 °C viene effettuata una lettura allo spettrofotometro, De Novix DS-11 FX, (OD600) per determinare la quantità di cellule presenti nella sospensione batterica e stimare il grado di crescita della coltura. A valori di assorbanza compresi tra 0,6-0,8 viene indotta la produzione di M<sup>pro</sup> aggiungendo IPTG 0,5 mM nel ceppo BL21(DE3) pLys e Rosetta<sup>TM</sup>(DE3).

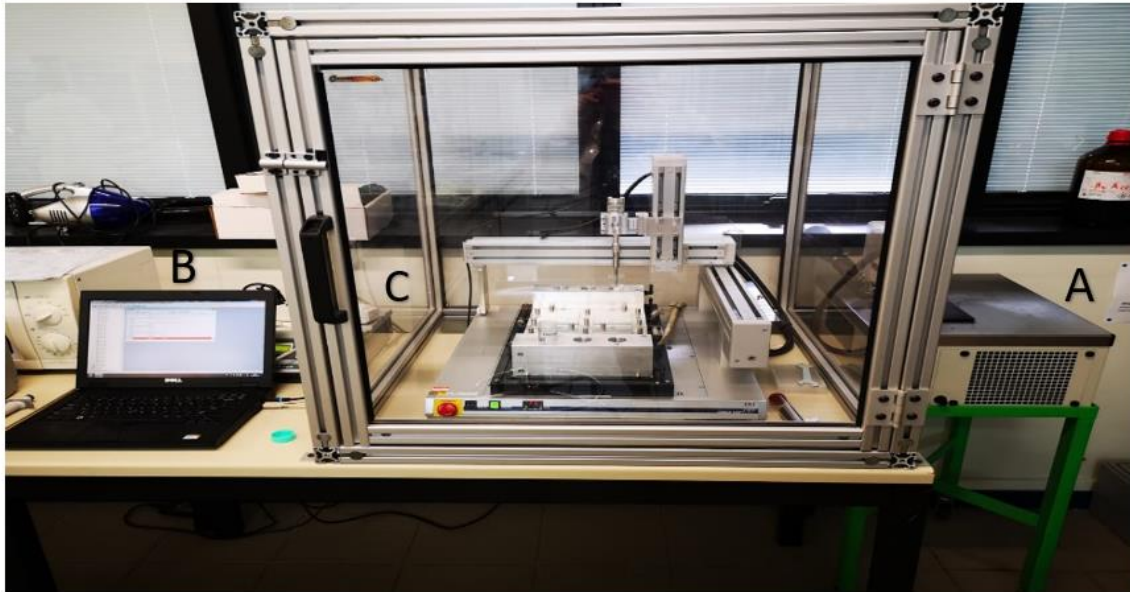


**Figura 13.** Innova 44, Incubator Shaker Series, utilizzato per la crescita microbica.

Trascorso un periodo di induzione di 3h, le cellule vengono recuperate mediante centrifugazione a 4000g per 20 min a 4 °C ed i pellet cellulari ottenuti vengono conservati a -80 °C.

### ***3.6 Estrazione di proteine totali***

La biomassa raccolta viene risospesa in Lysis Buffer (tabella 2.1) e sottoposta a 15 cicli di sonicazione (IAI Corporation, HAAKE Fisons C1, HAAKE Fisons K15, Sonics Vibraa Cella) da 5 secondi ciascuno, alternato da 5 secondi in ghiaccio ad ogni ciclo.



**Figura 14.** Sonicatore formato da: HAAKE Fisons C1(A) garantisce una temperatura uniforme nella camera, HAAKE Fisons k15, sonics vibra cell (B) emette ultrasuoni permettendo la lisi cellulare, IAI Corporation (C) la base nel quale viene inserito il campione trattato.

Il campione viene sottoposto a centrifugazione a 12.000g per 40 minuti a 4 °C. Il surnatante, contenente la frazione proteica solubile contenente la proteina di interesse, alla resina di Nickel e viene lasciato nell'incubatore rotante per 3 ore a 4°C.

### ***3.7 Cromatografia di affinità***

La purificazione di proteine mediante cromatografia di affinità “IMAC” (Immobilized metal ion Affinity Chromatography) o “cromatografia di affinità con ioni metallici immobilizzati”. Si basa sull’interazione tra la proteina di interesse e il metallo immobilizzato sulla resina. Nel caso di M<sup>pro</sup> l’interazione avviene tra la coda di istidine localizzata in C-terminale alla proteina e il nichel immobilizzato sulla resina. Il principio prevede l’eliminazione di ogni altra specie contaminante presente nell’estratto proteico totale e di concentrare la proteina di interesse per cui è fondamentale scegliere accuratamente il gruppo funzionale della resina costituente la colonna cromatografica: maggiore sarà la specificità di legame, migliore sarà l’efficienza del processo. Più in dettaglio, questa specificità è assicurata dalla presenza, nella proteina di interesse, di un tag di sei istidine il cui gruppo funzionale caratteristico è un anello di imidazolo. Questo eterociclo aromatico presenta una forte affinità per ioni metallici divalenti. Utilizzando una resina funzionalizzata con il Nichel in grado di legare le istidine attraverso il gruppo imidazolo, sarà possibile ottenere la proteina di interesse in forma pura.

La Resina utilizzata è Nickel Chelating Resin (GBiosciences), (**Fig.15**). La resina è conservata a 4°C in una soluzione contenente il 20% di etanolo. Prima dell’utilizzo, una aliquota di resina viene lavata 2 volte con il buffer di



lisi, in cui è risospeso l'estratto proteico, sia per rimuovere l'etanolo che per equilibrare la resina a 4°C (Tabella 2.1) per ca. 30 minuti. Una volta equilibrata, il buffer in eccesso viene rimosso e si aggiunge l'estratto proteico totale. La soluzione di resina ed estratto viene incubata a 4°C in rotazione per ca. 3 ore. Una volta trascorso il tempo necessario, la soluzione viene caricata su colonnina da banco e il non legato è lasciato fluire sfruttando la forza di gravità. Successivamente, la resina con la proteina immobilizzata è sottoposta a 3 lavaggi (10 volumi colonna per ciascuno) con 3 tamponi a diversa forza ionica (Tabella 2.2). Terminato il terzo lavaggio, la proteina viene eluita utilizzando un tampone contenente imidazolo ad elevata concentrazione (tabella 2.3).



**Figura 15.** (A) Nickel Chelating Resin (GBiosciences), resina utilizzata per la cromatografia di affinità, presenta una capacità di legare in 1ml 50mg di proteina con il His-tag; (B) La raccolta dei campioni: Il contenuto scende per gravità e la resina di impacchetta sul fondo della colonna dove sarà presente un filtro, ciò che passa viene detto flowthrough (FT). In seguito, si effettuano diversi lavaggi e in fine l'eluizione. Concluso il processo di purificazione, la resina verrà rigenerata e mantenuta in etanolo al 20% per evitare la disidratazione.

**Tabella 2.1** Tampone di lisi

Reagente	Concentrazione Finale
Tris HCl	0.02 M
NaCl (LLG-Labware)	0.3 M

**Tabella 2.2** Tamponi di Lavaggio

Reagente	Lavaggio 1	Lavaggio 2	Lavaggio 3
Tris HCl	20mM	20mM	20mM
NaCl	500mM	1M	300mM
Imidazole	10mM	10mM	

**Tabella 2.3** Tampone di eluzione

Reagente	Concentrazione Finale
Tris HCl	20mM
NaCl	100mM
Imidazole	300mM

**Tamponi per l'estrazione di M<sup>pro</sup> dai corpi inclusi nel ceppo Rosetta<sup>TM</sup>(DE3).**

**Tabella 2.4** Tampone di lisi, lavaggio 1 e lavaggio 2

Reagente	Concentrazione Finale
Triton x100	1%
Hepes	50mM a pH 8.0
NaCl	0.3M
Urea	1M

**Tabella 2.5** Tampone di denaturazione

<b>Reagente</b>	<b>Concentrazione Finale</b>
TRITON 100X	1%
Hepes	50 mM pH 8.0
NaCl	0.3 M
Urea	8 M

**Tabella 2.6:** Rigenerazione resina

<b>Reagente</b>	<b>Concentrazione</b>
MES	20mM pH 5.0 (10 CV)
H2O	20 CV
Etanolo 20%	10 CV e conservazione

### ***3.7.1 GST-Glutatione***

La cromatografia di affinità tra glutatione e glutatione-S-transferasi (GST) è stata utilizzata sia per la rimozione del prodotto di fusione GST- M<sup>pro</sup> non maturato in vivo sia per la rimozione della proteasi Prescission, fusa alla GST e utilizzata per la rimozione del tag di istidine da M<sup>pro</sup>. In questa fase è stata utilizzata una colonna Glutathione Sepharose Fast Flow (GSTrapFF) collegata

ad una siringa ed è stato seguito il protocollo di purificazione descritto dalla ditta produttrice (GE Healthcare).



**Figura 16.** GStrapFF – (GE Healthcare), colonna cromatografica di affinità tra glutatione-S-trasferasi (GST) e glutatione.

**Tabella 2.7.** Tampone di taglio della proteasi Prescission:

Reagente	Concentrazione Finale
TRIS HCl	50 mM
NaCl	150 mM
EDTA	1mM
DTT	1mM

**Tabella 2.8.** Tampone di eluizione per GST:

Reagente	Concentrazione Finale
TRIS HCl	50mM, pH 7.4
Glutatione	10mM

### **3.8 Dialisi**

La dialisi rappresenta una metodica molto utilizzata durante la fase di purificazione di proteine, consente di modificare il tampone in cui si trova la soluzione proteica di interesse oltre a permettere anche il refolding di proteine denaturate. In questo caso è stata utilizzata per rimuovere l'imidazolo dalle frazioni eluite dalla colonna cromatografica. Per il processo di dialisi è stata utilizzata una membrana in cellulosa rigenerata Spectrapor (cut-off 30 kDa) che viene prima immersa per qualche minuto in acqua deionizzata per eliminare il rivestimento di glicerina che viene utilizzato come agente umettante. La membrana, riempita con le frazioni contenenti di M<sup>pro</sup> eluite dalla colonna, è stata lasciata immersa ed in agitazione in una soluzione tampone di Dialisi (Tabella 2.9) per tutta la notte a 4 °C in modo da eliminare l'imidazolo presente in soluzione.

**Tabella 2.9** Tampone di Dialisi:

<b>Reagente</b>	<b>Concentrazione Finale</b>
TRIS HCl	50 mM, pH 7.6
NaCl	150mM

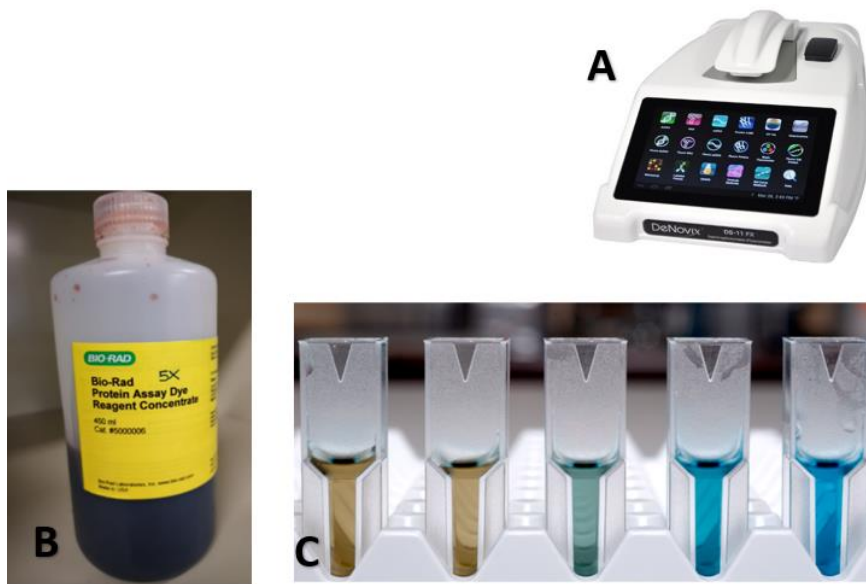
**Tabella 2.10** Tampone di Dialisi (Rosetta™(DE3))

Reagente	Concentrazione Finale
TRIS HCl	20 mM
NaCl	100 mM

### ***3.9 Quantificazione Proteine***

La concentrazione proteica è stata determinata sia mediante un metodo di Bradford sia mediante una valutazione spettrofotometrica a 280 nm (De Novix DS-11 FX).

Il metodo colorimetrico utilizzato è quello di Bradford (1976) impiegando il kit BioRadProtein assay (BioRad®) secondo le specifiche indicate dalla ditta ed utilizzando come riferimento la siero-albumina bovina (Sigma®). Per la valutazione spettrofotometrica invece è stata saggiata l'assorbanza dei campioni proteici alla lunghezza d'onda di 280 nm. (**19. Fig.17**). Come riferimento, è stata utilizzata la quantificazione generica in cui  $1A = 1 \text{ mg/ml}$  (Fig.17, A).

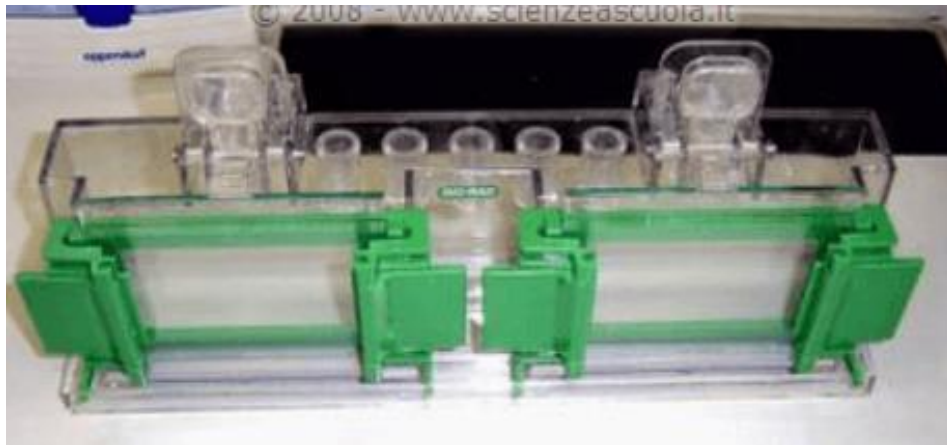


**Figura 17.** (B) Bio-Rad Protein Assay Dye, Reagent Concentrate; (C) Il colorante Coomassie Brilliant Blue G-250 esiste in tre forme: anionico (blu), neutro (verde) e cationico (rosso). In condizioni acide, la forma rossa del colorante viene convertita nella sua forma blu, legandosi alla proteina da saggiare. Se non ci sono proteine da legare, la soluzione rimarrà marrone; (A) L'assorbanza viene misurata allo spettrofotometro a 595 nm, De Novix DS-11 FX.



### 3.10 SDS- PAGE

L'SDS-PAGE è una tecnica analitica che permette l'analisi di preparati proteici, separando le proteine tramite elettroforesi in gel di poliacrilammide sulla base della loro massa molecolare. L'SDS è un forte agente denaturante delle proteine in quanto interferisce con le interazioni idrofobiche intramolecolari.



**Figura 18.** Vetrini fissati al supporto per gels di spessore pari a 0.75mm (Sistema **TETRA SYSTEM**).

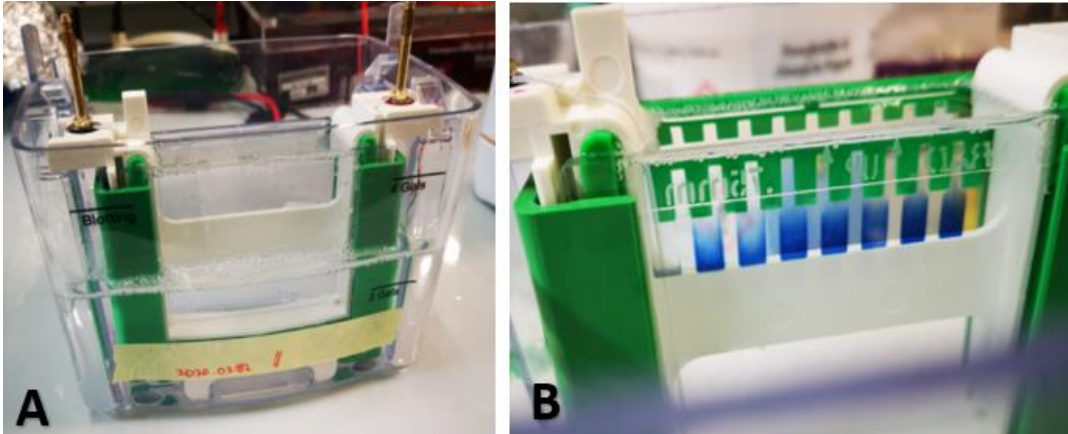
Il gel di poliacrilammide per l'SDS-PAGE è costituito da due parti:

Stacking gel: porzione superiore del gel che ha la funzione di concentrare il campione proteico caricato negli appositi pozzetti, creati dalla polimerizzazione della soluzione di acrilammide in presenza di pettini da 10

wells. In questo modo tutte le proteine iniziano la loro migrazione dallo stesso punto di partenza (Tabella 3.0. A).

Running gel: parte inferiore del gel ed ha la funzione di separare le proteine in funzione della loro massa molecolare. E' costituito dagli stessi componenti dello stacking gel ma con una quantità variabile di acrilammide in base alla taglia della proteina da separare (Tabella 3.0.B).

I campioni, prima del caricamento in gel, vengono trattati con Laemmli Sample Buffer (4×) (Tabella 3.2) a 100°C per 5 min. Nella camera da corsa si versa il running buffer (Tabella 3.1) e si caricano i campioni nei pozzetti. Si applica un potenziale di 80 V per 30', per la fase di stacking gel, ed 100 V per circa 60' per la fase running gel (Fig.19 A.B).



**Figura 19.** (A) La camera da corsa con i supporti posizionati e immersa nel tampone di corsa, (B) nei pozzetti vengono caricati i campioni che sono stati denaturati con il denaturante e mantenuti per 5 min a 100°C.

Conclusa la corsa elettroforetica, il gel viene colorato Stain (Tab. 3.3.A) e decolorato con il Destain (Tab. 3.3.B); L'SDS-PAGE GEL analizzato con il Chemidoc (Fig.21.B).

**Tabella 3.0** (A) Running gel al 12%; (B) Stacking gel

<b>Running Gel al 12%</b> <b>(A)</b>		<b>Stacking Gel 4%</b> <b>(B)</b>	
<b>Reagente</b>	<b>Quantità</b>	<b>Reagente</b>	<b>Quantità</b>
<b>H2O II</b>	1.34 ml	H2O II	1.5 ml
<b>ACRILAMMIDE</b> <b>30%</b>	1.60 ml	ACRILAMMIDE 30%	350 ul
<b>TRIS HCL 1.5M</b> <b>pH8.8</b>	1 ml	TRIS- HCL 0.5M pH6.8	625 ul
<b>SDS 10%</b>	40 ul	SDS 10%	25 ul
<b>APS 10%</b>	20 ul	APS 10%	25 ul
<b>TEMED</b>	2 ul	TEMED	2.5 ul
<b>Volume totale</b>	4ml	<b>Volume totale</b>	2.5 ml

**Tabella 3.1.** Il Running Buffer: TGS 10X

<b>Reagente</b>	<b>Quantità</b>
Tris base	30g
Glicina	144g
SDS	10g

**Tabella 3.2.** Laemmli Sample Buffer (4×)

<b>Reagente</b>	<b>Quantità</b>
Tris (1.0 M, pH 6.8)	5mL
SDS	2.0g
Glycerol	10mL
$\beta$ -Mercaptoethanol	5 mL
Bromophenol blue	0.05 g = 50 mg
H <sub>2</sub> O	25mL

**Tabella 3.3** Stain (A) e Destain (B)

<b>Stain (A)</b>		<b>Destain (B)</b>	
<b>Reagente</b>	<b>Quantità</b>	<b>Reagente</b>	<b>Quantità</b>
Methanol	500 ml	Methanol	250 ml
Glacial Acetic Acid	100ml	Glacial Acetic Acid	100ml
Comassie Brilliant Blue	1g	H <sub>2</sub> O	A vol 1L
H <sub>2</sub> O	A vol 1L		

### **3.11 Western Blotting:**

L'espressione proteica viene effettuata attraverso il western blotting, una tecnica che si basa sull'identificazione della proteina d'interesse con un anticorpo specifico. Il Western Blotting permette di evidenziare l'espressione proteica in una cellula. Per il trasferimento delle proteine, separate in SDS-PAGE, su membrana PVDF è stato utilizzato il tampone di trasferimento Trans-Blot Turbo Transfer buffer 5x (Biorad). La membrana con le proteine trasferite è ibridata con un anticorpo primario anti-istidina (6X-His Tag Monoclonal Antibody, Invitrogen). Un anticorpo secondario anti-Mouse HRP-coniugato è stato successivamente utilizzato per il legame con l'anticorpo primario e per il rilevamento del segnale chemiluminescente (Clarity™ Western ECL Substrate).

#### **Buffer usati per il Western Blotting:**

**Tabella 3.4:** Tampone di TBS

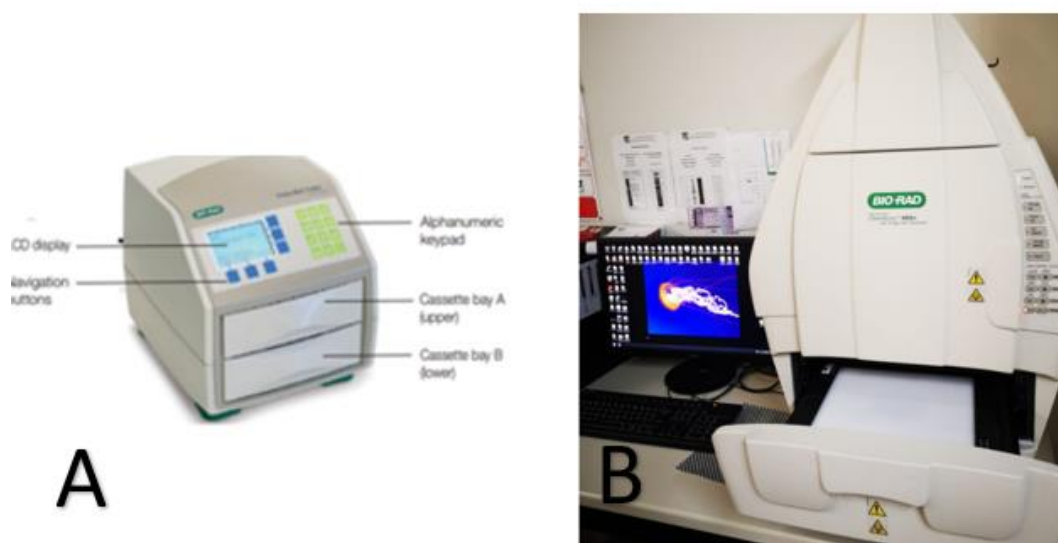
<b>Reagente</b>	<b>Concentrazione</b>
NaCl	137 mM
Tris HCl	0.02 M

**Tabella 3.5** Blocking solution TBS-T-BSA

<b>Reagente</b>	<b>Concentrazione</b>
TBS	Tab. 3.6
BSA	5 %
TWEN20	0,1 %

Per l'acquisizione delle immagini del Gel di acrilammide e del Western Blot viene utilizzato il ChemiDoc XRS+ System (Fig. 20B).





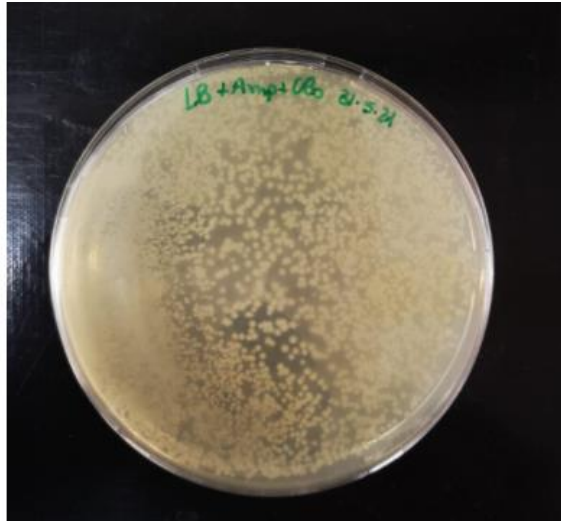
**Figura 20.** (A) BIO-RAD Trans-Blot Turbo Transfer System ed il (B) Chemidoc, usato per acquisire l'immagine del Western blot ed il gel di acrilammide.

## 4. Risultati

### 4.1 Trasformazione batterica del ceppo BL21(DE3)pLys con il plasmide pGEX-6p1- M<sup>pro</sup>

Il ceppo batterico utilizzato è il BL21 DE3 pLys, si tratta di cellule competenti di *Escherichia coli*, adatto per l'induzione di alto livello d'espressione di geni da qualsiasi vettore d'espressione basato sul promotore T7, è un ceppo lisogenico per un profago lambda (DE3) contenente un RNA polimerasi T7 inducibile che è sotto il controllo del promotore lacUV5. Le cellule hanno un'efficienza di trasformazione maggiore di  $1 \times 10^6$  cfu/ $\mu$ g. Il Frammento codificante per M<sup>pro</sup> è presente nel vettore Pgex-6p1 il quale verrà trasformato nelle BL21(DE3)pLys. Ad un'aliquota di cellule competenti si aggiunge 1ul (ca. 40ng) di Plasmide e si lascia per 10 min ghiaccio. Uno shock termico [42°C, 45 secondi; Termoblock (Fig. 12C)] permette la formazione di pori transienti sulle membrane delle cellule e l'ingresso del plasmide. Dopo lo shock termico si aggiungono 800ul di LB, e si lascia ad incubare in agitazione per 1h a 37 °C.

Le cellule trasformate sono poi seminate su piastra Petri contenente il terreno (LB) e gli antibiotici (Tab.2), per la selezione del plasmide pGEX-6p-1- M<sup>pro</sup> (Fig.10). In seguito, le piastre vengono incubate a 37°C over-night.



**Figura 21.** Piastra Petri con cloni di *E. coli* trasformati con il plasmide pGEX-6p1- M<sup>pro</sup>

Dalla piastra madre si replica in una piastra master le singole colonie (Fig.8B).

## ***4.2 Espressione ricombinante di M<sup>pro</sup>***

### ***4.2.1 Terreni e induzione***

Il terreno di coltura LB viene preparato secondo le quantità riportate in Tabella 1.

Il terreno viene sterilizzato in autoclave (Fig. 9B) e l'inoculo della biomassa viene effettuato sotto cappa sterile a flusso laminare per mantenere le condizioni di sterilità.

Il ceppo ricombinante di *E. coli*, prelevato dalla piastra master, viene inoculato in 10ml di LB+Amp+Clo (Tab.1, 2) e viene fatto crescere a 37°C per 16-18h in agitazione a 180 rpm.

Il giorno seguente la pre-coltura viene inoculata in 1L di LB+Amp+Clo (Tab.1; 2). La crescita microbica avviene a 37°C fino al raggiungimento di una densità ottica pari a  $OD_{600} = 0.6 - 0.8$ .

L'espressione ricombinante della  $M^{pro}$  viene indotta con IPTG 0,5 mM per 3 ore a 37°C, in agitazione (180 rpm). La biomassa viene poi raccolta mediante centrifugazione (4000-6000 rpm, 20 min, 4°C).

Oltre al ceppo BL21DE3pLys, è stato testato il ceppo Rosetta<sup>TM</sup>(DE3) per l'espressione proteica di  $M^{pro}$  in modo da comparare l'efficienza di espressione dei due ceppi di *E. coli*. In questo caso, dopo il raggiungimento della densità ottica precedentemente indicata, l'induzione è avvenuta per 16h a 16°C, in agitazione (180 rpm). Anche in questo caso le cellule sono state raccolte mediante centrifugazione.

## ***4.2.2 Valutazione dell'espressione di M<sup>pro</sup> in BL21(DE3)pLys e***

**Rosetta<sup>TM</sup>(DE3).**

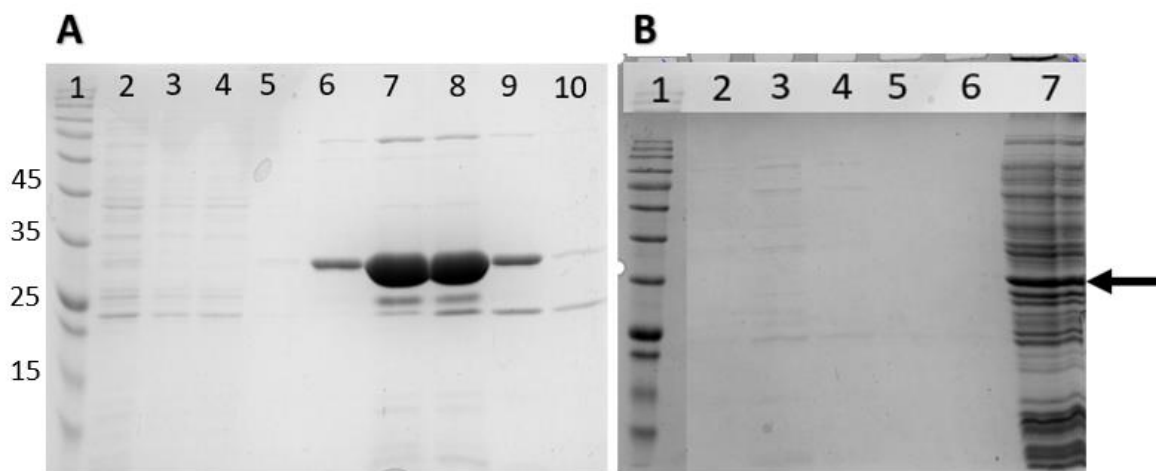
### ***4.2.2.1 Estrazione M<sup>pro</sup> in ceppi Rosetta<sup>TM</sup>(DE3) e BL21(DE3)pLys)***

Gli estratti proteici totali dalle cellule BL21(DE3)pLys e Rosetta<sup>TM</sup>(DE3) sono stati ottenuti mediante lisi della biomassa precedentemente raccolta. La lisi è stata effettuata mediante l'utilizzo del tampone di lisi sia per il ceppo BL21(DE3)pLys che per il ceppo Rosetta<sup>TM</sup>(DE3) (Tab.2.1). Una volta risospese le cellule nei tamponi precedentemente indicati, la biomassa è stata sonicata (sonicatore vibra cell Fig.14) per 15 cicli (2 secondi ognuno, AMP 60) e la sospensione è stata centrifugata per 1h a 4000 rpm in modo da separare i detriti cellulari dall'estratto proteico totale.

### ***4.2.3 Purificazione M<sup>pro</sup>***

L'estratto proteico totale viene incubato con la resina al Nichel a 4°C per 3h per permettere l'interazione tra il metallo e la coda di istidine in C-terminale ad M<sup>pro</sup>. Il processo di purificazione richiede l'utilizzo di colonne per la cromatografia, dove l'estratto proteico precedentemente incubato con la resina viene caricato sulla colonna stessa. Una volta che la resina si è impaccata per gravità e l'estratto proteico è passato attraverso le sue maglie,

si effettuano i lavaggi (Tab 2.2) e infine l'eluizione (Tab.2.3) in cui l'imidazolo compete per il legame del nichel all'istidina fusa alla M<sup>pro</sup>. Concluso il processo di purificazione, la resina verrà rigenerata come indicato nella tabella 2.4 e mantenuta in etanolo al 20% per evitare la disidratazione. Le singole frazioni proteiche eluite assieme a non-legato (FT; flowthrough) e lavaggi vengono poi quantificati a 280nm e successivamente vengono preparati i rispettivi campioni con il denaturante SB 4X (Tab.3.2) per la separazione elettroforetica in SDS-PAGE.



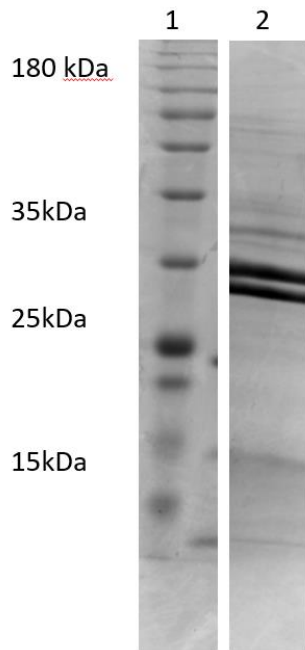
**Figura22.** (A) SDS-PAGE di BL21pLys+ M<sup>pro</sup>, Lane 1) standard proteico (prestained protein ladder sigma); lane 2, l'estratto totale; lane 3) ft; lane 4, lavaggio 1; lane 5, lavaggio 10; lane 6, eluato (EL) 1; lane 7, EL 2; lane 8, EL3; lane 9, EL4; lane 10, EL 5. (B) SDS-PAGE di Rosetta<sup>TM</sup>(DE3) + M<sup>pro</sup>, Lane 1, ladder; lane 2, el 1; lane 3, el 2; lane 4, el 3; lane 5, el 4; lane 6, el 5; lane 7, pellet. La freccia indica la presenza di M<sup>pro</sup> nei corpi inclusi e non in soluzione come invece avviene nel ceppo BL21(DE3)pLys (A).

Analizzando SDS-PAGE si è notato che  $M^{pro}$  in Rosetta<sup>TM</sup>(DE3) non era in soluzione, come in BL21(DE3)pLys, ma localizzata nei corpi inclusi (freccia in figura 22 B). Per questo motivo, è stata testata l'estrazione di  $M^{pro}$  dai corpi inclusi mediante l'utilizzo di urea come agente denaturante.

Il pellet ottenuto dal ceppo Rosetta<sup>TM</sup>(DE3) è stato lavato 2 volte con un tampone contenente U1M (Tab.2.4), per rimuovere  $M^{pro}$  eventualmente adesa alla membrana dei corpi inclusi. Il sovrantante è stato recuperato ogni volta per essere analizzato in SDS-PAGE.

Per l'estrazione definitiva delle proteine dai corpi inclusi, un tampone U8M (Tab.2.5) viene poi aggiunto al pellet ottenuto, in modo da rompere la membrana dei corpi inclusi e la soluzione è lasciata in agitazione per un massimo di 16h a temperatura ambiente. Il giorno seguente, l'estratto viene recuperato mediante centrifugazione a 14000 rpm per 10 min.

Il prodotto finale contenente  $M^{pro}$  estratta dai corpi inclusi è mostrato nel gel riportato in figura 23.



**Figura 23.** SDS-PAGE, lane 1, ladder; lane 2 M<sup>pro</sup> U8M, in cui si è evidenziata la presenza di M<sup>pro</sup> nei corpi inclusi.

Rosetta<sup>TM</sup>(DE3) - M<sup>pro</sup> derivante dai corpi inclusi non ha conformazione nativa ma risulterà misfoldata.

#### **4.2.4 DIALISI**

La dialisi viene effettuata utilizzando una membrana in cellulosa rigenerata, Spectrapor, che presentano un cut-off di circa 30kDa in maniera tale da non permettere il passaggio nella M<sup>pro</sup> ma solo delle molecole con le dimensioni inferiori di 30kDa. La dialisi avviene a 4°C per 18h in agitazione. Nel ceppo BL21(DE3)pLys viene effettuata per l'eliminazione dell'imidazolo, per evitare precipitazione della proteina (Tab.2.9). Mentre nel



ceppo Rosetta™(DE3) la dialisi viene effettuata per la rimozione della concentrazione elevata dell'urea, usata nei passaggi precedenti, ed il refolding della M<sup>pro</sup> (Tab.2.10).

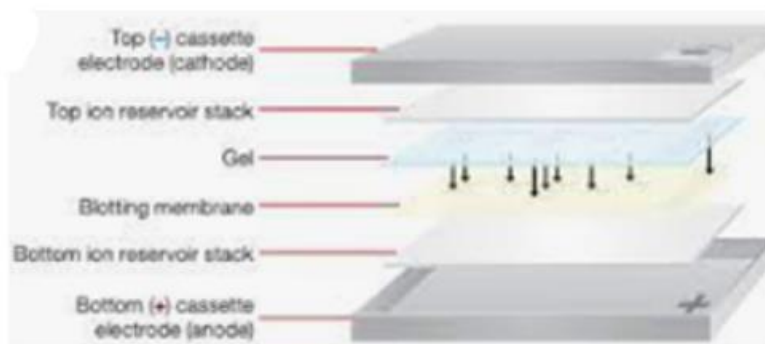
#### ***4.2.5 Western Blotting***

Il Western blot è stato utilizzato per evidenziare l'espressione della nostra proteina d'interesse.

Sono stati effettuati due SDS-PAGE: uno verrà utilizzato per la colorazione con il Comasse Blue Brilliant mentre l'altro per il Western Blotting

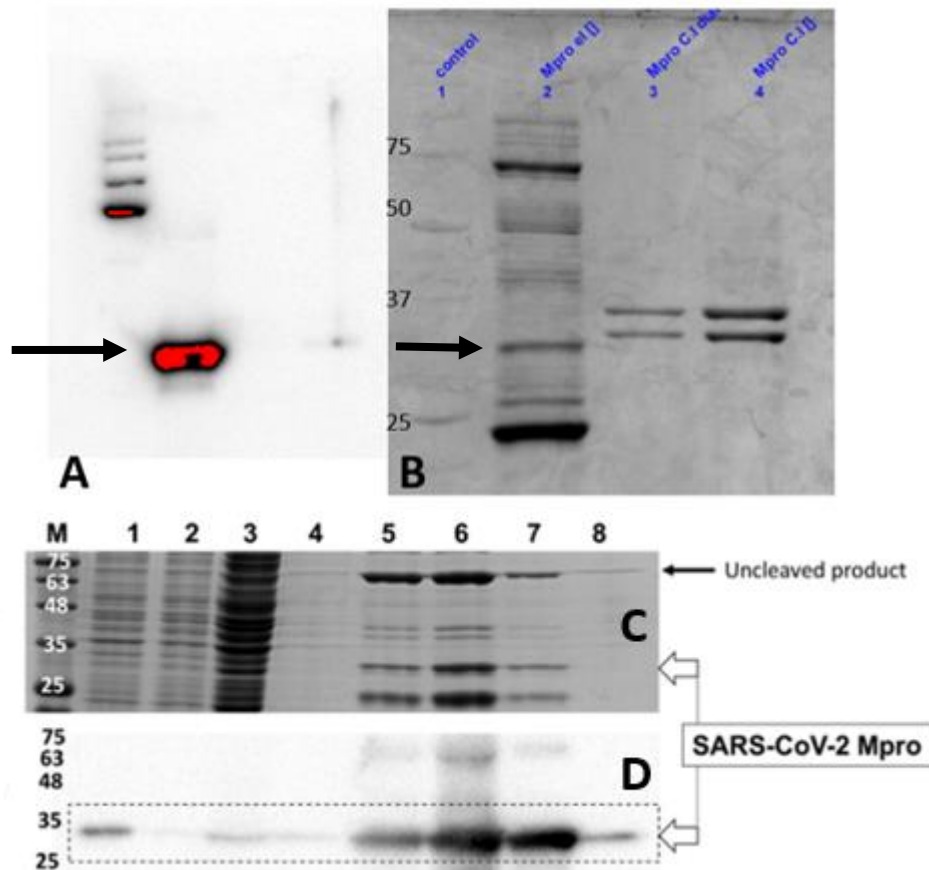
Quest'ultimo viene bagnato con TBS (Tab. 3.4); Viene rimosso il TBS nel quale è immersa la membrana e il gel, e viene sostituita con il Tampone di trasferimento.

Si prepara a sandwich nella cassetta (Fig 24) nel seguente ordine:



**Figura 24.** Disposizione a sandwich, foglio spesso, membrana di trasferimento, gel, foglio spesso e si passa con un rullo sopra per eliminare le bolle formate.

Si inserisce nel TransBlot Turbo si imposta Mixed MW 1.34°-25V-7M. il gel verrà immerso nel TBS e lasciato in agitazione per 5 min. Il TBS verrà in seguito sostituito con il TBS-T-BSA (Tabella 3.5) per 1 h in agitazione a temperatura ambiente. Viene rimosso in seguito il TBS-T-BSA in cui è immerso il gel aggiungendo TBS-T-BSA + l'Anticorpo e si lascia a 4°C in agitazione per 18h. Il giorno seguente si effettuano più lavaggi con la TBST e si rileva l'espressione proteica con il ChemiDoc, (Fig.20 B) sulla base di un legame con uno specifico anticorpo.

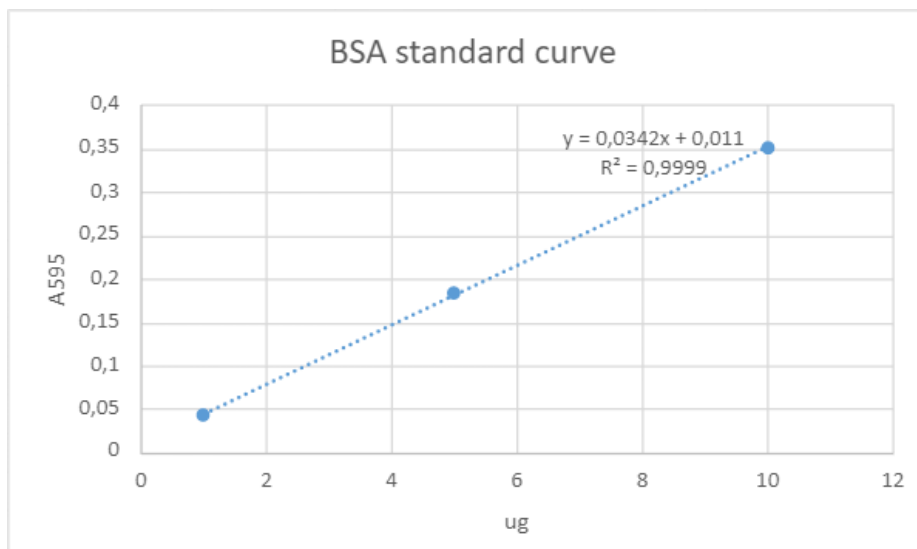


**Figura 25.** (A) WB in cui si osserva il legame dell'anti-mouse con l'his, Rosetta<sup>TM</sup>(DE3) + M<sup>pro</sup> (B) con il rispettivo SDS-PAGE, Rosetta<sup>TM</sup>(DE3) + M<sup>pro</sup>; (C) SDS-PAGE BL21(DE3)pLys + M<sup>pro</sup>; (D) WB BL21(DE3)pLys + M<sup>pro</sup> (20) Mpro è indicata con la freccia e presenta un peso di 30kDa.

Dall'immagine del gel mostrata in figura 25C, è possibile notare la presenza del prodotto di fusione GST- M<sup>pro</sup> non ancora processato durante il meccanismo di autocatalisi che avviene in *E.coli* mentre la proteina ricombinante viene prodotta (20-21).

### ***4.3 Quantificazione proteica***

Il dosaggio delle proteine Bradford si basa sul metodo Coomassie Blue G modificato. Il colorante forma un complesso blu con le proteine e l'intensità del colore misurata a 595 nm è direttamente proporzionale alla concentrazione di proteine nel campione. La proteina utilizzata come standard è l'albumina sierica bovina (BSA). Si Pipetta un campione di 20 µl della soluzione da analizzare in una provetta di reazione, quindi aggiungi 800 µl di acqua deionizzata e 200 µl di colorante blu brillante Coomassie. In presenza di proteine, la soluzione diventerà quasi immediatamente di colore blu ciò può essere misurato quantitativamente con un fotometro a 595 nm. Per il bianco, invece, si aggiunge 800ul di acqua a 200ul di Bradford, che verrà letto prima allo spettrofotometro (Fig.17A).



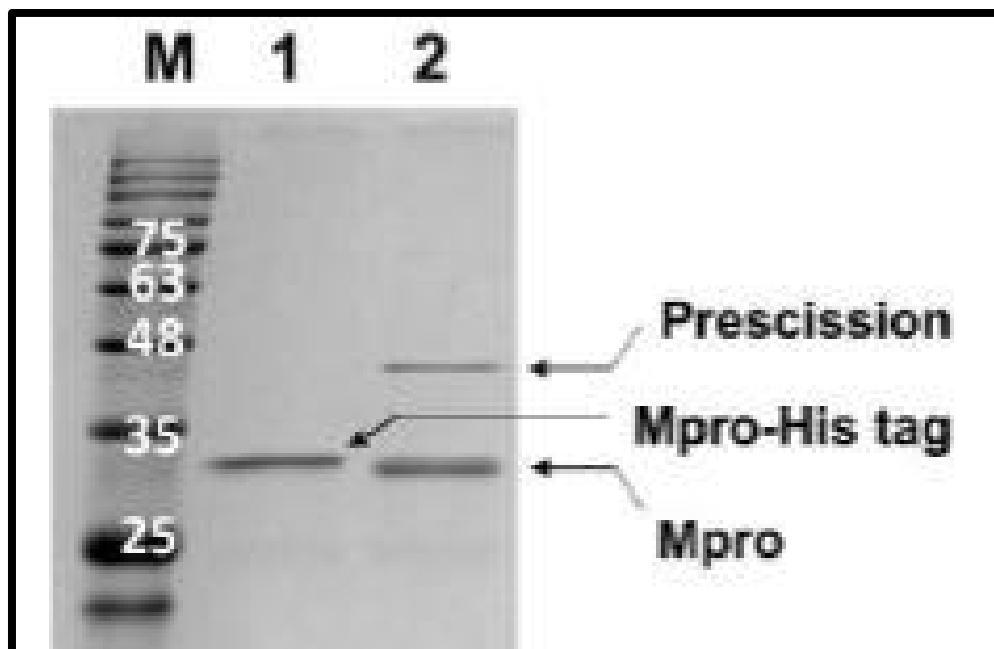
**Figura 26.** Vi è riportata la curva standard della BSA, dove in ascisse è presente la quantità di BSA e in ordinate la rispettiva lettura a 595 nm. L'equazione della retta è indicata così come l'R quadro.

#### ***4.4 Rimozione His tag***

La proteasi Prescission, utilizzata in questo lavoro di tesi per la rimozione del tag di istidine da M<sup>pro</sup>, è un enzima che presenta una fusione con GST in N-terminale. 1U digerisce 100μg di soluzione proteica e 1U equivale a 0.5ul. La reazione di rimozione dell'His tag avviene mediante incubazione della soluzione proteica con la corrispondente quantità della proteasi Prescission per 4h a 4°C. La colonna GSTrapFF è stata usata per la rimozione della proteasi Prescission dalla soluzione proteica una volta terminata la reazione di taglio dell'His tag. La colonna è stata lavata con PBS 1x a pH 7.4 ed è stata

caricata la soluzione proteica. Il non legato è stato raccolto poiché contiene la soluzione con M<sup>pro</sup> ma senza la proteasi Prescission che è rimasta attaccata alla colonna. Dopo alcuni lavaggi con PBS 1x, si aggiunge tampone di eluizione (TRIS HCl 50mM pH 7.6 + glutatione 10mM) per ripulire la colonna. La colonna è stata poi conservata a 4°C in etanolo al 20 %.

Del campione raccolto se ne prepara un campione per SDS-PAGE.



**Figura 27.** SDS-PAGE che mostra la rimozione attraverso la Prescission delle tag di 6 His all'estremità del C-terminale della M<sup>pro</sup>.

#### 4.5 Attività enzimatica di M<sup>pro</sup>

Una volta prodotta la proteina ricombinante M<sup>pro</sup> in *E. coli* è necessario valutare se M<sup>pro</sup> è ripiegata in modo corretto e quindi se è enzimaticamente attiva. L'attività enzimatica di M<sup>pro</sup> è stata monitorata sia per la proteina con il tag di istidine, sia nella proteina in cui il tag è stato rimosso.

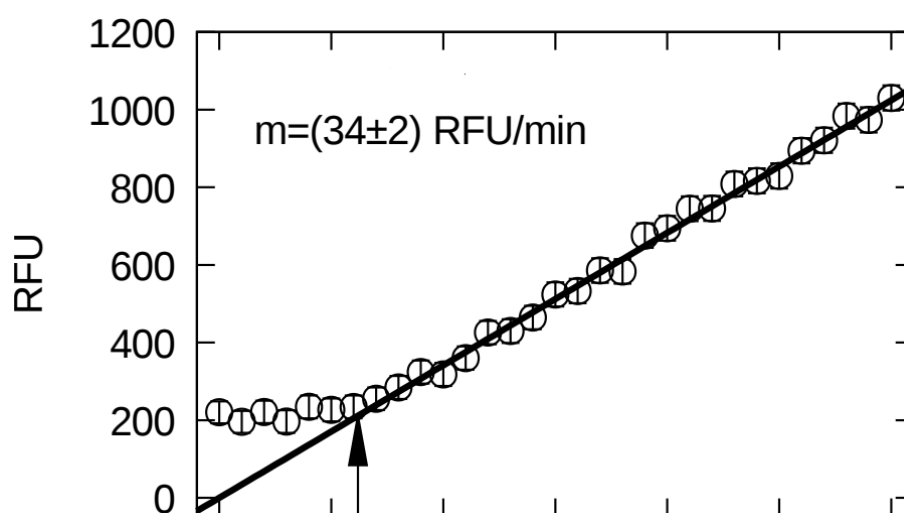
A tal fine è stato utilizzato il substrato specifico per M<sup>pro</sup> costituito da 9 residui amminoacidici MCA- AVLQ↓SGFRK -DnpK [(MCA=7-Methoxycoumarin-4-yl)acetyl];Dnp= 2,4-dinitrophenyl; GenScript].

Se attiva, M<sup>pro</sup> taglia tra i residui Q ed S liberando un segnale fluorescente che viene quantificato nel tempo tramite fluorimetro (Fig.17A).

L'attività enzimatica a 30°C corrispondente ai valori di fluorescenza liberata (eccitazione 320 nm, emissione 405 nm) è stata registrata ogni minuto per un totale di 20 minuti. I dati di fluorescenza sono rilevati da campioni composti da M<sup>pro</sup>, il substrato e il tampone con un volume finale di 200ul. Ai valori di fluorescenza raccolti verranno poi sottratti i valori del bianco, ugualmente raccolti nel tempo (Tab.4.0).

**Tabella 4.0**

<b>Bianco</b>	<b>Mix dell'attività</b>
Tris HCl 50 mM a pH 7.6	Tris HCl 50 mM a pH 7.6
Substrato specifico per M <sup>pro</sup>	Substrato specifico per M <sup>pro</sup>
	M <sup>pro</sup>



**Figura 28.** L'attività enzimatica, dove nell' ascissa è riportato il tempo e nell'ordinata è riportato RFU (Relative Fluorescence Units) (20).

Per verificare l'integrità di M<sup>pro</sup> dopo congelamento a -80°C e scongelamento, l'attività proteica è stata misurata come precedentemente indicato. L'attività proteica è stata monitorata utilizzando diverse quantità di proteina (Appendice I).

L'attività enzimatica è stata monitorata anche nel ceppo Rosetta<sup>TM</sup>(DE3) + M<sup>pro</sup>, mostrando che è inferiore al ceppo BL21DE3pLys+ M<sup>pro</sup>, dato che non

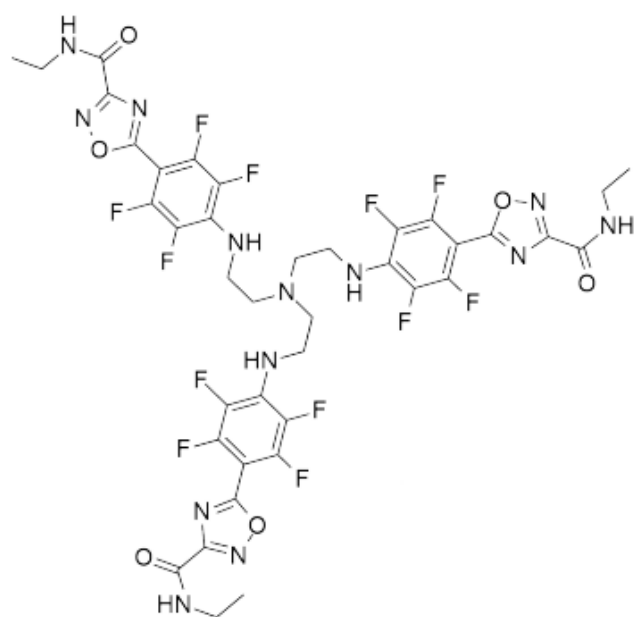


presenta una conformazione nativa e di conseguenza non sarà attiva (Appendice I).

#### ***4.6 Inibizione dell'attività enzimatica***

I test di attività enzimatica quantificano sia la capacità catalitica di SARS-CoV-2 M<sup>pro</sup> sia la presenza del dimero. Al fine di bloccare l'attività catalitica di M<sup>pro</sup> è fondamentale valutare gli effetti di potenziali inibitori che agiscono sull'interfaccia di dimerizzazione poiché il monomero risulta cataliticamente inattivo (15). In base a ciò, sono stati progettati *in silico* dei potenziali inibitori del sito catalitico e della dimerizzazione di M<sup>pro</sup>.

A tal fine è stato testato l'inibitore 7 NV 1399(Fig.29).



**Figura 29.** Inibitore NV 1399 (7) in silico, testato per inibire l'attività catalitica di  $M^{pro}$ .

L'inibizione dell'attività enzimatica è stata registrata in base ai valori di fluorescenza liberata (eccitazione 320 nm, emissione 405 nm) ed eseguita a 30°C monitorando ogni minuto per un totale di 20 min. Per l'inibizione dell'attività catalitica di  $M^{pro}$  la mix di reazione è stata preparata come indicato in Tabella 4.1.

**Tabella 4.1 Mix per l'inibizione**

<b>Bianco</b>	<b>Mix</b>
Tris HCl 50Mm a pH 7.6	Tris HCl 50Mm a pH 7.6
Substrato specifico per M <sup>pro</sup>	Substrato specifico per M <sup>pro</sup>
INIBITORE n°7	M <sup>pro</sup>
	INIBITORE

## 5. Discussione

M<sup>pro</sup> è stata espressa con rese differenti in due diversi ceppi di *E. coli*, osservando che la biosintesi di questa proteina varia dal ceppo BL21(DE3)pLys e dal ceppo Rosetta<sup>TM</sup>(DE3). M<sup>pro</sup> è stata ingegnerizzata in modo da avere un tag istidinico all'estremità C-terminale che ha permesso di purificarla mediante cromatografia di affinità e di rilevarne l'espressione mediante Western blot e SDS-page. Nei due ceppi testati, l'espressione è risultata differente: in BL21(DE3)pLys M<sup>pro</sup> viene prodotta in soluzione mentre nel ceppo Rosetta<sup>TM</sup>(DE3) è localizzata nei corpi inclusi. Questo risultato ha permesso di scegliere il ceppo migliore per avviare la produzione di M<sup>pro</sup> su ampia scala, alla base di qualsiasi studio strutturale di caratterizzazione proteica. Al fine di raccogliere dati strutturali mediante tecniche di caratterizzazione proteica come dicroismo circolare (effettuato dall'Università degli studi e CNR di Perugia) e SAXS (Small Angle X-Ray Scattering) della proteina in soluzione, una quantità di proteina pari a ca. 100 mg è stata prodotta per via ricombinante durante questo lavoro di tesi.

Al fine di valutare il corretto ripiegamento della struttura proteica sono stati effettuati dei saggi di attività enzimatica che, come indicato nel paragrafo 4.5, hanno dimostrato che la proteina M<sup>pro</sup> prodotta per via ricombinante risulta

essere attiva e in grado di tagliare il suo substrato specifico MCA-AVLQ↓SGFRK -DnpK [(MCA=7-Methoxycoumarin-4-yl)acetyl];Dnp= 2,4-dinitrophenyl; GenScript]. La misura della fluorescenza liberata durante l'attività enzimatica di M<sup>pro</sup> ha permesso di verificarne lo stato catalitico e di compararne l'efficienza con lo stesso saggio effettuato in presenza di un inibitore specifico per M<sup>pro</sup> stessa. L'inibitore NV1399 (sintetizzato dal dipartimento STEBICEF dell'Università degli Studi di Palermo) è risultato efficiente nell'inibire l'attività proteolitica di M<sup>pro</sup> verso il suo substrato. Questa evidenza supporta il principio di utilizzare M<sup>pro</sup> come target specifico di nuovi inibitori al fine di bloccare la replicazione virale.

## 6. Conclusioni

La produzione su larga scala della proteasi principale del SARS-CoV-2 ha assunto un tempo di notevole importanza.

È stato individuato un efficace protocollo di espressione ed estrazione della proteina da *Escherichia coli*. In particolare, M<sup>pro</sup> viene espresso con rese elevate in soluzione nel ceppo di E. coli BL21(DE3)pLys. L'attività enzimatica di M<sup>pro</sup>, ottenuta per via ricombinante, ha confermato che la proteina è enzimaticamente attiva. Inoltre, l'introduzione di un inibitore specifico per M<sup>pro</sup> ha permesso di quantificare l'effettiva inibizione della proteasi comparandola con i saggi di attività effettuati in assenza di inibitori.

I dati raccolti e dimostrati in questo lavoro di tesi hanno permesso di effettuare studi biofisiche di caratterizzazione strutturale della proteasi principale di SARS-CoV-2 e di indagare l'equilibrio monomero-dimero della proteina stessa in diverse condizioni di concentrazione e temperatura. Fondamentale sono gli studi di attività enzimatica finalizzati all'identificazione di un inibitore specifico in grado di bloccare il ruolo chiave di M<sup>pro</sup> nella replicazione virale.

## 7. Bibliografia

1. Cui, J., Li, F. & Shi, Z. L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
2. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*.
3. Deng, S. Q. & Peng, H. J. Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. *J. Clin. Med.* **9**, 575 (2020).
4. Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
5. McBride, R., van Zyl, M. & Fielding, B. C. The Coronavirus nucleocapsid is a multifunctional protein. *Viruses*, 2991-3018,
6. Cubuk, J., Alston, J.J., Incicco, J.J. et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* **12**, 1936 (2021).

7. Qiu Y, Xu K. Functional studies of the coronavirus nonstructural proteins. *STE Med.* 2020.
8. Ahmad Abu Turab Naqvi, Kisa Fatima, Taj Mohammad, Urooj Fatima, Indrakant K. Singh, Archana Singh, Shaikh Muhammad Atif, Gururao Hariprasad, Gulam Mustafa Hasan, and Md. Imtaiyaz Hassan (101 october 2020) Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim Biophys Acta Mol Basis Dis.*
9. .Lan, J., Ge, J., Yu, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
10. Kiemer, L., Lund, O., Brunak, S. & Blom, N. Coronavirus 3CLpro proteinase cleavage sites: possible relevance to SARS virus pathology. *BMC Bioinformatics*, (2004).
11. Berger, A. & Schechter, I. (1970). Mapping the active site of papain with the aid of peptide substrates and inhibitors *Philos Trans R Soc Lond B Biol Sci* 257(813), 249-264
12. Anand, K. et al. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra  $\alpha$ -helical domain. *EMBO* (2002).



13. Hilgenfeld R. et al. (2006) Structure and Dynamics of Sars Coronavirus Main Proteinase (M<sup>PRO</sup>). In: Perlman S., Holmes K.V. (eds) *The Nidoviruses. Advances in Experimental Medicine and Biology*, vol 581.
14. Jin, Z. et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582(7811), 289-293 (2020).
15. Yang, H. et al. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc Natl Acad Sci U S A* (2003).
16. St John, S. E., Tomar, S., Stauffer, S. R. & Mesecar, A. D. Targeting zoonotic viruses: structure-based inhibition of the 3C-like protease from bat coronavirus HKU4—the likely reservoir host to the human coronavirus that causes Middle East respiratory syndrome (MERS). *Bioorg Med Chem* 23(2015).
17. Wang, F. et al. Crystal structure of feline infectious peritonitis virus main protease in complex with synergetic dual inhibitors. *J Virol* (2015).
18. St John, S.E., Tomar, S., Stauffer, S.R., Mesecar, A.D. (2015) Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4-The likely reservoir host to the

human coronavirus that causes Middle East Respiratory Syndrome (MERS). *Bioorg.Med.Chem.*

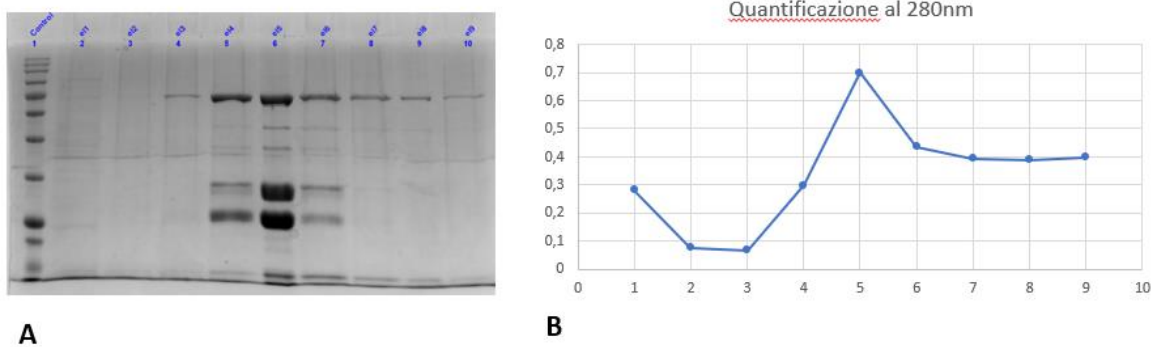
19. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry.*
20. Silvestrini L., Belhaj N., Comez L., Gerelli Y., Lauria A., Libera V. The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci Rep.* 2021;
21. Xue X, Yang H, Shen W, Zhao Q, Li J, Yang K, Chen C, Jin Y, Bartlam M & Rao Z (2007) Production of authentic SARS-CoV MPro with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. *J Mol Biol.*

## 8. Appendici

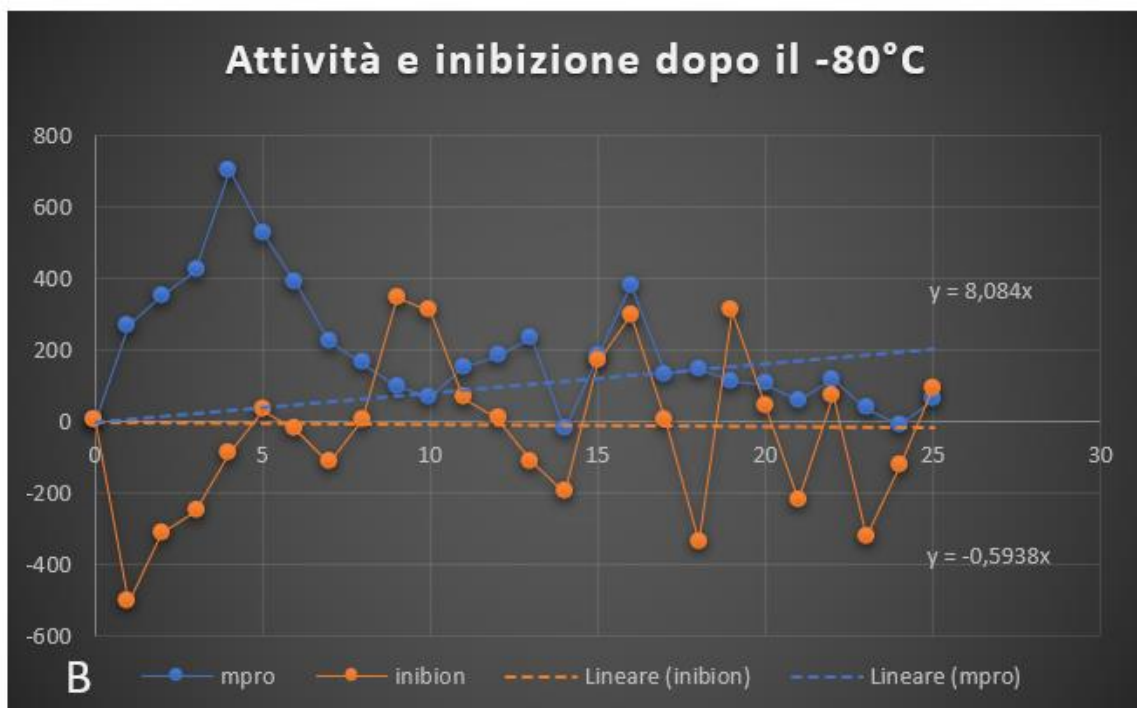
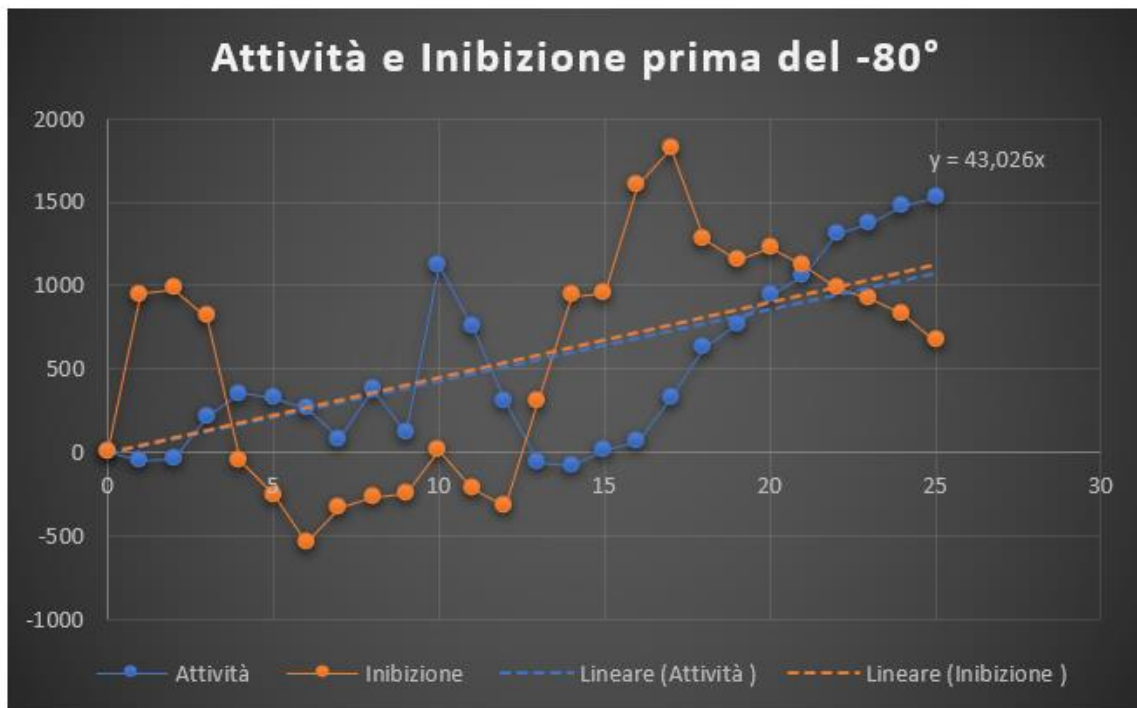
### *Appendice I*

In questo lavoro di tesi sono stati svolti ulteriori test:

1. L'attività di  $M^{pro}$ , con concentrazione proteica pari a 0.71, è stata analizzata a seguito al suo mantenimento a  $-80^{\circ}C$  con glicerolo al 20%, per valutare l'eventuale effetto legato al mantenimento a bassa temperatura. Si è potuto dimostrare che in seguito al congelamento  $M^{pro}$  presenta un'attività inferiore.

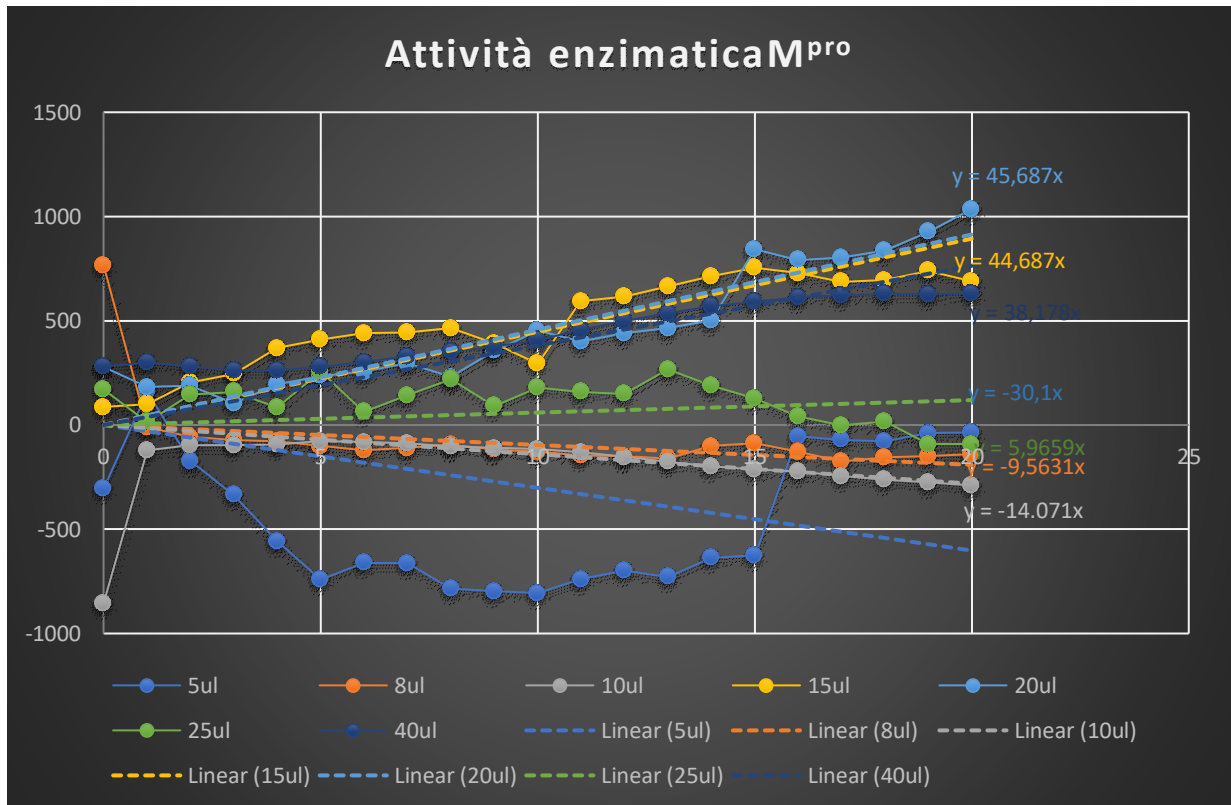


**Figura 30.** (A) SDS-PAGE, lane 1. Ladder; lane 2 E11; lane 3, E12; lane 4, E13; lane 5, E14; lane 6, E15; lane 7 E16; lane 8, E17; lane 9, E18; lane 10, E19. (B) Quantificazione al 280nm dei eluati.



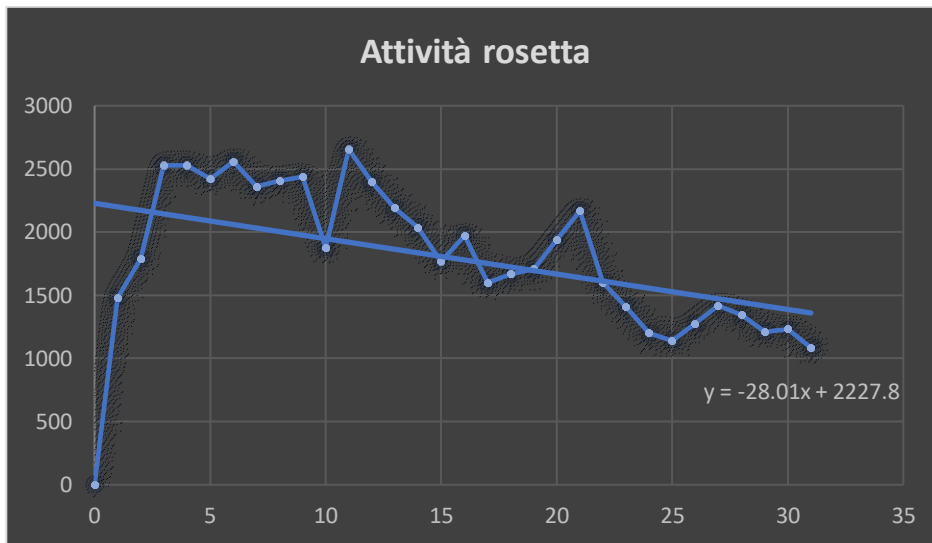
**Figura 31.** Attività enzimatica e inibizione enzimatica, (A) prima del congelamento (B) dopo il congelamento.

2. L'attività enzimatica è stata analizzata a differenti rapporti di concentrazione rispetto al substrato. Grazie all'utilizzo di una quantità costante di substrato (40 $\mu$ M), si è potuto osservare come la quantità proteica influisca sull'attività.



**Figura 32.** Attività enzimatica di M<sup>pro</sup> in BL21(DE3)pLys a quantità differenti (5ul, 8ul 10ul, 15ul, 20ul 25 ul, 40ul) monitorata per 25 min.

3. L'attività enzimatica è stata monitorata anche dopo produzione nel ceppo Rosetta™(DE3) + M<sup>pro</sup>, mostrando che è inferiore a quella di M<sup>pro</sup> ottenuta dal ceppo BL21(DE3)pLys + M<sup>pro</sup>, dato che nel primo caso non presenta la conformazione nativa e di conseguenza non sarà attiva.



**Figura 33.** Attività enzimatica di Rosetta™(DE3) + M<sup>pro</sup>: l'attività è bassa dato che l'estrazione della proteine è avvenuta dai corpi inclusi e di conseguenza la proteina non presenta il corretto ripiegamento e sarà inattiva.

## ***Appendice II***

Il lavoro di questa tesi è inserito in una linea di ricerca attivata durante la pandemia da COVID-19 e svolta in collaborazione con le Università di Perugia e di Palermo e il CNR di Perugia. Il tema di questa collaborazione è la ricerca di piccoli inibitori della Mpro che potessero bloccarne l'attività e l'analisi dei loro meccanismi di inibizione (in particolare, blocco della dimerizzazione o/e blocco del sito attivo). L'attività svolta ha portato alla pubblicazione di un lavoro nella rivista Scientific Report (Silvestrini, L., Belhaj, N., Comez, L. et al. The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Sci Rep* 11, 9283 (2021). <https://doi.org/10.1038/s41598-021-88630-9>) riportato qui in Appendice II.

# Dimer-monomer equilibrium of SARS-CoV-2 main protease as affected by small molecule inhibitors: a biophysical investigation

Lucia Silvestrini<sup>1</sup>, Norhan Belhaj<sup>1</sup>, Lucia Comez<sup>3</sup>, Yuri Gerelli<sup>2</sup>, Antonino Lauria<sup>4</sup>, Valeria Libera<sup>5</sup>, Paolo Mariani<sup>2</sup>, Paola Marzullo<sup>4</sup>, Maria Grazia Ortore<sup>2</sup>, Antonio Palumbo Piccionello<sup>4</sup>, Caterina Petrillo<sup>5</sup>, Lucrezia Savini<sup>1</sup>, Alessandro Paciaroni<sup>\*5</sup>, and Francesco Spinozzi<sup>\*2</sup>

<sup>1</sup>Marche Polytechnic University, Department of Life and Environmental Sciences, NY-MaSBiC laboratory, Ancona, 60131, Italy

<sup>2</sup>Marche Polytechnic University, Department of Life and Environmental Sciences, Ancona, 60131, Italy

<sup>3</sup>CNR-IOM c/o University of Perugia, Department of Physics and Geology, Perugia, 06123, Italy

<sup>4</sup>University of Palermo, STEBICEF Department, Palermo, 90128, Italy

<sup>5</sup>University of Perugia, Department of Physics and Geology, Perugia, 06123, Italy

\*alessandro.paciaroni@unipg.it, f.spinozzi@univpm.it

## ABSTRACT

The maturation of coronavirus SARS-CoV-2, which is the etiological agent at the origin of the COVID-19 pandemic, requires a main protease  $M^{pro}$  to cleave the virus-encoded polyproteins. Despite a wealth of experimental information already available, there is wide disagreement about the  $M^{pro}$  monomer-dimer equilibrium dissociation constant. Since the functional unit of  $M^{pro}$  is a homodimer, the detailed knowledge of monomer-dimer equilibrium thermodynamics is a key piece of information for possible therapeutic intervention, with small molecules interfering with dimerization being potential broad-spectrum antiviral drug leads. In the present study, we exploit small angle x-ray scattering (SAXS) to investigate the structural features of the SARS-CoV-2  $M^{pro}$  monomer-dimer equilibrium, by revealing the corresponding equilibrium dissociation constant and the associated thermodynamic parameters. Then, SAXS is used to study how the  $M^{pro}$  dissociation process is affected by a set of small inhibitors selected through combinatorial design. Our results show that a clear picture connecting the ability of inhibitors to disrupt the  $M^{pro}$  dimerization with the loss of catalytic activity cannot be provided, thus highlighting the possible role of allosteric effects for the regulation of  $M^{pro}$  functionality.

## Introduction

The COVID-19 pandemic is the ongoing worldwide health emergency caused by the coronavirus SARS-CoV-2 (severe acute respiratory syndrome-coronavirus-2).<sup>1,2</sup> Coronaviruses (CoVs) are enveloped positive-stranded RNA viruses; once the virion gets into the cell, the single-strand RNA translates into two overlapping polyproteins, termed pp1a and pp1ab, which mediate viral replication and proliferation. The virus maturation involves a highly complex cascade of proteolytic processing events on these polyproteins: most cleavage events are ruled by a nonstructural protein, the CoV main protease ( $M^{pro}$ , also known as 3CL<sup>pro</sup>), a three-domain (domains I to III) protein.<sup>3</sup> The enzyme shows first autolytic cleavage from pp1a and pp1ab, then starts processing the two polyproteins at no less than 11 conserved sites.<sup>3</sup>

Because of this mechanism of action, inhibiting  $M^{pro}$  might lead to an attenuation of the viral infection. Indeed, this enzyme is a very attractive target for anti-CoV drug design: the  $M^{pro}$  sequence is highly conserved among various CoVs,<sup>4</sup> as mutations of  $M^{pro}$  turn out to be often fatal for the virus.<sup>5</sup> Thus, the risk of mutation-mediated drug resistance is very low and inhibitors will display broad-spectrum antiviral activity. In addition,  $M^{pro}$  inhibitors are unlikely to be toxic because human proteases have different cleavage specificity. A second point should be however considered: the published X-ray structures of SARS-CoV-2  $M^{pro}$ , obtained both in the presence and in the absence of inhibitors,<sup>6,7</sup> revealed that two  $M^{pro}$  molecules form a functional active homodimer, as already detected in different coronaviruses,<sup>3</sup> which share with SARS-CoV  $M^{pro}$  almost all the amino-acids involved in the dimerization. In such homodimer, the two monomers are arranged almost perpendicular to each other<sup>7</sup> and each monomer comprises the catalytic dyad His41-Cys145 and the substrate-binding site located in a cleft between domains I and II. Domain III, which contains five  $\alpha$ -helices arranged into a globular cluster, is directly involved in controlling



dimerization of  $M^{\text{pro}}$  mainly through a salt-bridge between Glu290 of one monomer and Arg4 of the other.<sup>8</sup> Quite remarkably, while individual monomers are enzymatically inactive,  $M^{\text{pro}}$  is active in the dimeric form. The structural reason behind the functionality of the dimer is probably due to the interaction of the N-finger of each of the two monomers with Glu166 of the other monomer, which establishes the shape of the so-called S1 pocket of the substrate-binding site.<sup>9</sup> To approach this interaction site, the N-terminal amino acid residues are squeezed in between domains II and III of the parent monomer and domain II of the other monomer.<sup>7</sup>

According to these considerations, two different strategies have been considered for the development of therapeutic agents: first, direct inhibition of the catalytic site by using molecules targeting the substrate binding pocket; second, attenuation of the catalytic activity by using inhibitors targeting the dimerization site. The second alternative is related to the  $M^{\text{pro}}$  equilibrium between dimers and monomers in solution. The thermodynamic equilibrium of  $M^{\text{pro}}$  dissociation process, has been recently studied by analytical ultracentrifugation. Sedimentation velocity experiments provided a value of about  $2.5 \mu\text{M}$  for the apparent dimer dissociation constant  $K_D$ .<sup>7</sup> However, a more recent estimate, by mass-spectrometry based assays, established for  $K_D$  a much lower value of  $0.14 \pm 0.03 \mu\text{M}$ , indicating that  $M^{\text{pro}}$  has a stronger preference to dimerize in solution than expected. In the case of SARS-CoV  $M^{\text{pro}}$ , an even wider discrepancy among the different estimates of the dimer-monomer dissociation constants has been observed, with the values of  $K_D$  provided by various experimental techniques falling in a range from  $230 \pm 30 \mu\text{M}$ <sup>10</sup> down to  $0.19 \pm 0.03 \mu\text{M}$ .<sup>11</sup> In the light of the above mentioned uncertainty on the  $K_D$  value, we decided to further investigate the SARS-CoV-2  $M^{\text{pro}}$  dimer-monomer equilibrium. The study was also performed in the presence of a set of in-silico selected small inhibitors, whose activity was spectroscopically assayed, in order to simultaneously test their potential therapeutic strategy with respect to dimerization inhibition. By using small angle X-ray scattering (SAXS) and circular dichroism (CD) techniques we could characterize the large-scale structural features of SARS-CoV-2  $M^{\text{pro}}$  dimer-monomer equilibrium and the way they are affected by the presence of small inhibitors.

## Results

As the dimerization process is a pivotal key for the biological activation of the SARS-CoV-2  $M^{\text{pro}}$ , several therapeutic strategies against COVID-19 are based on inhibitors acting also (or only) at the  $M^{\text{pro}}$  dimerization interface.<sup>6,12,13</sup> In order to evaluate the effects of potential inhibitors targeting both the catalytic and dimer dissociation, we have first derived the thermodynamics parameters controlling the  $M^{\text{pro}}$  dimer-monomer equilibrium in solution by SAXS and CD spectroscopy techniques. Subsequently, we have investigated by SAXS experiments the  $M^{\text{pro}}$  dimerization in the presence of a series of potential inhibitors, selected from an *in-house* database containing commercial and synthetic compounds. Activity assays were also performed to correlate the  $M^{\text{pro}}$  activity to dimerization inhibition.

### $M^{\text{pro}}$ dimerization and thermal stability

The dimer-monomer equilibrium of SARS-CoV-2  $M^{\text{pro}}$  has been investigated at different protein concentrations by performing in-solution SAXS experiments in the temperature range between  $15^\circ$  and  $45^\circ$  C and far-UV CD measurements at room temperature. Far-UV CD spectroscopy was also used to study the  $M^{\text{pro}}$  thermal stability, monitoring the unfolding transition between  $10^\circ$  and  $80^\circ$  C.

### SAXS

SAXS data of SARS-CoV-2  $M^{\text{pro}}$  recorded at the B21 beam-line of the Diamond Synchrotron (Didcot, UK) at different protein concentrations and temperatures are shown as log-log plots in Fig. 1. We have assumed that SAXS curves arise from a system of interacting  $M^{\text{pro}}$  monomers and dimers, according to the thermodynamic equilibrium dissociation process given by the relationship:

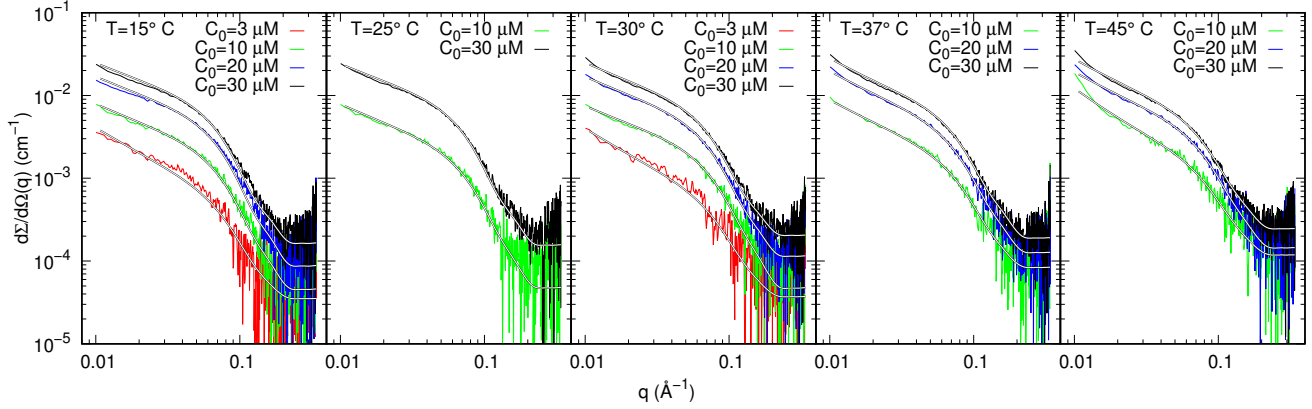


The corresponding equilibrium dissociation constant is

$$K_D = \frac{[M_1^{\text{pro}}]^2}{[M_2^{\text{pro}}]} = \frac{2Cx_1^2}{1-x_1} = e^{-\Delta G_D/(RT)}, \quad (2)$$

where  $C$  is the total molar concentration of monomers,  $x_1$  is the molar fraction of proteins that remain in the monomeric state,  $\Delta G_D$  is the dissociation Gibbs free energy change,  $R$  is the universal gas constant and  $T$  the absolute temperature. To note, Eq. 2 can be solved in terms of  $x_1$ ,

$$x_1 = K_D \frac{\sqrt{1+8C/K_D}-1}{4C}. \quad (3)$$



**Figure 1.** SAXS experimental data and best theoretical fits obtained by GENFIT software<sup>14</sup> (solid black and white lines) of SARS-CoV-2 M<sup>pro</sup>. Each panel reports a dataset obtained at the same temperature.

According to classical thermodynamics, the temperature dependence of  $\Delta G_D$  is

$$\Delta G_D = \Delta G_D^\circ + (\Delta C_{pD} - \Delta S_D^\circ)(T - T_o) - \Delta C_{pD} T \log \frac{T}{T_o} \quad (4)$$

where  $\Delta G_D^\circ = -RT_o \log K_D^\circ$  is the dissociation Gibbs free energy at the reference temperature  $T_o = 298.15$  K ( $K_D^\circ$  being the associated equilibrium constant),  $\Delta C_{pD}$  is the change of the constant pressure heat capacity upon dissociation (here supposed independent on  $T$ ) and  $\Delta S_D^\circ$  is the dissociation entropy at  $T_o$ .

The macroscopic differential scattering cross section, which is the experimental information provided by a SAXS curve, for a system of interacting monomers and dimers can be written as

$$\frac{d\Sigma}{d\Omega}(q) = N_A \kappa C_N P(q) S_M(q) + B, \quad (5)$$

$N_A$  being Avogadro's number,  $\kappa$  an unknown fraction of the nominal protein molar concentration  $C_N$  ( $C = \kappa C_N$ ),  $B$  an arbitrary flat background that takes into account possible uncertainties in the determination of transmissions of proteins and buffers samples.  $P(q)$  represents the average form factor of the system

$$P(q) = x_1 P_1(q) + \frac{1}{2} (1 - x_1) P_2(q), \quad (6)$$

where  $P_j(q)$  stands for the form factor (which is the orientational average of the excess squared X-ray scattering amplitude) of the M<sup>pro</sup> monomer ( $j = 1$ ) or dimer ( $j = 2$ ). We have calculated  $P_j(q)$  from the the crystal structure of SARS-CoV-2 M<sup>pro</sup> dimer recently determined<sup>7</sup> (PDB code 6y2e) considering one chain ( $j = 1$ ) or both chains ( $j = 2$ ) by means of the SASMOL method.<sup>15</sup> This method takes into account the contribution to the scattering due to the hydration water molecules around the protein, whose positions are found by embedding the atomic structure in a tetrahedral close packed lattice. For SARS-CoV-2 M<sup>pro</sup> monomer and dimer, 726 and 1243 hydration water molecules have been respectively calculated, suggesting that for the dimer formation about 200 water molecules are removed from the hydration shell of both monomers. Hence, the water molecules attributed to each monomer decrease from 726 to 621 upon M<sup>pro</sup> dimerization. This suggests that the dimerization process is accompanied with slight structural changes reducing the average area accessible to solvent. The  $S_M(q)$  term in Eq. 5 is the so-called ‘‘measured’’ structure factor, which describes the long range intermolecular interactions among all the particles in solution. For a sake of simplicity, here we consider a unique effective structure factor that takes into account monomer-monomer, monomer-dimer and dimer-dimer interactions. Considering that at low  $q$  all the experimental scattering curves (Fig. 1) show a positive deviation from a Guinier trend, indicative of the prevalence of protein-protein attraction with respect to repulsion, we have approximated the structure factor by the one of fractal distribution of inhomogeneities developed by Teixeira,<sup>16</sup> whose main parameters are  $D$ , the fractal dimension of the aggregates,  $r_0$ , the effective radius of the aggregating protein molecule and  $\xi$ , the correlation length, which can be interpreted as the average size of the aggregates (see Eqs. 10, 11 and 12).

The above described model, which combines SARS-CoV-2 M<sup>pro</sup> thermodynamic and structural features, has been adopted to simultaneously analyze the whole set of the SAXS curves shown in Fig. 1. Fitting parameters shared by all the curves are  $K_D^\circ$ , the dissociation equilibrium constant at  $T_o$ ,  $\Delta C_{pD}$ , the constant pressure heat capacity upon dissociation,  $\Delta S_D^\circ$  and the

dissociation entropy at  $T_0$ . Another parameter considered unique for all the curves is the relative mass density of the hydration water (in general higher than 1),  $d_h$ , which is taken into account in the SASMOL method.<sup>15</sup> The common fitted parameters are shown in Table 1, while all the other fitted parameters are reported in Table S1 of the Supporting Information.

$K_D^\circ$	( $\mu\text{M}$ )	$7 \pm 1$
$\Delta C_{pD}$	( $\text{kJ mol}^{-1} \text{K}^{-1}$ )	$1.7 \pm 0.7$
$\Delta S_D^\circ$	( $\text{J mol}^{-1} \text{K}^{-1}$ )	$50 \pm 20$
$d_h$		$1.100 \pm 0.006$

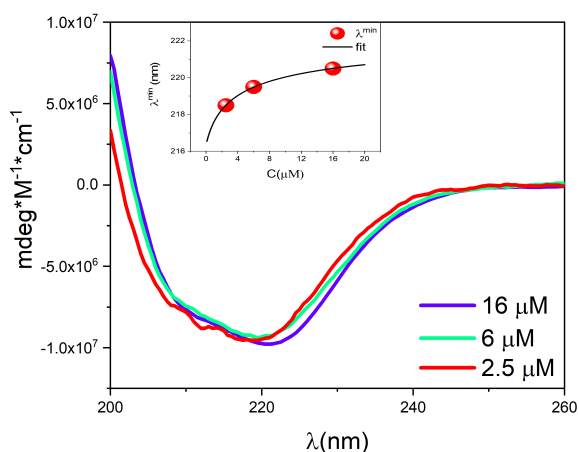
**Table 1.** Thermodynamic parameters resulting from the global fit of SAXS data for SARS-CoV-2 M<sup>pro</sup> at different temperatures and concentrations.

The most important parameter obtained by the simultaneous fit of SAXS data is the dissociation constant  $K_D^\circ$ , which resulted equal to  $7 \pm 1 \mu\text{M}$ , in good agreement with the value obtained by Graziano et al.<sup>17</sup> on the very similar main protease from SARS-CoV. The corresponding dissociation Gibbs free energy is  $\Delta G_D^\circ \simeq 30 \text{ kJ mol}^{-1}$ , a value quite similar to the one observed for the  $\beta$ -lactoglobulin dimer dissociation at neutral pH.<sup>18</sup> Regarding the dissociation entropy, we have obtained a positive value,  $50 \pm 20 \text{ J mol}^{-1} \text{K}^{-1}$ , but meaningfully smaller in respect to the one observed in the above mentioned  $\beta$ -lactoglobulin case.<sup>18</sup> It should be noticed that in a dissociation process, many factors other than translational and rotational motions contribute to a positive dissociation entropy and it is difficult to separate them. One such factor is, without doubts, the removal of about 200 hydration water molecules from the monomer-monomer interface when the dimer is formed.

The change of the heat capacity at constant pressure upon dissociation resulted positive and large. This parameter indirectly describes the monomer-monomer interface, as it can be attributed to the hydration and correlates with the interface size.<sup>19</sup> When dissociation heat capacities are positive and large, temperature meaningfully increases monomer-monomer affinities. This is our case: the M<sup>pro</sup> large dissociation heat capacity might be directly correlated with the SARS-CoV2 infective efficiency as a function of temperature. However, a further investigation on the monomer-monomer interface area and its relationship with the dissociation heat capacity<sup>20</sup> requires further calorimetric experiments in order to obtain a lower estimation errors. Finally, the relative density of the hydration shell is slightly larger than one, in agreement with previous literature results on globular proteins.<sup>21–23</sup> The determination of the thermodynamic features of the dimer-monomer equilibrium of M<sup>pro</sup> in conditions quite similar to those found *in vivo* is a fundamental step to investigate the effects of drugs aimed to inhibit dimerization and underlines the importance to further investigate M<sup>pro</sup> monomer-monomer interface by in-solution techniques.

### Far-UV CD

To provide further insights on the dimer-monomer equilibrium, we have measured the far-UV CD spectra of M<sup>pro</sup> at three different concentrations, as shown in Fig. 2. At the higher concentration of  $16 \mu\text{M}$ , the ellipticity shows a minimum wavelength



**Figure 2.** Far-UV CD spectra of SARS-CoV-2 M<sup>pro</sup> at three different concentrations. Inset: position of the minimum of the spectra as a function of the concentration (red circles). The continuous line represents an estimate of the minimum position based on Eq. 7. Results from the fit are:  $\lambda_{\text{mon}}^{\text{min}} = 216.4 \pm 0.1 \text{ nm}$  and  $\lambda_{\text{dim}}^{\text{min}} = 222.9 \pm 0.1 \text{ nm}$ .

$\lambda^{\text{min}}$  at about 221 nm and a shoulder centered at about 208 nm, which are typical of proteins with  $\alpha$ -helical and  $\beta$ -sheet

content,<sup>24,25</sup> fully consistent with the structural features<sup>24</sup> of the SARS-CoV-2 M<sup>pro</sup>,<sup>7</sup> and in agreement with CD measurements of the same enzyme<sup>26</sup> and of the very much similar SARS-CoV M<sup>pro</sup>.<sup>27</sup> As concentration decreases,  $\lambda^{\min}$  shifts towards lower values, thus reporting an increase of the  $\beta$ -sheet component at the expense of the  $\alpha$ -helical content.<sup>24</sup> Such an effect is related to the fact that M<sup>pro</sup> monomer and dimer have different secondary structure components. The  $\lambda^{\min}$  trend can be described in terms of the dimer-monomer equilibrium through the following expression,

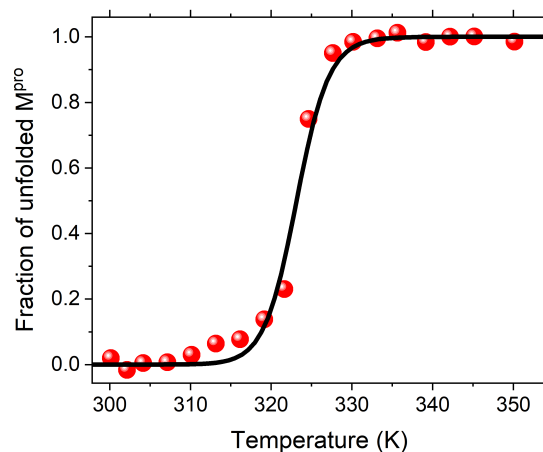
$$\lambda^{\min} = x_1 \lambda_{\text{mon}}^{\min} + (1 - x_1) \lambda_{\text{dim}}^{\min} \quad (7)$$

where we fixed  $K_D = 7 \pm 1 \mu\text{M}$  as estimated by SAXS, while  $\lambda_{\text{mon}}^{\min}$  and  $\lambda_{\text{dim}}^{\min}$  are the minimum wavelength parameters corresponding to the monomer and the dimer spectra, respectively. As shown in the inset of Fig. 2, the trend of the  $\lambda^{\min}$  values is fitted in an excellent way with Eq. 7.

The thermal stability of the M<sup>pro</sup> has been characterized by monitoring the signal at 221 nm of the M<sup>pro</sup> sample at 16  $\mu\text{M}$  concentration, which mainly consists of dimers. The rather sharp transition we have obtained is shown in Fig. 3 and clearly suggests a two-state model, where the dimer unfolds and yields two random-coil monomeric chains:



Considering the scheme 8, if we hypothesize that the dimer can unfold to two random-coil monomeric chains, we obtain an apparent melting temperature of 323 K, with a melting Van't Hoff enthalpy of  $\Delta H_v = 810 \pm 60 \text{ kJ/mol}$ . This value is in good agreement with the Van't Hoff enthalpy  $\Delta H_v \sim 880 \text{ kJ/mol}$  estimated through the equation  $\Delta H_v = 4RT_m^2 C_{P,\text{max}} / \Delta H_{\text{cal}}$  from DSC measurements.<sup>26</sup> It is also worth of note that by taking  $\Delta H_{\text{cal}} = 443 \text{ kJ/mol}$ ,<sup>26</sup> it turns out a ratio  $\Delta H_v / \Delta H_{\text{cal}} \sim 1.8$ : such a value larger than 1 is fully consistent with the unfolding transition coupled to the dimer dissociation.



**Figure 3.** Thermal melting of the SARS-CoV-2 M<sup>pro</sup> followed by monitoring the far-UV CD signal at 221 nm. The continuous line results from the theoretical fitting model arising from Eq. 8.

### M<sup>pro</sup> dimer-monomer equilibrium in presence of inhibitors

#### *In-silico inhibitor selection*

To identify new inhibitors of SARS-CoV-2 main protease from a large *in house* database, we applied the *in silico* protocol, recently proposed by some of us.<sup>28</sup> The flowchart of the adopted protocol is depicted in Fig. S1 of the Supporting Information. As a first step, we performed molecular docking studies on the compounds present in the database to analyze their binding capability in the catalytic active site of the SARS-CoV-2 M<sup>pro</sup> (PDB code 6y2f),<sup>7</sup> as detailed in the Materials and Methods section. Fig. 4 shows the 3D binding active site of SARS-CoV-2 M<sup>pro</sup> co-crystallized with the native inhibitor **13b**<sup>7</sup> covalently bonded to Cys145. The ligand binds to the enzymatic catalytic cleft of the protease located between domains I and II. The 3D binding site representation (Fig. 4) highlights the interactions with the amino acid residues involved in the inhibition mechanism, such as Met49, Met165, Glu166, His164, Phe140, Gly143 and the catalytic Cys145. It is noteworthy the presence of hydrogen bonds between the pyridone moiety of ligand and Glu166, which rules the catalytic activity driving the SARS-CoV-2 M<sup>pro</sup> to adopt an inactive conformation. The resulting best docked molecules have been selected based on a docking score cut-off of  $-6.5 \text{ kcal/mol}$  and submitted to ligand based approaches, by taking advantage of the web-service DRUDIT (DRUGs



Inhibitor	Prime_Energy	XPG_score	IFD_score
<b>1</b>	-11527.6	-8.580	-584.958
<b>2</b>	-11360.0	-10.772	-578.774
<b>3</b>	-11674.5	-7.895	-591.618
<b>4</b>	-11497.2	-5.969	-580.827
<b>5</b>	-11517.1	-8.918	-584.772
<b>6</b>	-11489.6	-9.176	-583.655
<b>7</b>	-11561.2	-10.409	-588.468
<b>13b</b>	-11736.3	-7.944	-594.758

**Table 2.** IFD results for the seven selected inhibitors compared with the **13b** compound.

amino-acids. This evidence suggests that these compounds are not covalently bound to the SARS-CoV-2 M<sup>pro</sup> catalytic site.

### M<sup>pro</sup> activity assays

The selected inhibitors have been tested for their efficacy to reduce the M<sup>pro</sup> activity. As reported in Fig. 7, the time dependence of substrate fluorescence after hydrolysis indicates that the catalytic activity of M<sup>pro</sup> in the presence of the selected compounds changes. In particular, compounds **2**, **4**, **5** and **7** induced an irreversible inactivation of the enzyme, while compounds **1** and **6** resulted rather inactive. For two of the most effective compounds (**2** and **7**) inhibition tests have been carried out as a function of the concentration. Unfortunately, we have not been able to perform this test for compound **4**, which shows the best inhibition efficacy, as it produces a fluorescence signal that partially obscures that of the substrate. Results are shown in Fig. 8. Percent inhibition data have been fitted with the Hill equation,  $p(C_1) = 100/(1 + (IC_{50}/C_1)^n)$ , to get the half maximal effective concentration, IC<sub>50</sub>, and the Hill slope  $n$ . We obtained IC<sub>50</sub> = (10.3 ± 0.2) μM for **2** and (15 ± 2) μM for **7**, with  $n = 5 ± 1$  and  $3 ± 1$ , respectively. These values on  $n$  larger than one indicate that the binding is positively cooperative, in agreement with other recent experimental results.<sup>32</sup>

### SAXS

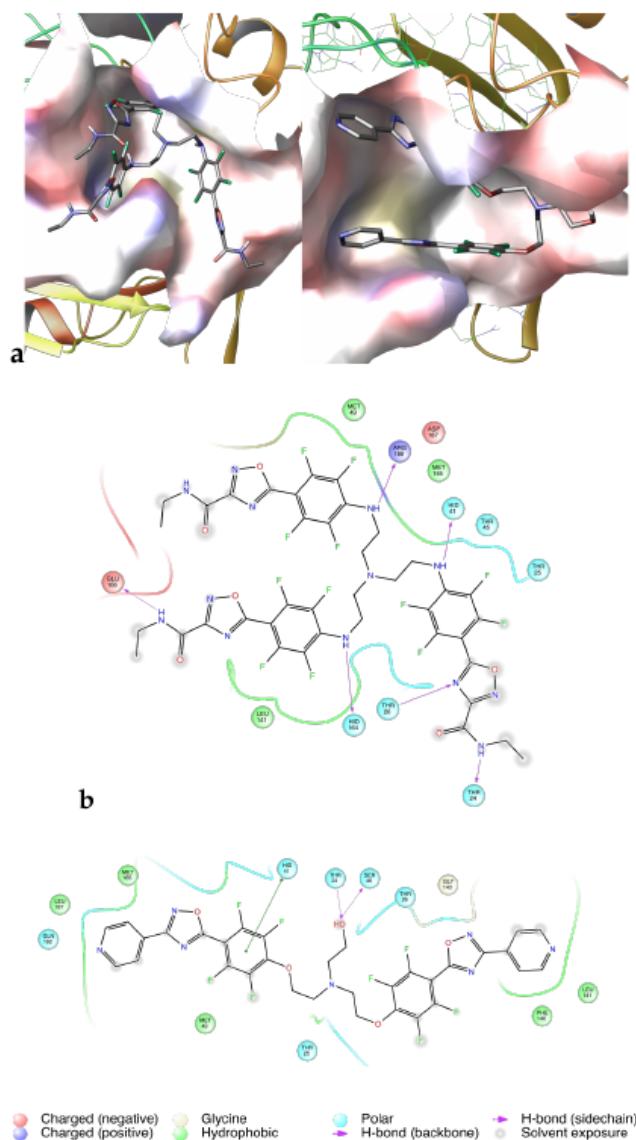
SAXS curves of SARS-CoV-2 M<sup>pro</sup> samples obtained in the presence of the seven selected potential inhibitors at two concentrations and at different temperatures, are reported in Fig. 9 as log-log plots. SAXS data have been analysed with the same approach adopted for data without inhibitors, with the further assumption that, for each compound, the thermodynamic parameters are linear functions of its concentration  $C_1$ , namely  $\Delta G_D^\circ = \Delta G_{D,0}^\circ(1 + \alpha_G C_1)$ ,  $\Delta C_{pD} = \Delta C_{pD,0}(1 + \alpha_{C_p} C_1)$ , and  $\Delta S_D^\circ = \Delta S_{D,0}^\circ(1 + \alpha_S C_1)$ .  $\Delta G_{D,0}^\circ = -RT_0 \log K_{D,0}^\circ$ ,  $\Delta C_{pD,0}$  and  $\Delta S_{D,0}^\circ$  are exactly the values already obtained from the analysis of SAXS data without inhibitors (reported in Table 1), and the three corresponding constant rates  $\alpha_G$ ,  $\alpha_{C_p}$  and  $\alpha_S$  are fitting parameters common to all the SAXS curves corresponding to the same inhibitor. The high quality of the fitting procedure can be appreciated in Fig. 9, where the calculated SAXS curves are superposed to the experimental curves and the resulting thermodynamic common fitting parameters are shown in Table 3. The inhibitors with the lowest values of  $\alpha_G$  (Table 3) are

Inhibitor	$\alpha_G$ (10 <sup>-2</sup> μM <sup>-1</sup> )	$\alpha_{C_p}$ (10 <sup>-2</sup> μM <sup>-1</sup> )	$\alpha_S$ (10 <sup>-2</sup> μM <sup>-1</sup> )
<b>1</b>	-0.18±0.02	-6±3	-4±1
<b>2</b>	-0.0±0.1	-1±3	-4±1
<b>3</b>	0.00±0.08	0±4	-8±4
<b>4</b>	0.0±0.1	-2±2	-1±5
<b>5</b>	-0.14±0.05	-1±4	-4±4
<b>6</b>	-0.20±0.05	-6±3	-1.7±0.8
<b>7</b>	-0.19±0.07	-3±3	-5±2

**Table 3.** Common thermodynamic fitting parameters of the analysis of SAXS data for SARS-CoV-2 M<sup>pro</sup> samples with inhibitors.

those that mostly favour dimer dissociation. Results reported in Table 4 suggest that compounds **1**, **6**, and **7** are, within the experimental error, mostly able to increase the dissociation equilibrium constant, which at  $C_1 = 30$  μM becomes as large as  $\approx 15$  μM and, at  $C_1 = 60$  μM almost doubles its value, reaching  $\approx 30$  μM. Inhibitor **5** is slightly less active: at  $C_1 = 60$  μM we found a dissociation equilibrium constant of  $\approx 20$  μM. The other three compounds, **2**, **3** and **4**, do not show any statistically significant difference with respect to the results in the absence of inhibitors. Despite the high uncertainties on  $\alpha_{C_p}$  and  $\alpha_S$ , their negative values suggest that upon dissociation there are changes of heat capacity and of entropy smaller than those observed



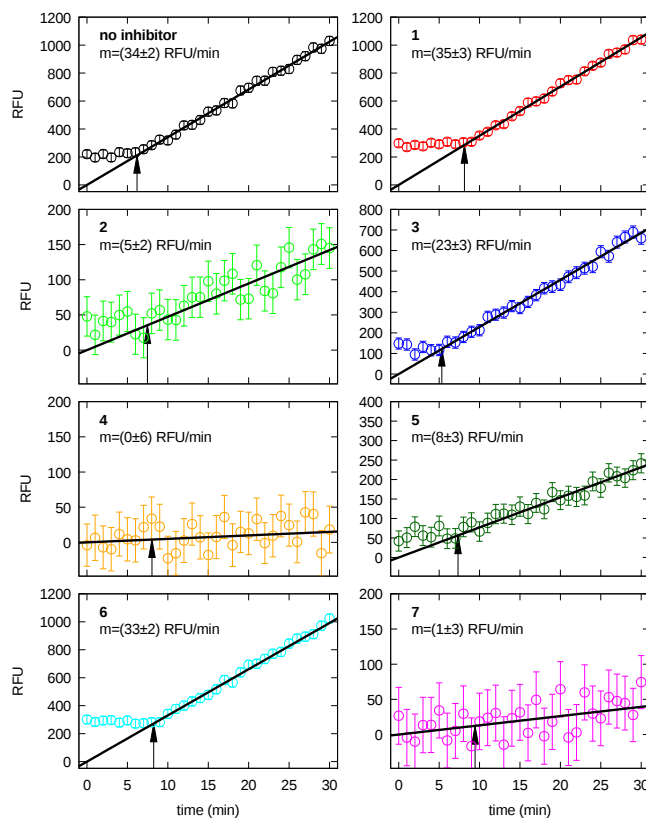


**Figure 6.** 3D binding modes of best scored compounds **3** and **7** into Covid-19 M<sup>Pro</sup> active site (**a**) and amino acid maps (**b**).

without inhibitors, indicating that inhibitors increase the monomer order. Looking to the single curve parameters, reported in Table S1 and S2 of the Supporting Information, we observe that for almost all cases the values of the correlation length  $\xi$  are similar for samples with and without inhibitors. The fractal dimension is  $\approx 2$ , suggesting a two-dimensional fractal growth of the aggregates in presence of inhibitors.

## Discussion

It is widely known that the active site of M<sup>Pro</sup> monomer, which is widely conserved among all coronavirus, is typically composed of four subsites, referred to as S1', S1, S2, and S4.<sup>33–35</sup> They accommodate the corresponding domains P1', P1, P2, and P4 of the substrate or the ones of the inhibiting compound that mimics the substrate.<sup>36</sup> The S1' subsite is constituted by the two residues Thr24 and Thr25. The S1 subsite (also referred to as the S1 pocket<sup>35</sup>) is formed by the side chains of residues Phe140, Asn142, Glu166, His163 and His172 and by the main chains of Phe140 and Leu141.<sup>36</sup> As discussed by Sacco et al.,<sup>35</sup> there are evidences that S1 can interact with both hydrophobic and hydrophilic groups: hence it is considered a promising target for an inhibiting compound. S2 is a hydrophobic subsite formed by the side chains of His41, Met49, Tyr54, Met165 and Asp187. S4 is a small hydrophobic pocket that involves the side chains of Met165, Leu167, Phe185, Gln192 and Gln189.<sup>36</sup> An unusual catalytic dyad, His41-Cys145, acts in the active site, where His41 is a proton acceptor whereas Cys145 is attacked



**Figure 7.** Fluorescence inhibition curves of the selected compounds, as indicated in each frame. The straight lines are the best fitting lines obtained considering data points comprised between the time indicated by the arrow and 30 min. The slope of the straight line is reported in each frame.

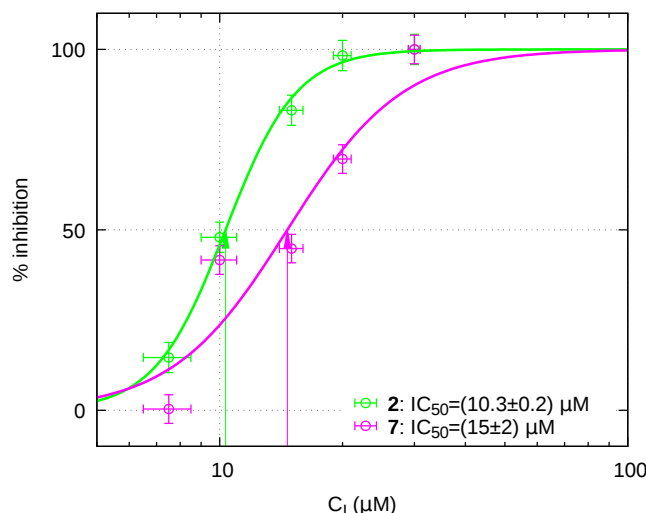
by the carbonyl carbon of the substrate. Hence, a signature of the inhibiting power of a compound is its capability to form a covalent bond with Cys145,<sup>33</sup> as very recently confirmed by Dai et al.,<sup>34</sup> who have found two promising inhibitors **11a** and **11b**. The importance of the protonation state of Cys145 as well as the network of hydrogen bonds between the catalytic site of  $M^{P70}$  and inhibiting compounds has also been recently discussed by Kneller et al.<sup>37</sup> by combining X-ray and neutron results.

On these grounds, we now discuss the experimental results obtained by the present study together with the structure of the seven inhibitors within the  $M^{P70}$  active site determined by the refined molecular docking.

The interaction map of inhibitor **1** is shown in Fig. S4 of the Supporting Information. There are a total of 11 contacts with amino acids of  $M^{P70}$  monomer (Ser1, Thr25, Thr26, Ser46, Asn119, Leu141, Asn142, Cys145, Pro168, Arg188, Gln192), 2 of them (Ser46, Asn119) are hydrogen bonds. The residues of the catalytic dyad and the four subsites in contact with **1** are: Cys145 (dyad, 1 of 2 (50%)); Thr25 (S1', 1 of 2 (50%)); Leu141 and Asn142 (S1, 2 of 6 (33%)); Gln192 (S4, 1 of 5 (20%)). To note, these contacts involve only one of the residues of the catalytic dyad, Cys145, without a hydrogen bond, whereas for inhibitor **13b** there is a hydrogen bond with Cys145 (Fig. 4). It also worth to notice that Ser1 is among the residue in contact with inhibitor **1**: since the mutual interaction of Glu166 of one monomer and Ser1 (the N-finger) of the other monomer has been proven to shape the catalytic cleft,<sup>7</sup> we argue that this compound could destabilize the dimer, as suggested by the high value of  $K_D^o = 26 \pm 4 \mu\text{M}$ . However, its enzymatic inhibition is very poor, as shown by the high similarity of the RFU slope with the one in the absence of inhibitors (Fig. 7). A possible explanation of this result could be the absence of any hydrogen bond with Cys145 as well as the absence of any contact with the residues of subsite S2.

Regarding inhibitor **2**, the map of contacts shown in Fig. S4 of the Supporting Information reveals a total of 10 interactions with the monomer chain (Met49, Asn142, Gly143, Cys145, Asp187, His164, Met165, Glu166, Arg188, Gln189); 4 of them are hydrogen bonds (Asn142, His164, Glu166, Gln189) that do not involve the catalytic site. More in detail, the residues of the catalytic dyad and the four subsites in contact with **2** are: Cys145 (dyad, 1 of 2 (50%)); Asn142 and Glu166 (S1, 2 of 6 (33%)); Met49, Met165 and Asp187 (S2, 3 of 5 (60%)); Met165 and Gln189 (S4, 2 of 5 (40%)). We also note that 2 residues of





**Figure 8.** Percent inhibition data of SARS-CoV-2 M<sup>pro</sup> as a function of the concentration of inhibitor **2** (green points) and **7** (magenta points). Best fits with the Hill equation are shown as solid lines.

30 $\mu$ M Inhibitor	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>6</b>	<b>7</b>
$K_D^\circ$ ( $\mu$ M)	14 $\pm$ 1	8 $\pm$ 3	7 $\pm$ 2	6 $\pm$ 2	15 $\pm$ 3	14 $\pm$ 3

60 $\mu$ M Inhibitor	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
$K_D^\circ$ ( $\mu$ M)	26 $\pm$ 4	8 $\pm$ 6	7 $\pm$ 4	5 $\pm$ 4	19 $\pm$ 7	30 $\pm$ 10	30 $\pm$ 10

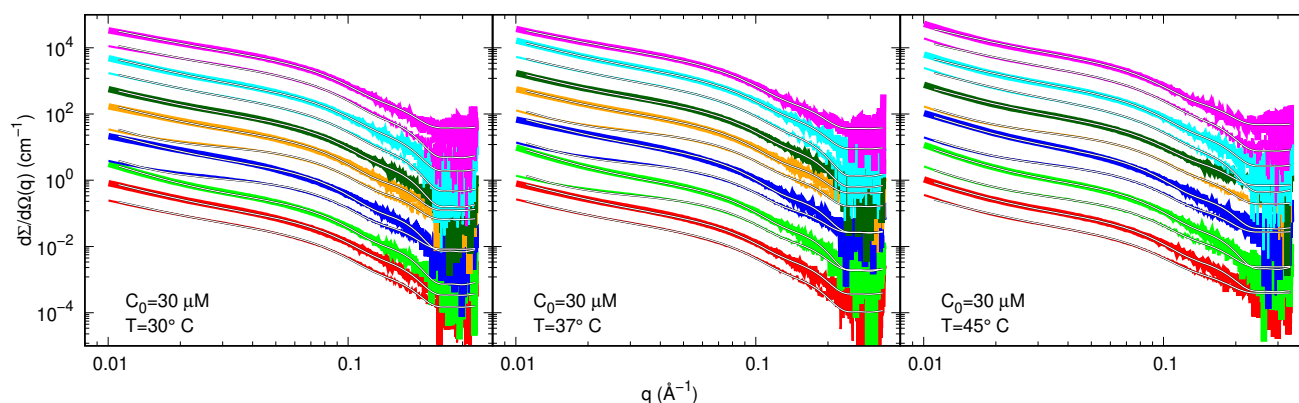
**Table 4.** Dissociation constants derived by the analysis of SAXS data for SARS-CoV-2 M<sup>pro</sup> samples with inhibitors.

S1, Asn142 and Glu166, interact with this inhibitor via a hydrogen bond. This evidence, together with the high number of contacts with S2 and S4, could explain the experimentally observed inhibition effect ( $m = 5 \pm 2$  RFU/min, Fig. 7). To note, this compound does not modify the dimer-monomer equilibrium (Table 4) and shows the worst IFD\_score parameter (Table 2).

Turning now to inhibitor **3**, despite it shows one of the best performance in terms of IFD\_score (Table 2), we have found that it does not alter the native dissociation equilibrium of M<sup>pro</sup> (Table 4) and also his inhibition effect observed by fluorescence analysis is weak ( $m = 23 \pm 3$  RFU/min, slightly lower than the value in absence of inhibitors). The contact map of compound **3** (Fig. S5 of the Supporting Information) show a total of 12 interactions with the monomer chain (Thr24, Thr25, Thr26, His41, Ser46, Met49, Phe140, Leu141, Gly143, Met165, Leu167, Gln192), 2 of them being hydrogen bonds (Thr24, Ser46) and one (His41) a  $\pi - \pi$  stacking CHECK. The residues of the catalytic dyad and the four subsites in contact with **3** are: His41 (dyad, 1 of 2 (50%)); Thr24 and Thr25 (S1', 2 of 2 (100%)); Phe140 and Leu141 (S1, 2 of 6 (33%)); His41, Met49 and Met165 (S2, 3 of 5 (60%)); Met165, Leu167 and Gln192 (S4, 3 of 5 (60%)). We notice that with respect to compound **2**, there are not hydrogen bonds with the five residues that stabilize the S1 pocket. This difference might be the reason for the weak inhibition effect.

Compound **4** is the most effective among the seven inhibitors ( $m \approx 0$ , Fig. 7), with a  $K_D^\circ$  slightly lower than the one without inhibitors (Table 4). The contact map (Fig. S5 of the Supporting Information) shows 11 interactions with the monomer chain (Thr25, Thr26, Leu27, His41, Ser46, Met165, Glu166, Pro168, Gln189, Thr190, Gln192), including one hydrogen bond (Gln189). The residues of the catalytic dyad and the four subsites in contact with **4** are: His41 (dyad, 1 of 2 (50%)); Thr25 (S1', 1 of 2 (50%)); Glu166 (S1, 1 of 6 (17%)); His41 and Met165 (S2, 2 of 5 (40%)); Met165, Gln192 and Gln189 (S4, 3 of 5 (60%)). Only one of the five residues that stabilize the S1 pocket is among the ones in contact with this inhibitor, Glu166, which does not form a hydrogen bond. On this ground, the high inhibition effect of compound **4** could be only justified by the contact with His41, one of the two residues of the catalytic dyad.

Results are different for compound **5**: it provokes a rather important increase of  $K_D^\circ$  (Table 4) and shows a moderate inhibition effect ( $m = 8 \pm 3$  RFU/min). Looking at the interaction map (Fig. S6 of the Supporting Information), we notice 11 interactions with monomer M<sup>pro</sup> (Thr25, Thr26, Leu27, His41, Asn142, Gly143, Met165, Glu166, Pro168, Arg188, Gln192), one of them regards Glu166, involved in two hydrogen bonds, and the other one His41, involved in two a  $\pi - \pi$  stacking interactions. The residues of the catalytic dyad and the four subsites in contact with **5** are: His41 (dyad, 1 of 2 (50%)); Thr25



**Figure 9.** SAXS data of SARS-CoV-2 M<sup>Pro</sup> with inhibitors at different concentrations and temperatures. Each panel reports curves at the same temperature. Red, green, blue, orange, dark-green, cyan and magenta refers to inhibitor **1**, **2**, **3**, **4**, **5**, **6** and **7**, respectively. Thin and thick lines refer to inhibitor concentrations of 30 and 60  $\mu\text{M}$ , respectively. Subsequent curves are multiplied by a factor 3.0 for clarity. Solid black and white lines are the best fits obtained by GENFIT.

(S1', 1 of 2 (50%)); Asn142 and Glu166 (S1, 2 of 6 (33%)); His41 and Met165 (S2, 2 of 5 (40%)); Met165 and Gln192 (S4, 2 of 5 (40%)). There are not hydrogen bonds involving the five residues that stabilize the S1 site. One could speculate that this inhibitor, probably due to its steric hindrance, provokes a modification of S1 that could interfere with the enzymatic activity of M<sup>Pro</sup>.

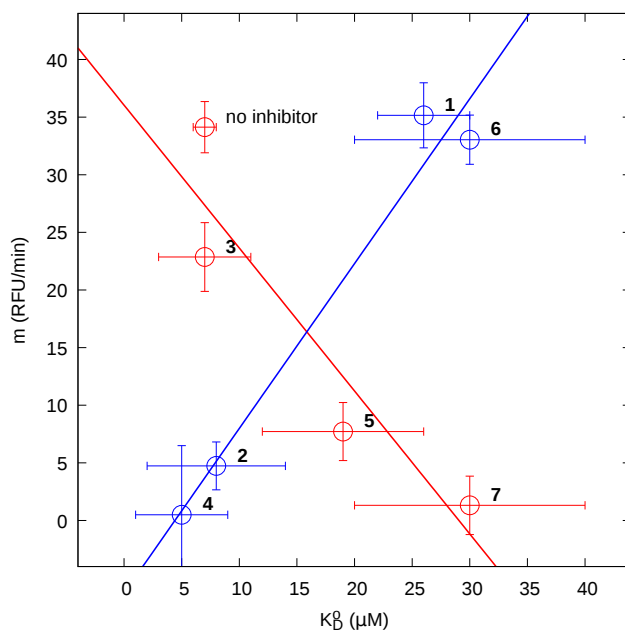
Compound **6** determines 11 contacts with the amino acid of the monomer (His41, Leu141, Ser144, Cys145, Met165, Glu166, Leu167, Arg188, Gln189, Ala191, Gln192, Fig. S6 of the Supporting Information), including two hydrogen bond (His41, Gln192). In particular, the residues of the catalytic dyad and the four subsites in contact with **6** are: His41 and Cys145 (dyad, 2 of 2 (100%)); Leu141 and Glu166 (S1, 2 of 6 (33%)); His41 and Met165 (S2, 2 of 5 (40%)); Met165, Leu167, Gln192 and Gln189 (S4, 4 of 5 (80%)). An almost absent inhibition effect is seen by fluorescence, being the slope of RFU (Fig. 7) very similar to the one determined in the absence of inhibitors. On the other side, compound **6** is able to modify the dimer-monomer dissociation, with one of the highest value of  $K_D^0 = 30 \pm 10 \mu\text{M}$  at  $C_1 = 60 \mu\text{M}$  (Table 4). To note, only one of the 6 amino acids that stabilize the S1 site are included in the list of residues interacting with compound **6**. Hence, the absence of its inhibition activity could be explained by the small size of its molecular structure, which might not be able to provoke important modifications of the S1 pocket and hence to modify the catalytic features of M<sup>Pro</sup>.

We finally turn to compound **7**, which reports the best value of IFD\_score (Table 2). It shows an opposite behaviour with respect to compound **6**: it is capable to change the dimer-monomer equilibrium at the same extent ( $K_D^0 = 30 \pm 10 \mu\text{M}$  at  $C_1 = 60 \mu\text{M}$ , Table 4) and displays a promising inhibition effect, with  $m \approx 1$ . For this compound, the map of contacts shows 12 interactions (Thr24, Thr25, Thr26, His41, Thr45, Met49, Leu141, His164, Met165, Glu166, Asp187, Arg188, Fig. S6 of the Supporting Information) with a large number of hydrogen bonds (Thr24, Thr26, His41, His164, Glu166, Arg188). The residues of the catalytic dyad and the four subsites in contact with **7** are: His41 (dyad, 1 of 2 (50%)); Thr24 and Thr25 (S1', 2 of 2 (100%)); Leu141 and Glu166 (S1, 2 of 6 (33%)); His41, Met49, Met165 and Asp187 (S2, 4 of 5 (80%)); Met165 (S4, 1 of 5 (20%)). Only one of the interacting residues (Glu166) is involved in the stabilization of the S1 pocket. We can argue that the high inhibition effect could be due to the high number of contact with S2 and to the presence of the 6 hydrogen bonds. An other hypothesis, which needs further insights to be confirmed, is that the fluorinated groups, which are present in a high number in compound **7**, may originate a new reactive warhead able to form a covalent bond with Cys145. We may also consider that one of them involves a residue of the catalytic dyad, His41, suggesting a possible important modification of the enzymatic activity.

In summary, our results show that the inhibition effect of compounds designed to bind the catalytic site of SARS-CoV-2 M<sup>Pro</sup> does not necessarily modify the dimer-monomer equilibrium. On the other side, we have also observed that compounds able to provoke dissociation do not always show inhibition effects. To better visualize the scenario presented by our results, we report in Fig. 10 the slope  $m$  of the fluorescence inhibition curve as a function of the dimer-monomer equilibrium constant  $K_D^0$  obtained at  $C_1 = 60 \mu\text{M}$  for the seven compounds. The points in this map could be organized in two groups, as represented in blue and in red. The red points refer to compounds that show an expected behaviour: the strongest is their capability to induce the dissociation of the M<sup>Pro</sup> dimer, the most important is their inhibition effect. For the blue compounds, the behavior is opposite: the increasing of the dissociation does not determine an increase of the inhibition effect. This apparently contradictory result can be in part explained by considering that, in all cases, the dissociation equilibrium is weak. As clearly shown in the Supporting Information Table S2 the fraction of monomers  $x_1$  is never larger than  $\approx 0.6$ , even in the case of the most effective

among our compounds. This means that, in the presence of a compound that alter the dimer-monomer equilibrium but that does not hamper the interaction with the substrate, there are always dimeric  $M^{pro}$  molecules that can exert their enzymatic activity when a substrate is available.

Very recently Suárez et al.,<sup>38</sup> through a 2  $\mu$ s Molecular Dynamics simulation of  $M^{pro}$  with and without a model peptide mimicking the enzyme substrate, have shown the importance of the dimerization in stabilizing the catalytic dyad and the overall contribution of protein flexibility in the binding of the protein with the substrate. The experimental work presented here provides a further evidence on the complex interplay between enzymatic activity inhibition and dimer dissociation. Furthermore, at the best of our knowledge, it shows for the first time the contribution of the SAXS technique, combined with advanced data analysis, for obtaining structural information about the SARS-CoV-2  $M^{pro}$  in solution and in the presence of promising inhibitors. Our results suggest that more experimental evidences about the impairment of monomer and dimer  $M^{pro}$  in the presence of inhibitors corroborated by computational result will be necessary for a deeper understanding of the  $M^{pro}$  allosteric mechanism.



**Figure 10.** Correlation map between the catalytic activity, represented by the RFU slope  $m$ , and dimer dissociation capability, measured by the dissociation constant  $K_D^O$ , of the seven SARS-CoV-2  $M^{pro}$  inhibitors.

## Materials and Methods

### $M^{pro}$ expression and western blot analysis

pGEX-6P-1 vector harboring the full length cDNA sequence encoding for SARS-CoV-2 Main Protease ( $M^{pro}$  NC\_045512) was purchased from GenScript (clone ID\_M16788F). The expressing vector was transformed into BL21DE3pLys *Escherichia coli* cells and the obtained clones were assayed both in small scale (5 mL) and medium scale (500 mL and 1 L) for the production of SARS-CoV-2  $M^{pro}$ . Transformants were grown onto LB medium containing 100  $\mu$ g/mL Ampicillin and 34  $\mu$ g/mL Chloramphenicol as selective antibiotics. Cultures were grown up to OD600 of 0.6-0.8 at 37° C, 200 rpm and then  $M^{pro}$  expression was induced by addition of 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG). Growth under induction was achieved both for 3 h at 37° C and 10 h at 16° C in order to test the best expressing condition. Cells were harvested by centrifugation at 6000 g. Cell pellets were resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 2 mM  $\beta$ -mercaptoethanol), and cell rupture was achieved by sonication (Sonics Vibra Cell sonicator) at 4° C. Cell debris was separated from the total protein extract by centrifugation at 6500 g for 1 h. Supernatant aliquotes were resuspended in Laemmli sample buffer, run onto 12% SDS-PAGE and transferred onto PVDF membrane for Western blot analysis.  $M^{pro}$  was decorated by 6 $\times$ -His tag monoclonal primary antibody (Invitrogen) and anti-mouse secondary antibody and detected by chemiluminescence (Clarity<sup>TM</sup> Western ICL Substrate, Biorad, Fig. S2, panels A and B).

### M<sup>Pro</sup> purification and His-tag cleavage

Total cell extract was loaded onto Ni-NTA affinity column (G-Biosciences) and washed by washing buffer (Tris-HCl 20 mM pH 7.6, NaCl 100 mM). M<sup>Pro</sup> was eluted by elution buffer (Tris-HCl 20 mM pH 7.6, NaCl 100 mM, 300 mM imidazole) in 5 fractions of 1 mL each. Aliquots of elution fractions were loaded onto 12% SDS-PAGE acrylamide gel and imidazole was removed by dialysis against Prescission cleavage buffer (Tris-HCl 50 mM pH 8.0, NaCl 150 mM, DTT 1 mM, EDTA 1 mM) through Amicon Ultra-4 centrifugal filters 30K (Merck Millipore). For M<sup>Pro</sup> C-terminal His-tag removal, the Prescission (1 U for 100  $\mu$ g of protein) cleavage reaction was performed at 4° C for 4 h and Prescission protease was then removed by GSTrap FF column (GE-Healthcare). The M<sup>Pro</sup> solution was further purified by FPLC size-exclusion chromatography on Superdex<sup>TM</sup>75 10/300 GL column (Fig. S3, panels A and B).<sup>26,36</sup>

### M<sup>Pro</sup> activity assay and inhibition

The fluorescently labelled auto-cleavage sequence of SARS-CoV-2 M<sup>Pro</sup>, Mca-AVLQ↓SGFRK(Dnp)K (purchased from GenScript), was utilized to monitor the recombinant M<sup>Pro</sup> kinetics (excitation 320 nm, emission 405 nm). The assay was started by mixing  $\approx 0.2 \mu$ M SARS-CoV-2 M<sup>Pro</sup> to different amounts of substrate (10, 20, 40  $\mu$ M) in order to set the best protein-substrate concentration to detect M<sup>Pro</sup> activity.<sup>33</sup> Fluorescence intensity was measured by DeNovix DS-11 FX+ fluorometer. The M<sup>Pro</sup> activity reported as reference for inhibition tests was obtained by linear fitting of the fluorescence curve in the presence of 40  $\mu$ M of substrate concentration.<sup>26,33</sup> Seven inhibitors dissolved in DMSO were tested at a final concentration of 30  $\mu$ M<sup>34</sup> (Fig. 7). Each reaction in a final volume of 200  $\mu$ L was firstly incubated for 20 min at 30° C without substrate. After substrate addition, fluorescence intensities were reported as relative fluorescence units (RFU) and monitored every minute for a duration of 30 min at 30° C.

### Circular Dichroism

Circular dichroism measurements were performed using Jasco J810 spectropolarimeter. Quartz cuvettes with path-length of 1 mm was used, in order to obtain the optimum signal-to-noise ratio for the M<sup>Pro</sup> samples with concentrations of 16, 6 and 2.5  $\mu$ M respectively. Each spectrum was collected in the range from 220 to 325 nm, with a scan speed of 50 nm/min. The thermal stability has been studied at 16  $\mu$ M M<sup>Pro</sup> concentration, by varying the temperature through a thermal bath from 300 to 350 K.

### Small Angle X-ray Scattering

SAXS experiments were carried out at the B21 beam-line of the Diamond Synchrotron (Didcot, UK), operating with a fixed camera length (4.014 m) at 12.4 keV ( $\lambda = 1.000 \text{ \AA}$ ) and with a flux of  $\sim 10^{12}$  photons per second. Samples were injected in the capillary (thickness 1.7 mm) by means of a robotic apparatus and measured 21 times with an exposure time of 1 min. The M<sup>Pro</sup> samples without inhibitors were measured at the nominal monomer molar concentration of 3, 10, 20 and 30  $\mu$ M and at temperature of 15°, 25°, 30°, 37° and 45° C. In the presence of inhibitors, SAXS curves were recorded at two M<sup>Pro</sup> monomer molar concentration, 30 and 60  $\mu$ M, and at three temperatures, 30°, 37° and 45° C.

SAXS data analysis approach has been described in the main text, with the exception of some minor points. Since in all conditions the nominal molar protein concentration is lower than 1 mM, its temperature variations can be considered to be only determined by the dependency with  $T$  of the relative mass density of water, which, according to literature results<sup>39</sup> is written as

$$d_w = e^{-\alpha_w(T-T_0)-\beta_w(T-T_0)^2/2}, \quad (9)$$

where, in our investigated range 15 – 45° C, the optimum value of the thermal expansivity at  $T_0$  is  $\alpha_w = 2.5 \cdot 10^{-4} \text{ K}^{-1}$  and the one of its first derivative is  $\beta_w = 9.8 \cdot 10^{-6} \text{ K}^{-2}$ . Accordingly,  $C_N = C_0 d_w$ ,  $C_0$  being the nominal protein concentration at  $T_0$ .

The measured structure factor  $S_M(q)$  has been obtained in relation to the protein-protein structure factor  $S(q)$  by:

$$S_M(q) = 1 + \beta(q)[S(q) - 1] \quad (10)$$

where  $\beta(q)$  is the coupling function

$$\beta(q) = \frac{|P^{(1)}(q)|^2}{P(q)} \quad (11)$$

and  $P^{(1)}(q)$  is the average of the protein excess scattering amplitude, a function provided, together with  $P(q)$  by the SASMOL method. According to Ref.,<sup>16</sup>  $S(q)$  has been written as

$$S(q) = 1 + \frac{1}{(qr_0)^D} \frac{D\Gamma(D-1)}{[1+(q\xi)^{-2}]^{D(D-1)/2}} \sin[(D-1)\tan^{-1}(q\xi)], \quad (12)$$

where  $\Gamma(x)$  is the gamma function,  $D$  is the fractal dimension (comprised between 1 and 3) of the aggregates,  $r_0$  is the effective radius of the aggregating protein and  $\xi$  is the correlation length.

## In-silico design

### Ligand Preparation

The default setting of the LigPrep tool implemented in Schrödinger's software (version 2017-1) was used to prepare the ligands for docking.<sup>40</sup> All possible tautomers and combination of stereoisomers were generated for pH  $7.0 \pm 0.4$ , using the Epik ionization method.<sup>41</sup> Energy minimization was subsequently performed using the integrated OPLS 2005 force field.<sup>42</sup>

### Protein Preparation

The crystal structure of SARS-CoV-2 M<sup>PRO</sup> in complex with ligand **13b** (PDB code 6y2f)<sup>7</sup> was downloaded from the Protein Data Bank.<sup>43</sup> The cocrystal ligand, covalently bonded to Cys145, was treated by breaking the covalent bond and filling in open valence. Protein Preparation Wizard of Schrödinger software was subsequently employed for further preparations of the protein structure using the default settings.<sup>44</sup> Bond orders were assigned, and hydrogen atoms as well as protonation of the heteroatom states were added using the Epik-tool (with the pH set at biologically relevant values, i.e. at  $7.0 \pm 0.4$ ). The H-bond network was then optimized. The structure was subjected to a restrained energy minimization step (RMSD of the atom displacement for terminating the minimization was 0.3 Å), using the Optimized Potentials for Liquid Simulations (OPLS) 2005 force field.<sup>42</sup>

### Docking Validation

Molecular Docking was performed by the Glide program.<sup>30,45,46</sup> The receptor grid preparation was performed by assigning the original ligand (**13b**) as the centroid of the grid box. The generated 3D conformers were docked into the receptor model using the Standard Precision (XP) mode as the scoring function. A total of 5 poses per ligand conformer were included in the post-docking minimization step, and a maximum of 2 docking poses were generated for each ligand conformer. The proposed docking procedure was validated by the re-dock of the crystallized **13b** within the receptor-binding pockets of 6y2f by Glide covalent docking. The results obtained were in good agreement of the experimental poses, showing a RMSD of 0.75 Å.

### Biotarget Finder module (DRUDIT)

The refined selection of suitable SARS-CoV-2 M<sup>PRO</sup> inhibitors was performed through the module Biotarget Finder as available in the www.drudit.com webserver.<sup>29</sup> The tool allows to predict the binding affinity of candidate molecules versus the selected biological target. The template of the biological target was built as previously reported. Thus, the in house database was submitted to the Biological Predictor module by setting the DRUDIT parameters, *N*, *Z*, and *G*, using the crystallized structure of **13b**, as previously reported.<sup>28</sup>

### Induced Fit Docking

Induced fit docking simulation was performed using the IFD application as available<sup>31,47</sup> in the Schrödinger software suite,<sup>48</sup> which has been demonstrated to be an accurate and robust method to account for both ligand and receptor flexibility.<sup>49</sup>

The IFD protocol was performed as follows:<sup>50,51</sup> the ligands were docked into the rigid receptor models with scaled down van der Waals (vdW) radii. The Glide Standard Precision (XP) mode was used for the docking and 20 ligand poses were retained for protein structural refinements. The docking boxes were defined to include all amino acid residues within the dimensions of 25 Å × 25 Å × 25 Å from the centre of the original ligands. The induced-fit protein-ligand complexes were generated using Prime software.<sup>52,53</sup> The 20 structures from the previous step were submitted to side chain and backbone refinements. All residues with at least one atom located within 5.0 Å of each corresponding ligand pose were included in the refinement by Prime. All the poses generated were then hierarchically classified, refined and further minimized into the active site grid before being finally scored using the proprietary GlideScore function defined as follows:  $XPG\_score = 0.065 \text{vdW} + 0.130 \text{Coul} + \text{Lipo} + \text{Hbond} + \text{Metal} + \text{BuryP} + \text{RotB} + \text{Site}$ , where vdW is the van der Waals energy term, Coul is the Coulomb energy, Lipo is a lipophilic contact term that rewards favourable hydrophobic interactions, Hbond is an H-bonding term, Metal is a metal-binding term (where applicable), BuryP is a penalty term applied to buried polar groups, RotB is a penalty for freezing rotatable bonds and Site is a term used to describe favourable polar interactions in the active site.

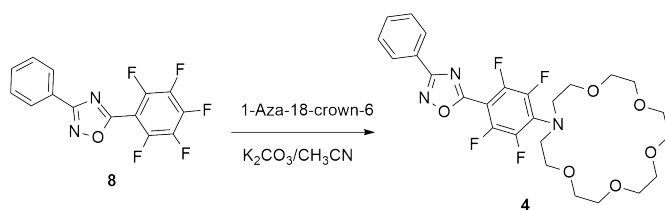
Finally, IFD\_score (IFD\_score =  $XPG\_score + 0.05 \text{Prime\_Energy}$ ), which accounts for both protein-ligand interaction energy and total energy of the system, was calculated and used to rank the IFD poses. More negative IFD\_score values indicated more favourable binding. Results are shown in Table 2.

## Chemical synthesis of inhibitors

Inhibitors **1**,<sup>54</sup> **3**,<sup>54</sup> **5**<sup>54</sup> and **6**<sup>55</sup> have been prepared as previously reported. Inhibitor **2** is commercial. Inhibitors **4**<sup>54</sup> and **7**<sup>54</sup> have been synthesized as described in detail in the next paragraphs. All solvent and reagents were used as received, unless otherwise stated. Melting points were determined on a hot-stage apparatus. <sup>1</sup>H-NMR and <sup>13</sup>C-NMR spectra were recorded at indicated frequencies, residual solvent peak was used as reference. Chromatography was performed by using silica gel (0.040-0.063 mm) and mixtures of ethyl acetate and petroleum ether (fraction boiling in the range of 40-60° C) in various ratios (v/v). Compounds **8**<sup>54</sup> and **9**,<sup>56</sup> used in the synthesis of inhibitors **4** and **7**, have been prepared as previously reported.



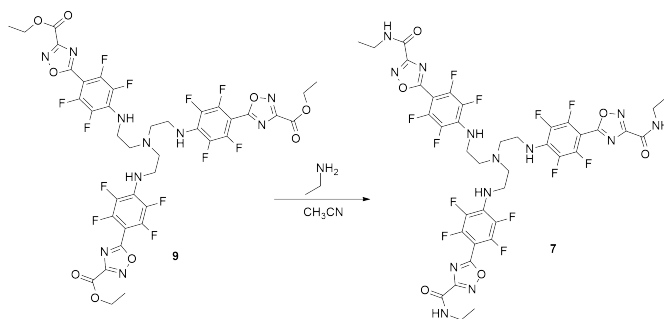
### Synthesis of inhibitor 4



**Figure 11.** Synthetic procedure for the obtainment of inhibitor 4.

Inhibitor **4** was synthesized through a nucleophilic aromatic substitution ( $S_NAr$ ) of 5-pentafluorophenyl-1,2,4-oxadiazole **8** with 1-Aza-18-crown-6 in para position (Fig. 11). Oxadiazole **8** (312 mg, 1 mmol) was dissolved in acetonitrile (5 mL). 1-Aza-18-crown-6 (289 mg, 1.1 mmol) and potassium carbonate (152 mg, 1.1 mmol) were added and the suspension was stirred at room temperature for 24 h. The reaction was monitored by TLC. The reaction mixture was dried under vacuum and treated with  $H_2O$  (50 mL) before extraction three times with EtOAc (50 mL each). The combined organic layers were dried with  $Na_2SO_4$  and then concentrated in vacuo to give the crude product, which was recrystallized from EtOH. **16-(2,3,5,6-tetrafluoro-4-(3-phenyl-1,2,4-oxadiazol-5-yl)phenyl)-1,4,7,10,13-pentaoxa-16-azacyclooctadecane**: yield: 73%. m.p.: 60-62° C.  $^1H$ -NMR (250 MHz,  $CDCl_3$ )  $\delta$ : 3.68-3.80 (m, 24H, overlapped  $-CH_2-$  signals), 7.51-7.54 (m, 3H, Ar), 8.16-8.20 (m, 3H, Ar). FTIR (Nujol) 1647, 1529, 1518  $cm^{-1}$ ; HRMS-ESI  $[(M+H)^+]$ :  $m/z$  calculated for  $(C_{26}H_{30}F_4N_2O_6)^+$ : 556.2060; found, 556.2046.

### Synthesis of inhibitor 7



**Figure 12.** Synthetic procedure for the obtainment of inhibitor 7.

Tripodal oxadiazolylamide (inhibitor **7**) was easily obtained by means of nucleophilic displacement with ethylamine from tripodal ester **9**, which was previously reported as heavy metal fluorescent sensor (Fig. 12).<sup>56</sup> Tripodal **9** (101 mg, 0.1 mmol) was dissolved in acetonitrile (3 mL). Ethylamine (2 M in MeOH, 150  $\mu$ L, 0.3 mmol) was added and the solution was stirred at room temperature for 24 h. The reaction was monitored by TLC. The reaction mixture was dried under vacuum and treated with  $H_2O$  (50 mL) before extraction three times with EtOAc (50 mL each). The combined organic layers were dried with  $Na_2SO_4$  and then concentrated in vacuo to give the crude product, which was purified by chromatography. **5,5',5''-(((nitriлотris(ethane-2,1-diyl))tris(azanediyl))tris(2,3,5,6-tetrafluorobenzene-4,1-diyl))tris(N-ethyl-1,2,4-oxadiazole-3-carboxamide)**: yield: 66%. m.p.: 188-189° C.  $^1H$ -NMR (300 MHz,  $DMSO-d_6$ )  $\delta$ : 1.18 (t, 9H,  $J = 6.9$  Hz,  $CH_3$ ), 2.85 (bs, 6H,  $NCH_2CH_2NH-$ ), 3.34-3.48 (m, 6H,  $CH_3CH_2NH-$ ), 3.56 (bs, 6H,  $NCH_2CH_2NH-$ ), 6.91 (bs, 3H,  $NCH_2CH_2NH-$ ), 9.02 (t, 3H,  $J = 5.7$  Hz,  $NHCH_2CH_3$ ). FTIR (Nujol) 3325, 1701, 1680, 1647  $cm^{-1}$ ; HRMS-ESI  $[(M+H)^+]$ :  $m/z$  calculated for  $(C_{39}H_{34}F_{12}N_{13}O_6)^+$ : 1007.2485; found, 1007.2518.

## References

1. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473 (2020).
2. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270+ (2020).
3. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. & Hilgenfeld, R. Coronavirus main proteinase (3CL(pro)) structure: Basis for design of anti-SARS drugs. *Science* **300**, 1763–1767 (2003).

4. Morse, J. S., Lalonde, T., Xu, S. & Liu, W. R. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem* **21**, 730–738 (2020).
5. Zhang, D. *et al.* Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. *Infection genetics and evolution* **10**, 84–88 (2010).
6. Goyal, B. & Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Combinatorial Science* **22**, 297–305 (2020).
7. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. *Science* **368**, 409–412 (2020).
8. Shi, J. & Song, J. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *FEBS Journal* **273**, 1035–1045 (2006).
9. Anand, K. *et al.* Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *Embo Journal* **21**, 3213–3224 (2002).
10. Chen, S. *et al.* Severe acute respiratory syndrome coronavirus 3C-like proteinase n terminus is indispensable for proteolytic activity but not for enzyme dimerization - Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *Journal of Biological Chemistry* **280**, 164–173 (2005).
11. Chou, C. *et al.* Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* **43**, 14958–14970 (2004).
12. Wei, P. *et al.* The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochemical and Biophysical Research Communications* **339**, 865–872 (2006).
13. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y. & Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *Journal of Medicinal Chemistry* **59**, 6595–6628 (2016).
14. Spinozzi, F., Ferrero, C., Ortore, M. G., Antolinos, A. D. M. & Mariani, P. GENFIT: software for the analysis of small-angle X-ray and neutron scattering data of macromolecules in-solution. *J. App. Cryst.* **47**, 1132–1139 (2014).
15. Ortore, M. G. *et al.* Combining structure and dynamics: non-denaturing high-pressure effect on lysozyme in solution. *J. R. Soc. Interface* **6**, S619–S634 (2009).
16. Teixeira, J. Small-angle scattering by fractal systems. *J. Appl. Cryst.* **21**, 781–785 (1988).
17. Graziano, V., McGrath, W. J., Yang, L. & Mangel, W. F. SARS CoV Main Proteinase: The Monomer-Dimer Equilibrium Dissociation Constant. *Biochemistry* **45**, 14632–14641 (2006).
18. Apenten, R., Khokhar, S. & Galani, D. Stability parameters for  $\beta$ -lactoglobulin thermal dissociation and unfolding in phosphate buffer at pH 7.0. *Food Hydrocolloids* **16**, 95 – 103 (2002).
19. Janin, J. Elusive affinities. *Proteins: Structure, Function, and Bioinformatics* **21**, 30–39 (1995).
20. Horton, N. & Lewis, M. Calculation of the free energy of association for protein complexes. *Protein Science* **1**, 169–181 (1992).
21. Svergun, D. *et al.* Protein hydration in solution: experimental observation by X-ray, neutron scattering. *Proc. Natl. Acad. Sci. USA* **95**, 2267–2272 (1998).
22. Sinibaldi, R. *et al.* Preferential hydration of lysozyme in water/glycerol mixtures: A small-angle neutron scattering study. *J. Chem. Phys.* **126**, 235101–235109 (2007).
23. Sinibaldi, R. *et al.* Sans/saxs study of the bsa solvation properties in aqueous urea solutions via a global fit approach. *Eur. Biophys. J.* **37**, 673–681 (2008).
24. Greenfield, N. J. & Fasman, G. D. Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry* **8**, 4108–4116 (1969).
25. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* **1**, 2527–2535 (2006).
26. Abian, O. *et al.* Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. *International Journal of Biological Macromolecules* **164**, 1693 – 1703 (2020).
27. Shi, J., Wei, Z. & Song, J. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme - Defining the extra domain as a new target for design of highly specific protease inhibitors. *Journal of Biological Chemistry* **279**, 24765–24773 (2004).

28. Martorana, A., Gentile, C. & Lauria, A. In Silico Insights into the SARS CoV-2 Main Protease Suggest NADH Endogenous Defences in the Control of the Pandemic Coronavirus Infection. *Viruses* **12**, 805 (2020).
29. Lauria, A. *et al.* DRUDIT: web-based DRUGs DIScovery Tools to design small molecules as modulators of biological targets. *Bioinformatics* **36**, 1562–1569 (2019).
30. Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **47**, 1739–1749 (2004).
31. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *Journal of Medicinal Chemistry* **49**, 534–553 (2006).
32. Lee, J. *et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *NATURE COMMUNICATIONS* **11** (2020).
33. Yang, H. *et al.* Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLOS Biology* **3**, e324 (2005).
34. Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the sars-cov-2 main protease. *Science* **368**, 1331–1335 (2020).
35. Sacco, M. D. *et al.* Structure and inhibition of the sars-cov-2 main protease reveal strategy for developing dual inhibitors against mpro and cathepsin l. *Science Advances* **6** (2020).
36. Jin, Z. *et al.* Structure of M<sup>PRO</sup> from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020).
37. Kneller, D. W., Phillips, G., Kovalevsky, A. & Coates, L. Room-temperature neutron and X-ray data collection of 3CL M<sup>Pro</sup> from SARS-CoV-2. *Acta Crystallographica Section F* **76**, 483–487 (2020).
38. Suárez, D. & Díaz, N. Sars-cov-2 main protease: A molecular dynamics study. *JCIM* **0**, null (2020).
39. Kell, G. S. Density, thermal expansivity, and compressibility of liquid water from 0° C to 150° C. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data* **20**, 97–105 (1975).
40. Schrödinger Release 2017, N. Y., LLC. LigPrep (2017).
41. Schrödinger Suite 2017-2, N. Y., LLC. Epik, Protein Preparation Wizard (2017).
42. Banks, J. L. *et al.* Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of Computational Chemistry* **26**, 1752–1780 (2005).
43. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nature Structural Biology* **10**, 980–980 (2003).
44. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **27**, 221–234 (2013).
45. Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry* **49**, 6177–6196 (2006).
46. Halgren, T. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* **47**, 1750–1759 (2004).
47. Sherman, W., Beard, H. S. & Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. *Chemical Biology Drug Design* **67**, 83–84 (2006).
48. Schrödinger Release 2012, N. Y., LLC. Maestro, version 9.3 (2012).
49. Zhong, H., Tran, L. M. & Stang, J. L. Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. *Journal of Molecular Graphics and Modelling* **28**, 336–346 (2009).
50. Wang, H., Aslanian, R. & Madison, V. S. Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17 $\alpha$  pocket flexibility. *Journal of Molecular Graphics and Modelling* **27**, 512–521 (2008).
51. Luo, H.-J., Wang, J.-Z., Deng, W.-Q. & Zou, K. Induced-fit docking and binding free energy calculation on furostanol saponins from *Tupistra chinensis* as epidermal growth factor receptor inhibitors. *Medicinal Chemistry Research* **22**, 4970–4979 (2013).
52. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics* **55**, 351–367 (2004).



53. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of Molecular Biology* **320**, 597–608 (2002).
54. Buscemi, S., Pace, A., Palumbo Piccionello, A. & Vivona, N. Synthesis of fluorinated first generation starburst molecules containing a triethanolamine core and 1,2,4-oxadiazoles. *Journal of Fluorine Chemistry* **127**, 1601–1605 (2006).
55. Martorana, A., Palumbo Piccionello, A., Buscemi, S., Giorgi, G. & Pace, A. Synthesis of 4(5)-phenacyl-imidazoles from isoxazole side-chain rearrangements. *Organic and Biomolecular Chemistry* **9**, 491–496 (2011).
56. Pibiri, I. *et al.* Fluorescent Hg<sup>2+</sup> sensors: Synthesis and evaluation of a tren-based starburst molecule containing fluorinated 1,2,4-oxadiazoles. *European Journal of Organic Chemistry* **2010**, 4549–4553 (2010).

## Acknowledgements

We wish to thank Charlotte Edwards-Gayle for her assistance during the SAXS experiments at DIAMOND.

## Author contributions statement

P.Mariani, L.Si., A.P. and F.S. conceived and designed the work. L.Si., L.Sa. and N.B. synthesised M<sup>Pro</sup>. A.P., C.P., L.C. and V.L. collected and analysed CD data. A.L. performed docking. F.S., Y.G., P.Mariani and M.G.O. collected and analyzed SAXS data. A.P.P. and P.Marzullo synthesized inhibitors. A.P. and F.S. took overall responsibility for the manuscript. All authors approved the final version of the manuscript.