UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVI CICLO DEL DOTTORATO DI RICERCA IN CHIMICA

Integrative approach in drug discovery against SARS-CoV2: a biochemical and biophysical study on the Papain-like protease

Settore scientifico-disciplinare: CHIM/06

DOTTORANDO

STEFANO MORASSO

COORDINATORE

PROF. ENZO ALESSIO

SUPERVISORE DI TESI

Dr. PAOLA STORICI

CO-SUPERVISORE DI TESI

PROF. FEDERICO BERTI

ANNO ACCADEMICO 2022/2023

Table of content

1.Abstract

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2), the etiological agent of Covid-19 syndrome, caused one of the most impacting pandemics in human history. Vaccination campaigns offered the solution to the global emergency, but up to now only a few treatments of the disease have been approved. Among potential pharmacological targets of the SARS-CoV2, the Cysteine proteases are one of the most promising, due to their essential role in viral replication and additional functions in the infection development: the more explored Main protease (Mpro) and the Papain-like protease (PLpro). The second one is a domain of the sizeable non-structural protein 3 (nsp3) and besides the cleavage of the viral polyprotein pp1a in three sites, shows deISGylating and deUbiquitinating activity. PLpro represents indeed one of the main viral protecting systems against the immune response, interfering with the cytokines pathways regulated by the Ubiquitin and ISG15 modification system. Despite the number of drug screenings performed on this target, only a few molecules were validated as inhibitors. In this thesis are presented our efforts to identify ensured repurposing inhibitors of this enzyme and the characterization of differences between the commonly used construct of the PLpro and the double-domain construct PLpro_NAB, containing the Nucleic-Acid Binding domain of nsp3, both produced in the Elettra Protein Facility. The project was developed in collaboration with valuable partners who provided their skills, leading to the identification of interesting repurposing inhibitors. After critically evaluating the first results, applying an integrative approach of biochemical, computer-aided and biophysical techniques, we elucidated the mechanism of action of false-positive compounds which at first looked promising, highlighting instead a real inhibitor, the CPI-169. The μM affinity was suitable to be studied by ligand-based NMR techniques, which allowed the validation of the computational and biochemical readouts. The interaction with the preferred human substrate of the PLpro, the ISG15, was also investigated in this project. The production of ISG15 and its precursor (proISG15) was necessary to highlight interesting discrepancies in the binding and catalytic activity of the PLpro_NAB mutants $S_{466}R$ and $T_{467}K$, two mutations discovered in the Delta-variant of SARS-CoV2. Our results show that the point mutations located on the NAB domain alter the neighbour domain, the PLpro, and how a single point mutation on another domain of the nsp3 can also affect the PLpro.

RIASSUNTO

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2), l'agente eziologico della sindrome Covid-19, ha causato una delle pandemie più impattanti della storia umana. Le campagne di vaccinazione hanno offerto la soluzione all'emergenza globale, ma finora sono stati approvati solo pochi trattamenti della malattia. Tra i potenziali bersagli farmacologici del SARS-CoV2, le cistein-proteasi sono tra le più promettenti, grazie al loro ruolo essenziale nella replicazione virale e per le loro funzioni aggiuntive nello sviluppo dell'infezione: la più studiata Main protease (Mpro) e la Papain-like protease (PLpro). La seconda è un dominio della proteina non strutturale 3 (nsp3) e, oltre al clivaggio della poli-proteina virale pp1a in tre siti, mostra attività deISGilasica e deUbiquitinasica. La PLpro rappresenta infatti uno dei principali sistemi di protezione del coronavirus contro la risposta immunitaria, interferendo con le vie delle citochine regolate dal sistema di modificazione dell'Ubiquitina e dell'ISG15. Nonostante il numero di screening farmacologici effettuati su questo target, solo poche molecole sono state validate come inibitori. In questa tesi vengono presentati i nostri sforzi per identificare inibitori accertati di questo enzima e la caratterizzazione delle differenze tra il costrutto comunemente usato della PLpro e il costrutto a doppio dominio PLpro_NAB, contenente il Nucleic-Acid Binding domain dell'nsp3, entrambi prodotti nella protein facility di Elettra. Il progetto è stato sviluppato in collaborazione con validi partner che hanno messo a disposizione le loro competenze, portando all'identificazione di interessanti inibitori di riposizionamento. Dopo aver valutato criticamente i primi risultati, applicando un approccio integrativo di tecniche biochimiche, computazionali e biofisiche, abbiamo delucidato il meccanismo d'azione di composti falsi-positivi che all'inizio sembravano promettenti, evidenziando invece un vero inibitore, il CPI-169. Grazie all'affinità μM della molecola, è stato possibile studiare l'interazione con tecniche di ligand-based NMR, che hanno permesso di validare i risultati computazionali e biochimici. In questo progetto è stata investigata anche l'interazione con il substrato umano di elezione della PLpro, l'ISG15. La produzione di ISG15 e del suo precursore (proISG15) è stata necessaria per evidenziare interessanti differenze nel legame e nell'attività catalitica dei mutanti $S_{466}R$ e $T_{467}K$ della PLpro_NAB, due mutazioni scoperte nella variante Delta del SARS-CoV2. Le nostre osservazioni dimostrano che le mutazioni puntiformi localizzate sul dominio NAB alterano il dominio vicino, la PLpro, e come una singola mutazione puntiforme su un altro dominio dell'nsp3 possa influenzare anche il PLpro.

2. Introduction

2.1 Severe Acute Respiratory Syndrome Coronavirus 2

2.1.1 SARS-CoV2 overview

Coronaviruses are positive-sense single-stranded RNA retroviruses, etiological agents of respiratory and intestinal syndromes in animals and humans. These viruses are a subfamily of *Coronaviridae*, consisting of four genera on the basis of their phylogenesis: Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus [1]. While Gamma- and Deltacoronaviruses infect birds and rarely mammals, Alpha- and Betacoronaviruses drew the attention of humans in the early 2000s when the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) appeared, qualifying as first coronavirus to cause an impacting syndrome in humans, followed by Middle-East Respiratory Syndrome coronavirus (MERS-CoV) [2].

The third coronavirus that achieved a burden on human health is the "Severe Acute Respiratory Syndrome Coronavirus 2", which rapidly and masterly caused one of the most dramatic pandemics that humans faced in history. SARS-CoV2 is a betacoronavirus which shares 79% genome sequence identity with SARS-CoV and 50% with MERS-CoV [3]. Considering the syndrome it causes, the phylogenetic analysis and the high homology with SARS-CoV, the new coronavirus was clustered in SARS-related coronaviruses and inherited the name from the first one [4]. The SARS-CoV2 genome is ca. 30 kb long with 14 open reading frames (ORFs) and encodes for 29 proteins. Approximately two-thirds of the genome from 5' encodes for two overlapping polyproteins, pp1a and pp1ab, which encode for the non-structural proteins (nsp) that will be released and activated by proteolytic cleavage (Figure 1a). Most SARS-CoV2 nsp(s) have more than 85% amino acid sequence identity with SARS-CoV nsp(s). Close to the 3' end of the viral genome, there are the structural proteins: the Nucleocapsid protein (N), the Spike protein (S), the Membrane protein (M) and the Envelope protein (E). Between the genes encoding for the structural proteins, there are ancillary genes encoding for the accessory proteins (ORF 3a, 3b, 6, 7a, 7b, 8b, 9b, 14) [5].

The S protein is a highly glycosylated homotrimer, which could be cleaved by a furin-like protease into two subunits (S1 and S2). The subunits interact with each other even after the entrance into the cell. The Receptor-Binding Domain (RBD) is located in the first subunit, composed of a five-stranded antiparallel β-strand core. Between, the receptor binding motif interacts with the human Angiotensin-Converting Enzyme 2 (hACE2), exposed on the host cell membrane. When the RBD of S1 undergoes a transient and less stable hinge-like conformation ("up" conformation), the determinant region of the RBD located between β4 and β5 is exposed to the hACE2 and allows the recognition with the receptor [6]. Once the virus is endocytosed into the

cell, thanks to the interaction of the S protein to the hACE2 the E protein mediates the viral lysis with the consequent genome release. Once translated, the two cysteine proteases release the nsp(s) from polyproteins ORF1a and 1ab, enabling replication. The replication and transcription complex (RTC) is formed by the complex of nsp12, which contains the catalytic core, the nsp8 and the nsp7. These last ones ensure continuous production of RNA, activating the nsp12. The subgenome synthesis is completed inside the host cell, therefore the Elongation RTC is formed, after the check performed by the exonuclease nsp14. Two coupled copies of nsp13 helicase (nsp13-1 and nsp13-2) join the complex, interacting in the Zn-binding domains with the N-terminal helices of the nsp8. The RNA-binding protein nsp9 interacts with the nsp12 and recruits the nsp10-14 complex to conclude the final capping RTC assembly (Figure 1) [7,8].

The structural proteins assemble and assist the formation of the virions between the endoplasmic reticulum and the Golgi compartment. Unlike other coronaviruses, some betacoronaviruses, including the SARS-CoV2, egress the host cell via lysosomal trafficking instead of exocytosis [9].

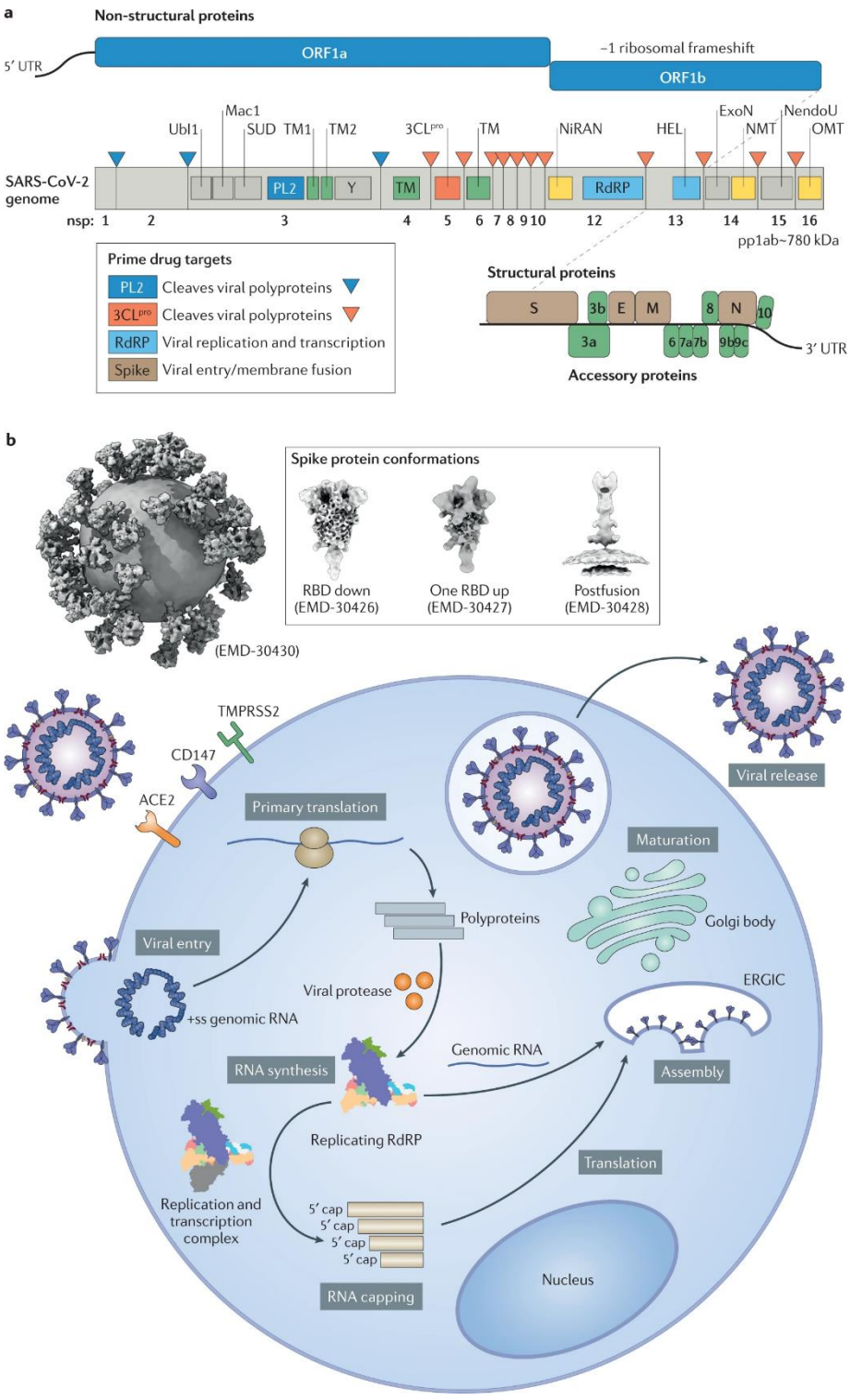


Figure 1: Viral replication cycle of SARS-CoV2. [a] Schematic representation of the viral genome with encoded proteins; [b] Viral cycle of the SARS-CoV2: entry, replication, assembly, and release (from Yang H. and Rao Z., 2021) [8]

2.1.2 COVID-19: syndrome and therapy

Since the first isolation and identification in December 2019, SARS-CoV2 has caused more than 6.9 million deaths worldwide. Even if the pandemic emergency was affirmed "over", in the month between the 10th of July and the 6th of August of this year nearly 1.5 million new cases of Covid-19 were reported (WHO Weekly Epidemiological Update, August 2023).

SARS-CoV2 causes the syndrome called Covid-19. The symptoms onset after a few days from the infection, with a flu-like syndrome that can evolve into severe pneumonia, dyspnea, organ dysfunction and, in worst cases, death. Since more than 273 million cases of Covid-19 were confirmed, risk and protective factors were outlined, including demographic (age, sex, etc) and clinical factors (cardiovascular pathologies, respiratory syndromes, immunodeficiency) [10]. Thanks to the provision of time-record vaccines, the large campaigns of immunization led to the progressive reduction of hospitalizations and severe syndrome rates [11]. However, despite the number of active compounds against SARS-CoV2 targets that have been discovered, only a few molecules have been approved for the treatment of the disease, any of those with a clear and structured development. Approved drugs could be divided into two groups: i) SARS-CoV2 selective drugs, with proven effectiveness towards the viral replication cycle; ii) drugs towards host proteins involved in viral replication. In the first group, we can find small molecules and biologics which target the RNA-dependent polymerase, the Main-protease or the interaction between the S protein and the hACE2. The selectivity is high, but could not become effective for new variants. In the second type of molecules, since the target is not viral but human, they may exhibit broad-spectrum antiviral activity, decreased selectivity and potentially poor safety profiles though. In the same group also ancillary treatments are present, especially towards the cytokine storm induced by the infection, which may result necessary to limit the excessive inflammatory reaction [12].

Table: Approved drugs to treat the Covid-19 (from Li G. et al., 2023)

Drug name	Type (delivery route)	Use	Eligible patients	Resistance likelihood	Status
RdRp inhibitors					
Remdesivir (Veklury)	Small molecule (i.v.)	Tx	Outpatients ≤7 days of symptom onset, or inpatients	Low	Approved by the FDA, EUA in many countries (including EU)
Molnupiravir (Lagevrio)	Small molecule (oral)	Tx	Outpatients ≥18 years old and ≤5 days of symptom onset	Low	Approved in the UK, EUA in many countries
JT001 (VV116)	Small molecule (oral)	Tx	Outpatients ≤5 days of symptom onset	Low	Approved in Uzbekistan
M^{pro} inhibitors					
Nirmatrelvir-ritonavir (Paxlovid)	Small molecule (oral)	Tx	Outpatients ≤5 days of symptom onset	Low	Approved in the UK and EU; EUA in many countries
Ensitrelvir (Xocova)	Small molecule (oral)	Tx	Outpatients ≤5 days of symptom onset	Low	EUA in Japan, phase III
Inhibitors that block the spike-ACE2 interaction					
Bebtelovimab	mAb (i.v.)	Tx	Outpatients ≤7 days of symptom onset	High (e.g., BQ.1, BQ.1.1)	EUA by the FDA; paused owing to resistance

Regdanvimab (Regkirona)	mAb (i.v.)	Tx	Outpatients ≤7 days of symptom onset	High (e.g., Omicron, Gamma, Beta)	EUA in many countries; paused owing to resistance
Sotrovimab	mAb (i.v.)	Tx	Outpatients ≤7 days of symptom onset	High (e.g., Omicron)	Approved or EUA in many countries; paused owing to resistance
Amubarvimab and romlusevimab	mAbs (i.v.)	Tx	Outpatients ≤10 days of symptom onset	High (e.g., Omicron)	Approved in China; discontinued
Bamlanivimab and etesevimab	mAbs (i.v.)	Tx	Outpatients ≤10 days of symptom onset	High (e.g., Omicron, beta)	EUA in many countries; paused owing to resistance
		PEP	Certain individuals at high risk of COVID-19		
Casirivimab and imdevimab (REGEN-COV)	mAbs (i.v. or s.c.)	Tx	Outpatients ≤10 days of symptom onset	High (e.g., Omicron)	EUA in many countries, paused owing to resistance
		PEP	Certain individuals at high risk of COVID-19		
Cilgavimab and tixagevimab (Evusheld)	mAbs (i.m.)	PrEP	Certain individuals at high risk of COVID-19	High (e.g., Omicron)	Approved or EUA in many countries, paused owing to resistance
Glucocorticoids					

Dexamethasone	Small molecule (i.v.)	Tx	Inpatients requiring oxygen support	No	Recommended by COVID-19 guidelines
Hydrocortisone	Small molecule (i.v.)	Tx	Inpatients requiring oxygen support	No	Recommended by COVID-19 guidelines
Janus kinase inhibitors					
Baricitinib	Small molecule (oral)	Tx	Inpatients requiring oxygen support	No	Recommended by COVID-19 guidelines
Tofacitinib	Small molecule (oral)	Tx	Inpatients requiring oxygen support	No	Recommended by COVID-19 guidelines
Cytokine antagonists					
Tocilizumab	Anti-IL-6R mAb (i.v.)	Tx	Inpatients receiving systemic corticosteroids and requiring oxygen support	No	Recommended by COVID-19 guidelines
Sarilumab	Anti-IL-6R mAb (s.c.)	Tx	Inpatients receiving systemic corticosteroids and requiring oxygen support	No	Recommended by COVID-19 guidelines
Anakinra	IL-1R antagonist (s.c.)	Tx	Inpatients requiring oxygen support	No	EUA by the FDA; authorized in the EU
Anticoagulants					

Various drugs (such as low-molecular-weight heparin)	Various (i.v., s.c. or oral)	Tx, TP	Non-ICU inpatients with no pregnancy	No	Recommended by COVID-19 guidelines
Anti-C5a inhibitors					
Vilobelimab	mAb (i.v.)	Tx	Hospitalized adults initiated ≤48 hours of oxygen support	No	EUA by the FDA

2.1.3 Non-structural protein 3

The non-structural protein 3 (nsp3) is the largest nsp protein of SARS-CoV2 (1945 amino acids), consisting of up to 17 domains, depending on the definition of domain borders. The nsp3 is cut from the entire polyprotein by the PLpro present in the middle of the nsp3 itself, then cooperates with the nsp4 to form double-membrane vesicles, which protects the replication machinery from host proteases [13]. Two transmembrane domains create a pore through the vesicle which allows the release of the new RNA to be packed into the virions [14]. This is not the unique function of the nsp3, the complexity of this multidomain protein is indeed in the multiplicity of interactions and functions of its single domains, which makes this protein the "Swiss army knife" of the coronavirus, as it was defined by Lea C. von Soosten et al. (Figure 2) [15].

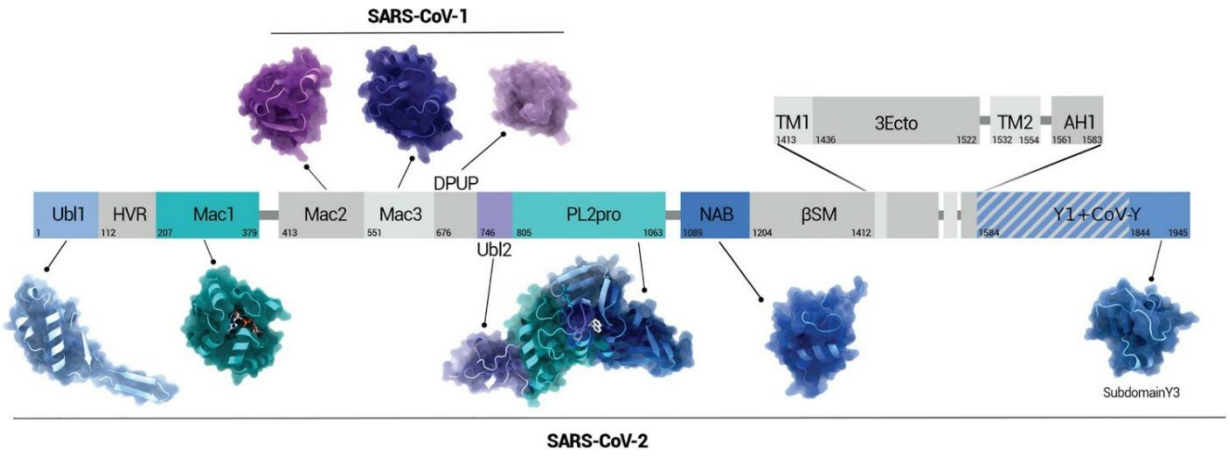


Figure 2: Nsp3 domains in sequence, from Von Soosten LC et al. Mac2 and Mac3 structures of SARS-CoV2 are now deposited in the PDB.

Nsp 3a

The nsp3a is represented by the Ubiquitin-like domain 1 (Ubl1) and a Glu-rich region, also known as the hypervariable region. The function of Ubl1 has not been completely understood, there is evidence that interacts with the N protein, driving the encapsulation of the RNA into virions [16]. Due to its similarity with ubiquitin and the ISG15, Ubl1 interferes with host proteins involved in antiviral modulation that target ubiquitinated or ISGylated proteins [17].

Nsp3b

The nsp3b is the Macrodomain1, a mono(ADP-ribosyl)-hydrolase. Despite other coronaviruses, which bind and hydrolase poly-ADP-ribosyl chains, SARS-CoV macrodomain1 recognises the monomer and thus antagonizes the action of the hPARP14, which rybosilates and activates host proteins involved in the antiviral response [18]. Therefore, this could represent an interesting target for therapy. The overall structure is formed by two layers of helices and β-strands, with the active site (ADP-ribose binding site) in the middle of the sheet layer and the N-terminal of the first helix (PDB 7KQO). Several structures are available in the PDB, as apo form, with the substrate in the catalytic pocket or with inhibitors.

Nsp3c

SARS-unique domains (SUD) are typical of sarbecoviruses, including SARS. In SARS-CoV2, two SUD (N and M) are in sequence that are macrodomains according to their fold. Studies on SUD-N and M of SARS-CoV1 suggest that SUD-N interacts with important host factors in the translation process, interacting with the PAIP1. SUD-M binds oligo(G)-containing nucleic acids which tend to form G-quadruplex. Both domains are necessary for the viral replication [19].

Nsp3d

The Nsp3d consists of the Ubl2 and the PLpro, which will be explored in more detail in the next chapters.

Nsp3e

The small Nucleic Acid Binding domain and the betacoronavirus-specific marker (βSM) are unique to betacoronaviruses. Only one structure of the SARS-CoV2 NAB domain is available (PDB 7LGO), but no corresponding published data. The investigation of the activity of SARS-CoV1 NAB domain revealed that this small domain binds guanine-rich ss-RNA, specifically at the level of positive amino acids on its surface [17].In early 2021, the Delta variant of SARS-CoV2 (B.1.617.2) quickly spread from India to Europe, exhibiting a more severe COVID-19 syndrome concerning Alpha or Omicron variants. In the Delta-variant (δ-variant), among the relevant mutations detected, two mutations in the nsp3 NAB domain were highlighted in terms of frequency and related to a significant aggressiveness of the disease, the $T_{2016}K$ and $S_{2015}R$ [68,69]. The molecular relevance of such point mutations as well as their involvement in the disease have not been understood yet.

Transmembrane domains and ectodomain

The transmembrane domains TM1 and TM2 anchor the nsp3 to the vesicle or the endoplasmic reticulum membrane, the ectodomain is thus exposed to the lumenal side. This domain in between seems to interact with the nsp4 and allows the formation of the pore on the vesicle [15].

Nidovirus-conserved domain and coronavirus-specific carboxyl-terminal domain

The C-terminal region of the nsp3 is composed of the Nidovirus-conserved domain of unknown function (Y) and the coronavirus-specific carboxyl-terminal domain (CoV-Y), divided into two domains (Y2 and Y3). The Y domain is conserved among the order of nidoviruses, the CoV-Y is typical of coronaviruses. Studies have shown that these C-terminal domains promote nsp3 and nsp4 interaction in the vesicle formation [17].

2.2 SARS-CoV2 Cys proteases

2.2.1 Main protease (Mpro)

The Mpro, also named 3-chymotrypsin-like protease or 3CL-pro, is one of the two Cys-proteases of SARS-CoV2. It is part of the viral polyprotein 1a, representing the nsp5. Its main function is to catalyse the cleavage of viral proteins including the nsps, starting from the excision of itself from the polyprotein [9]. The Mpro is plainly active as a homodimer that folds in a "heart-shaped" conformation (Figure 3a), whereas the single monomer is almost inactive [20]. Each monomer is composed of three domains: two β-barrel domains (residues 8-101 and 102-184) and a five α-helix third domain which largely contributes to the dimer formation (Figure 3b). The non-canonical Cys145-His41 catalytic dyad is located between the two β-barrel domains, exposed to the protein's surface. A conserved water molecule proximal to the His41 seems fundamental to the catalytic activity (Figure 3c) [21].

The essential role in the viral replication cycle, the highly conserved structure among coronaviruses and the low homology with human proteins make the Mpro a suitable target for pharmacological development. Its druggability is proven by the huge number of co-structures with ligands deposited in the PDB (more than 800), becoming one of the main targets of interest of the SARS-CoV2. As a matter of fact, one of the few approved therapies against Covid-19 is the Nirmatrelvir, a potent inhibitor of the Mpro, formulated with Ritonavir [22]. Besides this commercially available formulation and the novel orally available inhibitor from Pfizer PF-07321332 [23], a number of compounds were tested towards this target. Even if some allosteric cavities were identified, most of the ligands discovered bind the Mpro in the active-site cavity, which allows a variety of different chemical moieties to be bound, due to its malleability [24]. Therefore, various novel, repurposing and natural compounds appeared bound in Mpro crystal structures, with covalent and non-covalent interactions. The Elettra Protein Facility also worked on this target, describing the crystal structure in complex with a number of interesting compounds identified through repurposing screenings such as natural flavonoids (i.e. Myrycetin) or peptidomimetic inhibitors (MG-132, Figure 4) [27] or through de-novo synthetic design. In this context, I joined the Mpro project working on the characterization of novel inhibitors, in collaboration with the Medicinal Chemistry group of Prof. Summa from the University of Naples [113].

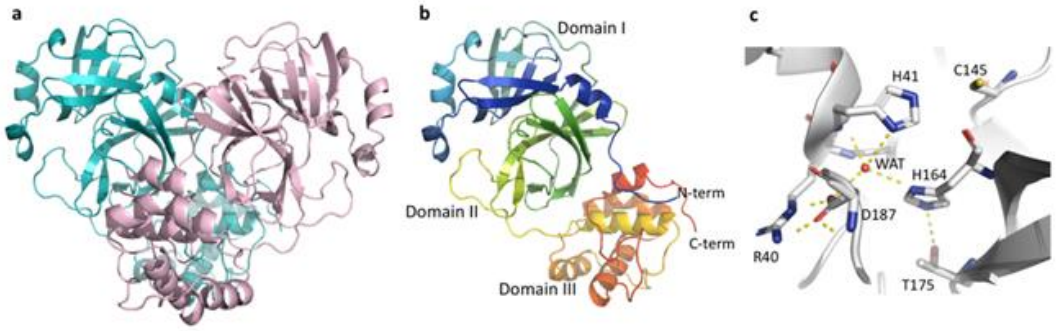


Figure 3: Mpro crystal structure (PDB 7BB2). a. Homodimer (chains A and B are coloured in cyan and pink); b. structure of the monomer with the indication of the three domains; c. Active site of the Mpro (from Morasso S et al, 2023) [25,26]

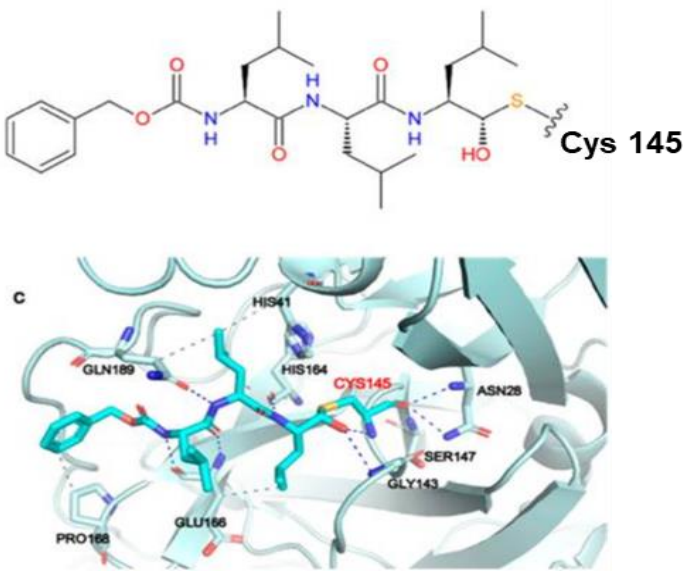


Figure 4: Crystal structure of MG-132 covalently bound to the Cys145 of Mpro (from Costanzi E. et al, 2021) [27]

2.2.2 Papain-like protease (PLpro)

PLpro is the second Cys-protease of SARS-CoV2. It is well-conserved among coronaviruses, usually in two copies, but in SARS-CoV1 and 2 only one is present [17]. The PLpro is located between the SUD and NAB domains in the nsp3 and is indispensable for viral replication as much as the Mpro. It recognizes the consensus sequence "LXGG↓XX" (X generic amino acid) in three sites on the pp1a, between nsp1 and nsp2 (LNGG↓AYTR), nsp2 and nsp3 (LKGG↓APTK), nsp3 and nsp4 (LKGG↓KIVN) [28]. The same sequence can be found in host proteins. In particular, the sequence constitutes the C-terminal of Ubiquitin (Ub) and the Ubiquitin-like protein ISG15. PLpro shows indeed deubiquitinating and deISGylating activity, targeting the modified host proteins involved in inflammation and immune system response pathways (Figure 5). PLpro is able to interfere with the NF-kB pathway and prevents the translocation to the nucleus of the interferon regulatory factor 3 (IRF3), both involved in the production of antiviral cytokines [17]. Therefore, PLpro is not necessary to the virus just for its essential role in the replication, but also to escape the immune system reaction. Notably, SARS-CoV1 and SARS-CoV2 PLpro have differences in preferences for their substrates, even though they share more than 80% identity. The first one favours chains of Ub, while SARS-CoV2 PLpro shows a much higher affinity for the ISG15. Intriguingly, the cleavage efficiency toward the ISG15 is ca 2500-3000 times higher than towards the pp1a [29].

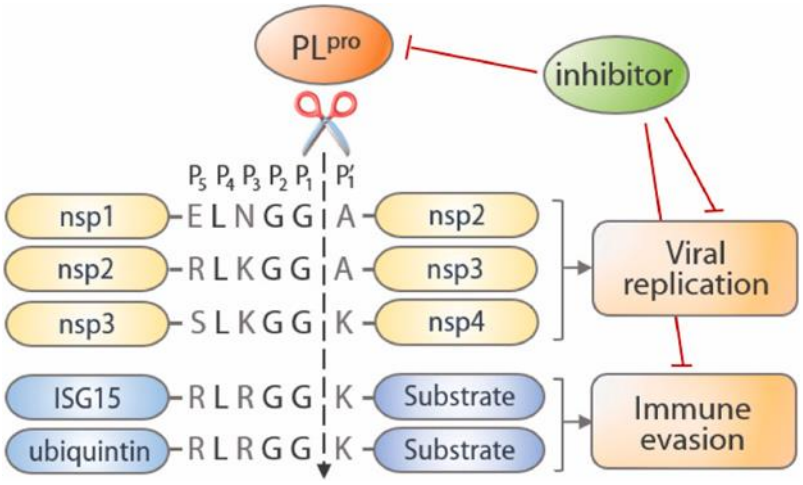


Figure 5: Schematic representation of cleaving activity of the PLpro towards viral and host targets (from Lingyu Li et al., 2022) [30]

Structural insights

SARS-CoV2 PLpro follows a particular architecture common in the deubiquitinases, in particular in the ubiquitin-specific protease family (USP), called “finger-thumb-palm”, from the shape reminding an open right hand. Before this triple domain, at the N-terminal is present a small Ubl domain (Ubl2). The 60 amino acids long Ubl2 function remains unclear, but its high conservation among coronaviruses suggests it could have an important regulatory role [17]. Ubl2 consists of five β-strands, one α-helix and one 3_{10}-helix; a small loop projects to the catalytic core, formed by residues 9-14. The “thumb” domain comprises six helices and a small β-hairpin. Due to its major flexibility, the “finger” conformation is the most variable among PLpro structures. It is made of six β-strands and two α-helices, with a Zinc ion coordinated by four cysteines on the extremity of the finger (Cys 189, 192, 224, and 226) (Figure 6). The Zinc ion is essential for the PLpro folding and, therefore, for its activity. The catalytic triad (Cys111, His 272, and Asp286) lies in the “palm” domain, at the interface with the thumb. This domain is organized into six β-strands, with a mobile β-loop formed by Gly266–Gly271, known as “blocking loop 2” (BL2), involved in the recognition of the consensus sequence of substrates (Figure 7) [17, 31].

Two binding sites could be identified (S1 and S2) for the cleavage of the substrates. The S2 site consists of the Phe69 and binds preferentially poly-chains of Ub linked at its residue Lys48. Compared to the SARS-CoV1 S1 site, six residues are mutated, explaining the major deubiquitinating activity of SARS-CoV1 PLpro than SARS-CoV2 PLpro [29]. The S1 binding site is more difficult to locate, varying according to the substrate: Ub binds between the palm and the finger domain, with its C-terminus entering the catalytic pocket; the ISG15 binds the palm domain as well, but interacts with the thumb domain instead of the finger, precisely the 7^{th} α-helix of the thumb and the Trp123 and Pro130/Glu132 of ISG15 [32].

Currently, 56 structures of the PLpro are available in the PDB, most of them realized using the inactive mutant of the catalytic Cys111 into Ser.

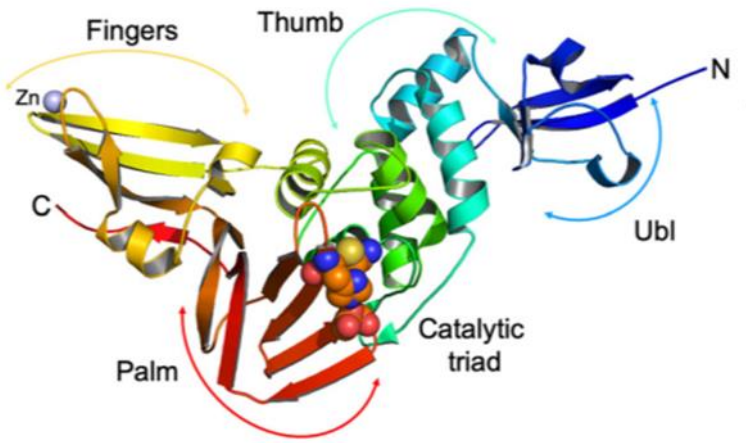


Figure 6: Structure of the PLpro the subdomains differently coloured (from Osipiuk J et al., 2021) [31]

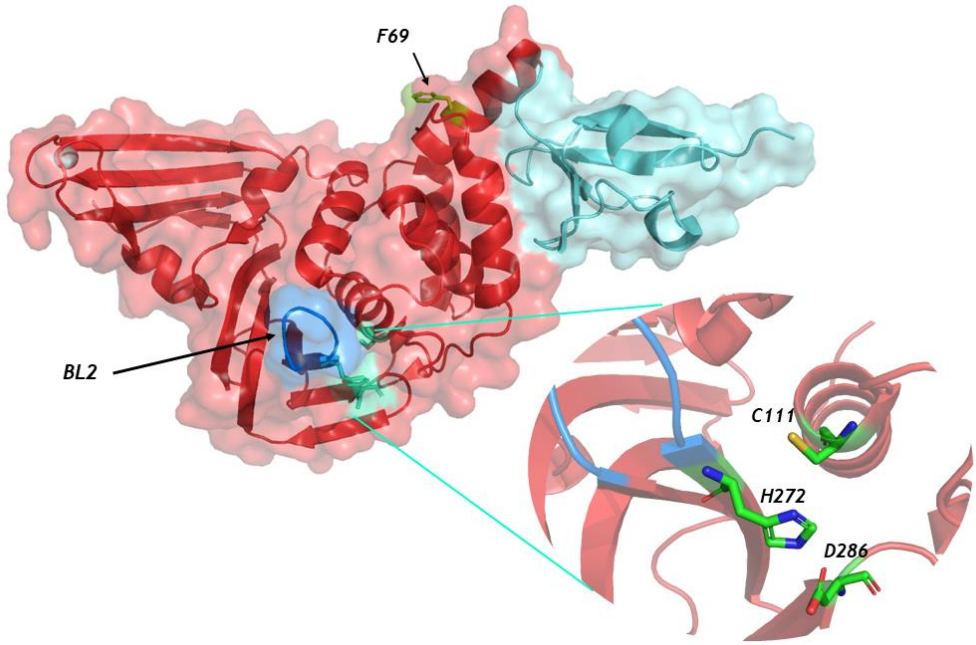


Figure 7: Structure of the PLpro (PDB 6WZU). Ubl2 domain is coloured in cyan; finger-thumb-palm domain is coloured in red. F69 of S2 binding site is coloured in lemon; the BL2 is coloured in light blue; the catalytic triad is coloured in green-cyan and zoomed on the side.

2.3 Drug discovery in response to the pandemic

2.3.1 Early Drug Discovery Process

The early drug discovery process could be described as the first stages of pharmaceutical research in response to the necessity of treating a disease or a clinical condition with no suitable medical products. Such stages comprise the target identification and validation, the primary selection of candidate compounds (screening) and the hit identification. The hit(s) will then be considered to be validated and developed during pre-clinical progression. A different approach of compound screening could be considered if the target of interest has not been discovered or characterized yet, the phenotype-based screening. This method requires a model of disease to be used for compound testing, the precise target of the molecules of interest could be identified in a second moment (in the favourable cases) [33]. However, target-based drug discovery is often preferable, since offers a more rational and less time-consuming approach to the drug discovery process. Concerning Covid-19, various methods were taken into consideration in response to the global emergency, but three were mostly applied: virtual screening, high-throughput screening, fragment screening and fragment-based drug design.

High-throughput (HT) screening

The HT screening is widely used to identify hit series from chemical or biochemical libraries. It is also the most preferred if little is known about the target (e.g. target structure). One of the most commonly used techniques is the biochemical assay. The assay requires the purified target of interest to develop a suitable assay for the target itself and the molecule libraries that will be tested (Supp. figure 1). Fluorescent methods largely replaced radioligand-based, using a fluorescently labelled substrate of the target as a probe. The assay format is thus competitive [34]. Depending on the specific technique used to set up the activity assay, some phenomena could generate artefacts and false positives or negatives more frequently with respect to the radioligand methods, like auto-fluorescence of molecules, quenching, and poor half-lifetime of the probe [35]. Biophysical techniques that directly detect the protein-ligand interaction could also be considered for HT screening. Given that such techniques usually require more material and time than the biochemical assays, libraries to be screened in this way are usually smaller. They are often used though for hit validation after the biochemical screening. Mass Spectrometry, ligand-based and protein-based NMR, and binding techniques (Thermal Shift Assay, Surface Plasmon Resonance, Microscale Thermophoresis,…) are often used as primary or secondary methods for hit discovery [36].

Fragment screening and fragment-based drug design

Fragment-Based Drug Design (FBDD) is a rational approach to designing novel potent inhibitors from "fragments" (small molecules with usually weight < 300 Da). The common suggestion in the choice of fragment library is to follow the "rule of three": molecular weight under 300 Da, ClogP under 3, and less than three hydrogen donors or acceptors [37]. The binding of fragments to the target of interest must be confirmed. Due to the usual low affinity of the fragments, biophysical assays could be applied, especially for primary screening. One elected tool for this aim is NMR, especially ligand-based NMR. STD and WaterLOGSY techniques are often applied for primary and secondary screening of fragments, well-suitable due to the low affinity of the molecules. Labelling the ligands with ^{19}F makes applying ^{19}F-NMR techniques to the ligand identification, possible because there is no ^{19}F in biological molecules and therefore no background in NMR spectra [38,39]. Probably the most exhaustive technique is still crystallography, which furnishes high-resolution information about the binding with detailed structure insights useful for a rational drug design and improvement. Fragments are usually soaked into formed crystals, thus the bottleneck of the procedure is obtaining a number of protein crystals, ideally in conditions suitable for the soaking [40]. The workflow of the FBDD is reported in the Supp. figure 2.

Several fragment screenings were performed in the context of the SARS-CoV2, focusing on the most druggable targets, including nsp14, Mpro, Macrodomain 1, and the helicase (nsp13).

Virtual screening and computational methods

Computer-Aided Drug Discovery (CADD) is an ensemble of *in silico* methodologies to design or select candidate compounds before their synthesis and biological evaluation. The employment of molecular modelling allows the analysis of large libraries of compounds in a short time and the simulation of pharmacokinetic and toxicological profiles, knocking down the time and costs of the classical screening approach [41]. Among CADD methods, Virtual Screening and Structure-based Virtual screening (SBVS) are popular and solid methods to accelerate hit identification. The technique requires a high-resolution structure of the target of interest, better if obtained experimentally but nowadays artificial intelligence-based simulators (eg AlphFold) are able to produce well-predicted structures. The compounds are selected according to their affinity for the receptor [42]. Score functions based on the calculated energy of the interaction are used to value the likelihood of the binding [41]. Even though SBVS is a powerful tool for ligand screening, experimental validation is inevitably required.

2.3.2 PLpro as a target in drug repurposing

Drug repurposing could be described as the research of new therapeutical indications for already approved drugs or to advance non-approved drugs previously studied [43]. The repurposing is claimed to be faster, less risky and less expensive since the chemical and safety properties of the molecules are already profiled. For this reason, in the context of the pandemic, drug repurposing was extensively considered to accelerate the discovery of effective drugs against SARS-CoV2.

Due to its essential role in viral replication as well as the interactions with host proteins and immune response pathways, the PLpro was one of the major interests of drug repurposing campaigns. However, two main challenges occurred: the S1 and S2 sites bind tightly close to glycines of substrates, which are difficult to be mimicked by a small molecule; second, similar substrate binding motifs are used by host proteins, which leads to selectivity problems [17]. In fact, a number of *in silico* and *in vitro* repurposing screenings were conducted on this target, but only a few compounds were validated, proven also by the low number of deposited co-structures with repurposing inhibitors.

Classes of compounds described in the literature include: inhibitors reacting with the Cys111, notably peptidomimetic inhibitors; thiopurine compounds; allosteric inhibitors including natural products, naphthalene inhibitors and piperidine carboxamide compounds [44, 45]. Two peptidomimetics were optimized from SARS-CoV1 PLpro known inhibitors (PDB structures 6WUU, 6WX4). Cys111 engages in Michael Addition to the β−carbon of the vinyl group of VIR250 and VIR251, forming a covalent thioether linkage [46]. The compound Acriflavine was reported to be a potent nanomolar inhibitor of SARS-CoV-2 PLpro in enzymatic, cell-based and *in vivo* studies. Using NMR and a crystal structure of PLpro with proflavine, the active component of acriflavine, Napolitano et al. showed that two proflavine molecules occupy the substrate binding pocket also used for the C-term recognition motif RLRGG in ISG15 [47]. Unfortunately, proflavine is known to have a mutagenic effect since it intercalates DNA. Thiopurine inhibitors (6-Mercaptopurine and 6-Thioguanine) have been reported to inhibit SARS-CoV1 PLpro protease activity with an IC_{50} in low μM range [48]. Both 6-mercaptopurine and 6-thioguanine are currently used in clinics as anti-cancer treatments. The mode of action of these compounds was reported as slow-binding, competitive, reversible, and selective for SARS-CoV2 PLpro [49]. Various natural products were also described as PLpro inhibitors: tanshinones, diarylheptanoids, geranylated flavonoids, chalcones and coumarins, as well as polyphenols. Such compounds present selectivity problems though, like tanshinones which inhibit several human deubiquitinases [50]. YM155, an imidazolium-based inhibitor of antiapoptotic protein survivin, was investigated by

Zhang et al. In the co-crystal structure obtained by soaking with the YM155 powder (PDB 7D7L), they found three binding sites, the first one in the outer portion of the S1 [51]. However, recent publications have raised doubts on the mechanism of action of many proposed inhibitors and initial reports of the inhibitors from multiple classes should be viewed critically. The study by Ma and Wang revealed the limitation of reported inhibitors by profiling them through a pipeline of enzymatic, binding and cellular activity assays. This study showed the tanshinone-family, YM155, SJB2-043, 6-thioguanine, and 6-mercaptopurine could not be confirmed to induce positive shifts in thermal stability upon incubation with PLpro and were not potent inhibitors in a cell-based (flipGFP) assay monitoring intracellular protease activity [52].

2.3.3 GRL-0617 and derived inhibitors

Compounds with naphthalene scaffold represent a solid class of inhibitors of the PLpro, with available structure-activity relationship (SAR) insights. The "father" of this class is the GRL-0617 (5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide) (Figure 8a), known to inhibit the SARS-CoV1 PLpro with good potency (low μM range) and low toxicity in the Vero6-cell model [53]. GRL-0617 was extensively studied towards the SARS-CoV2 PLpro, measuring an IC_{50} around 2 μM and four co-crystal structures were deposited (PDB 7JIR, 7JRN, 7CMD, 7CJM). The compound binds the subsite recognizing P3 and P4 residues, near the active site but not close enough to interact with the catalytic triad (minimum distance from Cys111 in crystal structure is 7.5 Å). Crystal structures clarified the nature of the interaction. Three hydrogen bonds stabilize the binding: between the -OH group of Y268 and the amino group on the benzene ring of GRL0617; the other two have shorter distances than the first one, which are formed by the backbone amino group of Q269 and the carbonyl oxygen of GRL0617, and by the carboxyl group of the D164 and the H11 of the GRL-0617 amide bridge. A T-shaped π-π stacking of the BL2 loop with the naphthalene moiety of the GRL-0617 closes the loop. This conformation at the bound state together with the competitive behaviour of GRL-0617 against the last two Gly of Ub and ISG15 C-terminus, explain the inhibitory mechanism of this molecule (Figure 8b) [31, 54]. A recent NMR study of the binding elucidated a discrepancy between the crystal structures 7JIR and 7JRN. In the first one, the L162 assumes an open conformation (*trans*) with its methyl distant 6.41 Å from the compound; in the second one the L162 methyl is much closer (*gauche+* conformation), 3.63 Å, directly interacting with the GRL methyl substituent on the aromatic amine. ^{1}H-^{13}C HMQC and ^{13}C-edited NOESY-HMQC validated the closer conformation of the L162, proving a fundamental hydrophobic interaction with the molecule (Figure 9) [55]. Concerning the various derivatives of the GRL-0617, only a few showed a sub-micromolar IC_{50} but gave co-crystal structures [31]. Their antiviral activity, including GRL-0617, was demonstrated by cellular infection models, reporting an EC_{50} usually ten times or higher than the IC_{50} measured in biochemical and biophysical assays [56].

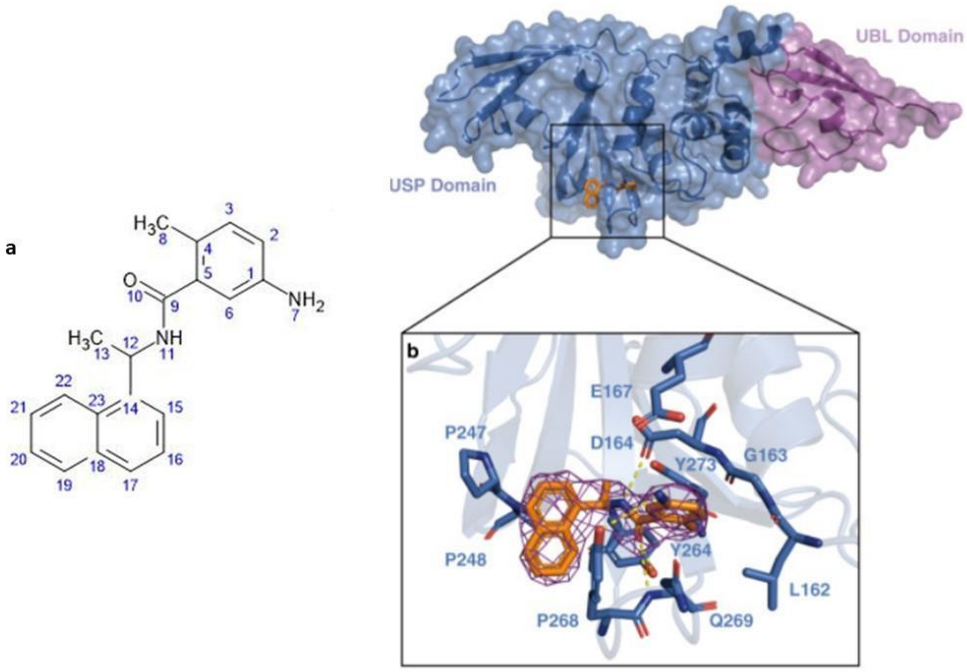


Figure 8: [a] Structure of GRL-0617; [b] GRL-0617 in the binding site and interactions with PLpro residues (from Fu Z et al, 2021) [54].

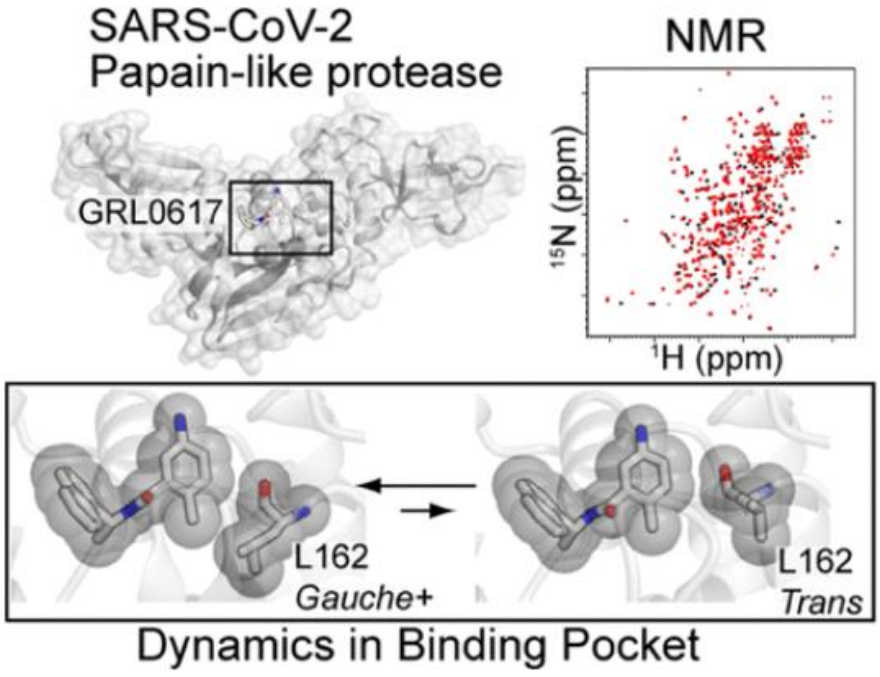


Figure 9: Dynamic study on the L162 of the PLpro at bound state with the GRL-0617 (from Shiraishi Yutaro and Ichio Shimada, 2023) [55].

2.4 The Interferon-Stimulated Gene 15 protein

2.4.1 Role of ISG15 in the human cell

Ubiquitin (Ub) is the most conserved protein in eukaryotics, with a complete identity from Zebrafish to Humans. Ub is conjugated to cellular proteins to regulate a broad spectrum of processes, from protein degradation to cell cycle regulation, thanks to the different combinations of Ub moieties on the modified target [57]. However, Ub is not the only globular small protein to be used as post-translational modification, a larger family of Ub-like proteins share this property. Proteins of this family show similar architecture and the same conjugation process: an ATPase-E1 enzyme generates a thioester intermediate which will react with the E2 conjugating enzyme, and then the E3 ligase will catalyze the transfer of the Ub-like protein to the final target [58]. ISG15 is a member of the Ub-like family, formed by two Ub-like tandem units linked by a highly charged and flexible linker (Asp76-Lys77-Cys78-Asp79-Glu80), which enables multiple conformations. The E1 enzyme UBA7 is unique for ISG15, the bottleneck in the ISGylation, indeed the E2 is the UbcH8 which is able to conjugate Ub as well and more than one enzyme can operate as E3, but the dominant E3 ligase for ISG15 in human is HERC5 [59]. The expression and conjugation of ISG15 are stimulated by interferon-depending pathways, therefore ISG15 plays an important role in pathogen response mechanisms, but plays a role in cancer development as well, due to its regulatory function in autophagy, protein translation and DNA repair. The mechanisms have not been completely understood yet [60]. Even if its essential role is known in infections, a lot about the molecular mechanisms in infected cells must still be discovered. Various studies show that based on the cellular model and the infective agent, the appreciable effect is different [61]. The common result is that ISG15, in one way or another, interferes with cellular processes that are conducive to viral replication.

Many viruses have developed strategies to survive the immune responses, often interfering with the ISG15 system. Coronaviruses, including SARS-CoV2, produce proteases targeting the ISG15 system. More precisely, USP-like enzymes such as the PLpro. In human cells, the putative deISGylating enzyme is the USP18, which recognizes the C-terminal Ubl domain of ISG15 leaving the distal N-terminal domain free for other potential interaction [62]. By contrast, SARS-CoV PLpro can interact with both Ubl subunits of ISG15, making the protease more versatile to the substrates (ISG15, monoUb, poly Ub). Interestingly, there is evidence that ISG15 can also be secreted by cells, acting like a cytokine-like intercellular signal. *In vitro* studies show indeed that INFγ production is enhanced by the presence of free ISG15 in the

extracellular environment [63]. The beneficial balance of free and conjugated ISG15 in infected cells must still be elucidated (Figure 10).

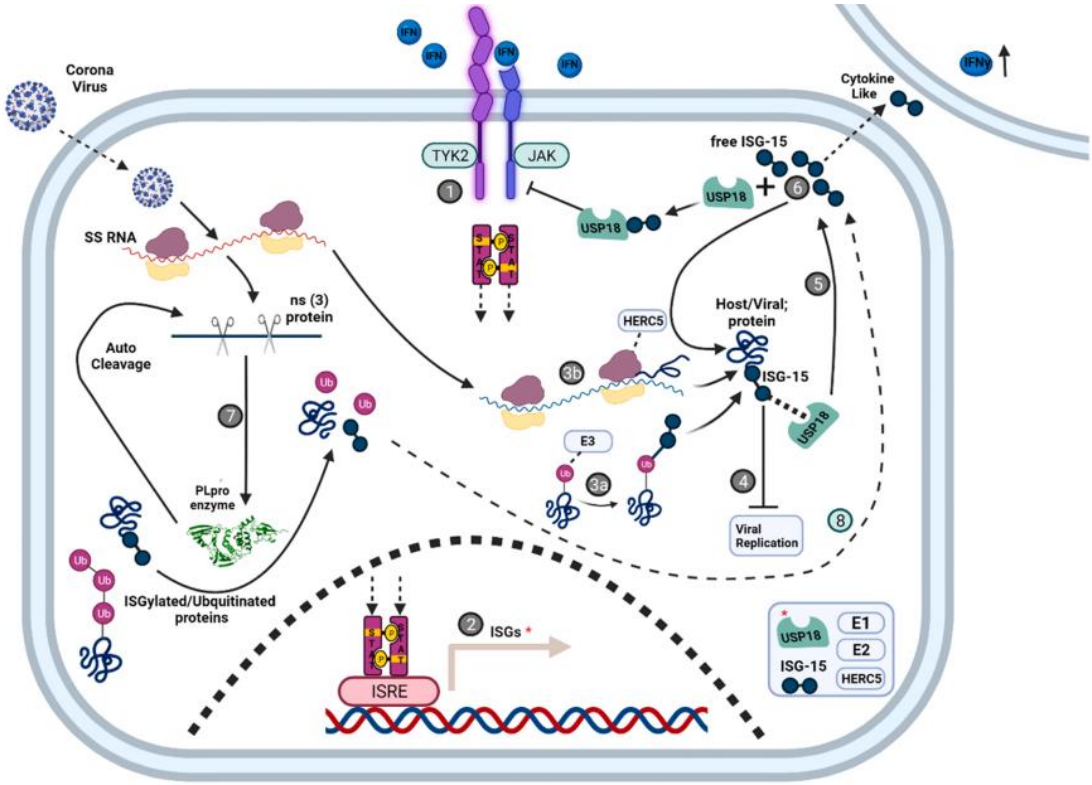


Figure 10: Schematic representation of factors balancing the conjugated and free ISG15 in the cell (from Gold IM et al, 2022) [59].

2.4.2 Structural properties and interaction with the PLpro

ISG15 presents two Ub-like subunits mimicking the head-to-tail fold of di-Ub (Ub2). Sequence and folds are similar to the K48-linked Ub2 thus. Each subunit contains five mixed β-strands intercalated by a single α-helix. The N-terminal domain could be superimposed upon the C-terminal if rotated 45° along y-axis and 60° clockwise along the z-axis, showing interaction between the two domains 3_{10} element at the α-helix/3rd β-strand of the N-terminal and β4/β5 strands of C-terminal domain (Figure 11) [64]. The variability in sequence among different species suggests a variability of tertiary structure as well, with the consequent diverse interaction with viral and host proteins [65].

The interaction of ISG15 with the PLpro was deeply described by Wydorski and Osipiuk. Comparing the crystallographic analysis of the complex PLpro:ISG15 and PLpro:K48-linked Ub2, they appreciated a good conservation of the binding mode at the substrate C-terminus in the S1 pocket, the N-terminal Ub is rotated of 27° towards the PLpro finger instead. The electron density of the substrate at the S2 site is also weaker for the N-terminus of Ub2, suggesting a poorer interaction with the PLpro compared to the ISG15 N-terminus (Figure 12). NMR titration of the ISG15 full-length and the separated two subunits shows a major affinity of the C-terminal moiety to the PLpro than the N-terminal, but the overall chemical shifts in the titration of the N-terminal subunit showed a similar attenuation as for the full-length, suggesting that even if weaker, the interaction of the N-terminal subunit is necessary for the full-length ISG15 binding to the PLpro. In the study, another difference between ISG15 and K48-Ub2 was highlighted. Combining molecular dynamic simulations with cross-linking mass spectrometry, Wydorski and Osipiuk demonstrated different natures of interaction at the S2 level, specifically nonpolar- vs electrostatics-driven distal UBL (F69) and Ub (E70) contacts in the hISG15 and Ub2 complexes, respectively [66].

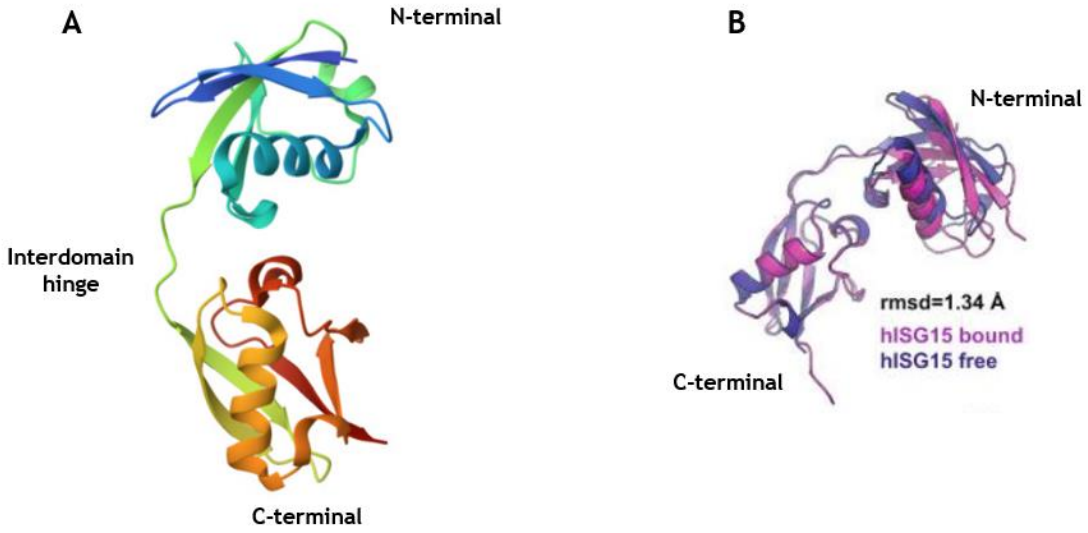


Figure 11: A) Crystal structure of ISG15 (PDB 7S6P); B) Overlap of free and bound crystal structures of ISG15 (from Wydorski et al, 2023) [66].

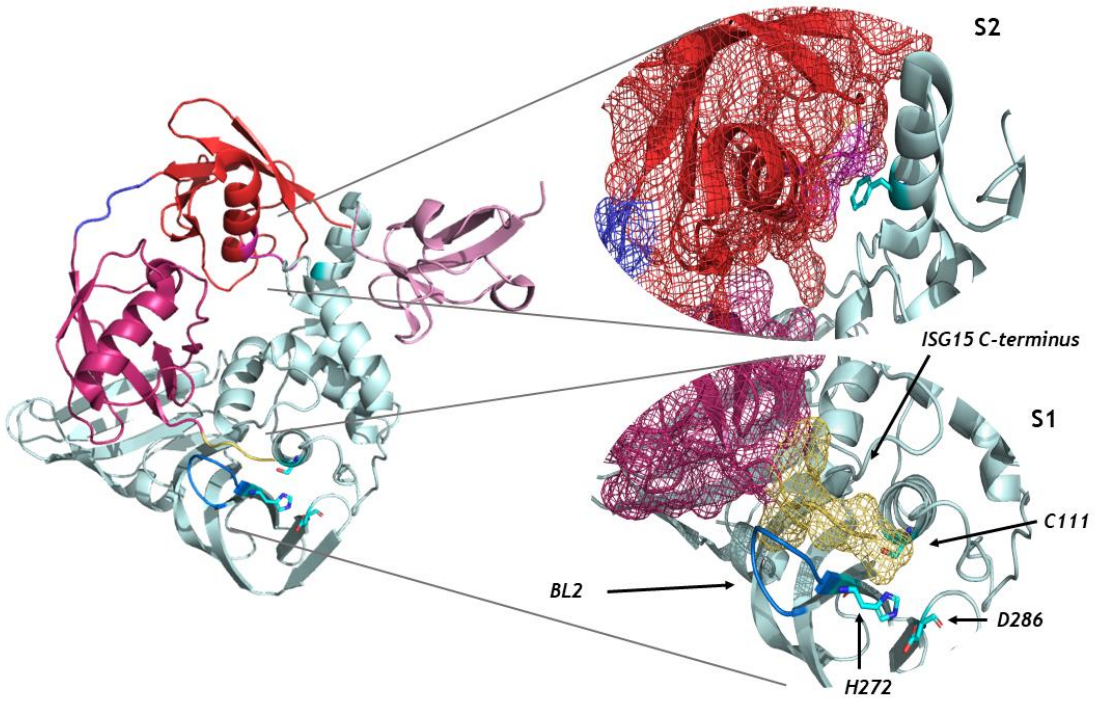


Figure 12: PLpro in complex with ISG15 (PDB 7RBS). Ubl2 of PLpro is coloured pink; finger-thumb-palm domain is coloured light cyan; N-terminal Ub-like subunit of ISG15 is coloured red; C-terminal Ub-like subunit of ISG15 is coloured magenta; LRGG C-terminal tail of ISG15 is coloured yellow; the Bl2 is coloured blue. The two binding sites are zoomed, S2 and S1; catalytic triad next to the S1 and F69 of S2 are highlighted as sticks.

2. Aim of the project

The SARS-CoV2 Papain-like protease (PLpro) is considered a valuable target for drug development. It is well-conserved among Betacoronaviruses and is a key role player in the viral cycle. Moreover, it is attractive from a pathological point of view, protecting the viral replication from the immune response aggression. PLpro has been proving to be a challenging target, both in the discovery of existing molecules via repurposing screening campaign and in the design of new inhibitors [67]. Up to now, only a few molecules have been identified to be effective against this protease and even fewer have been brought to the preclinical level.

With the beginning of the pandemic, researchers and scientists worldwide rallied together, pooling their expertise, knowledge, and resources to unravel the mysteries of the novel coronavirus. This monumental endeavor marked a new era in medical research, where the urgency of the situation spurred groundbreaking innovations and accelerated the development of anti-COVID-19 drugs. Dompé Farmaceutici likewise headed the EU's H2020 funded project called EXaSCale smArt pLatform Against paThogEns for CoronaVirus (Exscalate4CoV or E4C, www.exscalate4cov.eu, Grant 101003551) to provide the Exscalate platform for drug discovery, a virtual platform which combines artificial intelligence and supercomputing to screen molecules virtually and reach faster the lead optimization. In parallel to the in-silico strategy, it was run a more classical discovery workflow that involved state-of-the-art experimental facilities up through clinical validation. The Elettra Protein Facility joined the consortium, working mainly on the two proteases, Mpro and PLpro, to support HT-screening campaigns to identify repurposed drugs and new inhibitors and to provide biophysical and structural tools to characterize the drug-target interactions. For the PLpro, the molecules were identified via HT-screening on a library of almost 9000 repurposing compounds, thanks to an intense collaboration with the Drug-Discovery Platform of Fraunhofer Institute (Hamburg). Due to the various challenges faced, the PLpro branch project became our major objective in this context and the main focus of this PhD thesis.

The project aims to obtain and characterize highly pure and active recombinant proteins suitable for all the assays necessary to identify inhibitors, starting from the repurposing compounds. The hit compounds are then validated using a broad range of biochemical, biophysical and structural tools, such as activity assays, binding assays, spectroscopy, and structural techniques to understand the interaction of the hits with the PLpro. Besides the commonly used PLpro construct, a longer construct including

the NAB domain at the C-terminus of the PLpro was generated, with the aim to highlight differences in the ligands' behaviour in presence of the nsp3 further domain. A second objective is focused on investigating the binding properties of PLpro with its natural substrates, especially the ISG15. The influence of the NAB domain, especially of its mutations found in the Delta-variant, is explored to provide new insights into the interaction with the ISG15.

Even if the PLpro structure is solved already, new structural as well as functional knowledge of the target can lead to more rational and therefore successful drug development.

4. Materials and methods

4.1 Construct design and protein expression

4.1.1 PLpro constructs expression

The PLpro exists as a subdomain of the large multi-domain protein nsp3 (Figure 2). The common construct used for either structural or pharmaceutical purposes reported in literature published so far is the domain of the only PLpro, corresponding to the region 1564-1881 of nsp3. Besides this construct, a second one was designed and produced including the following domain in the nsp3 sequence, the Nucleic Acid Binding domain (NAB), overall corresponding to the 1564-2048 primary sequence of the pp1a and 746-1224 region of the nsp3. The second construct, called PLpro_NAB, aspires to highlight differences between the two constructs regarding the inhibitory activity of tested compounds or interaction with the substrate ISG15 (Figure 13).

Figure 13 Schematic representation of PLpro and PLpro_NAB constructs

The DNA sequence encoding for the PLpro_NAB construct (485 aa) was inserted by NdeI and XhoI cloning in *pET24b* (Novagen, supp. figure 3) expression vector with N-term 6xHis-tag and TEV cleavage site. Protein expression was optimized in *E. coli* strain BL21(DE_3), grown in LB medium supplemented with kanamycin (50 µg/mL) at 37°C and protein was induced at OD_{600}~0.7 for 18 hours at 20°C by adding 0.25 mM IPTG and 50 µM $ZnSO_4$. Cells were therefore harvested by centrifugation at 3000 xg 30 minutes, 4°C; cell pellets were washed with PBS+10% Glycerol, centrifuged again and stored at -80°C until usage.

The inactive mutant of the catalytic cysteine, Cys111 into Ser ($C_{111}S$), was obtained by site-direct mutagenesis using primers 5'-GGACAACAACAGCTATCTGGCGA and 5'-GCCCATTTGATGCTGGTC. The Q5 polymerase (NEB) was used as high-fidelity polymerase to amplify the mutated products during the 35 cycles of PCR. DH5α were transformed and colonies selected for the sequencing check. The expression was performed following the same protocol as the wild-type.

According to the interest in the δ-variant mutations on the NAB domain, $T_{2016}K$ and $S_{2015}R$, the corresponding PLpro_NAB mutants ($T_{T467}K$ and $S_{466}R$ in our construct) were obtained via site-direct mutagenesis using primers 5'-TTGCCTGTGGAGCAAAAAGCCGGTGG and 5'-CGGATGCACCAGGTGGTC for Thr to Lys and

5’-TTGCCTGTGGCGCACCAAGCCGGTGG and 5’- CGGATGCACCAGGTGGTC for Ser to Arg, amplified in PCR programmed following the NEB protocol for the Q5 polymerase. After the sequencing check of the products, the expression was performed following the same protocol as the wild-type.

The PLpro construct was kindly given by Joachimiak A. from Structural Biology Center, Argonne National Laboratory (IL, USA), cloned in the *pMCSG53* vector (supp. figure 4), as well as the $C_{111}S$ mutant cloned in the same vector. The expression of both WT and $C_{111}S$ mutant was tested in small-scale trials performed in 24 deep well plates, in 3 mL of culture medium considering different conditions:

- Medium: LB, TB, autoinduction
- *E. coli* strain: BL21(DE3), Rosetta2(DE3), BL21_pLyS(DE3), BL21(DE3)_GOLD
- IPTG concentration: 0.1 mM, 0.3 mM, 0.5 mM, 1 mM
- $ZnCl_2$ concentration: 10 μM, 50 μM
- Temperature: 20°C, 37°C

Due to a poor yield and solubility of the protein, the expression was further optimized by co-expressing with chaperonin GroE complex that is composed of 2 subunits: GroES (10 kDa) and GroEL (57 kDa). GroE supports protein expression and folding ability, helping proteins to form tertiary structures upon translation. Bl21(DE3) competent cells were transformed with *pGro7* vector (Takara Bio, supp. figure 5), and new stock of chemically competent cells was prepared resulting in strain $BL21(DE_3)$_pGro7.

Small-scale expression was repeated using the new strain with the following conditions:

- Medium: LB, TB
- IPTG concentration: 0.05 mM, 0.1 mM, 0.2 mM, 0.5 mM
- $ZnCl_2$ concentration: 10 μM, 50 μM
- Temperature: 20°C, 37°C
- L-Arabinose concentration: 0.5 mg/mL

The best condition was scaled up and PLpro was expressed in LB medium supplemented with 0.1 mg/mL of ampicillin and 34 μg/mL of chloramphenicol. Chaperonins GroEL and GroES expression was induced at 0.5 mg/mL of L-Arabinose during cell growth, while the PLpro was induced with 0.5 mM 1-thio-ß-D-galactopyranoside (IPTG) and 10 μM $ZnCl_2$ at OD_{600}~0,6 and left growing at 20°C for 18-20 h. Cell culture was harvested by centrifugation at 3000 xg, 4°C for 30’, cell

pellets washed with PBS+10% Glycerol, centrifuged again, and stored at -80°C until usage.

10	20	30	40	50	60
SEVRTIKVFT	TVDNINLHTQ	VVDMSMTYGQ	QFGPTYLDGA	DVTKIKPHNS	HEGKTFYVLP
70	80	90	100	110	120
NDDTLRVEAF	EYYHTTDPSF	LGRYMSALNH	TKKWKYPQVN	GLTSIKWADN	NCYLATALLT
130	140	150	160	170	180
LQQIELKFNP	PALQDAYYRA	RAGEAANFCA	LILAYCNKTV	GELGDVRETM	SYLFQHANLD
190	200	210	220	230	240
SCKRVLNVVC	KTCGQQQTTL	KGVEAVMYMG	TLSYEQFKKG	VQIPCTCGKQ	ATKYLVQQES
250	260	270	280	290	300
PFVMMSAPPA	QYELKHGTFT	CASEYTGNYQ	CGHYKHITSK	ETLYCIDGAL	LTKSSEYKGP
310	320	330	340	350	360
ITDVFYKENS	YTTTIKPVTY	KLDGVVCTEI	DPKLDNYYKK	DNSYFTEQPI	DLVPNQPYPN
370	380	390	400	410	420
ASFDNFKFVC	DNIKFADDLN	QLTGYKKPAS	RELKVTFFPD	LNGDVVAIDY	KHYTPSFKKG
430	440	450	460	470	480
AKLLHKPIVW	HVNNATNKAT	YKPNTWCIRC	LWSTKPVETS	NSFDVLKSED	AQGMDNLACE

DLKPVS

Aminoacidic sequence of PLpro_NAB. PLpro domain is highlighted in red; NAB domain is highlighted in green.

4.1.2 ISG15 and proISG15

Two constructs of the human ISG15 were designed: the mature form of the ISG15, of which the crystal structure and the co-crystal structure in complex with the PLpro are already available in the protein data bank (PDB codes 7S6P, 7RBS), and the precursor form of the ISG15 which includes an 8aa-long tail at the C-term after the PLpro cleavage site. Both constructs have the mutated Cys78 into Ser ($C_{78}S$), which prevents the disulfide-linked dimerization and stabilizes the protein in solution [64].

```
         10         20         30         40         50         60
SMGWDLTVKM LAGNEFQVSL SSSMSVSELK AQITQKIGVH AFQQRLAVHP SGVALQDRVP

         70         80         90        100        110        120
LASQGLGPGS TVLLVVDKSD EPLSILVRNN KGRSSTYEVR LTQTVAHLKQ QVSGLEGVQD

        130        140        150        160
DLFWLTFEGK PLEDQLPLGE YGLKPLSTVF MNLRLRGGGT EPGGRS
```

Sequence of proISG15. The mature form is highlighted in blue; the C-term tail of the precursor is highlighted in yellow; the PLpro cleavage site is underlined.

The *E. coli* codon-optimized genes encoding for the mature and precursor forms of ISG15 were provided by GenScript cloned in the expression vector *pET21d(+)*. An extensive small scale expression trials were conducted in 3 mL of culture medium:

- Medium: LB, TB, autoinduction medium
- *E. coli* strains: BL21(DE3), BL21(DE3)_pLyS, Rosetta2(DE3), BL21(DE3)_GOLD, Lemo21, BL21(DE3)_shuffle, BL21(DE3)_pGro7
- IPTG concentration: 0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM, 1 mM
- Temperature: 18°C, 37°C

After failing to obtain soluble protein, both DNA sequences coding for ISG15 and the proISG15 were amplified and inserted by LIC cloning into the selected vectors (supp. figures 6-8):

- *pNIC_ZB*: N-6xHis-ZB-TEV
- *pNH-Trxt*: N-6xHis-Trxt-TEV
- *pGTvL2*: N-6xHis-GST-TEV

Primers for LIC cloning were ordered from Eurofins Genomics. Primers sequences are reported in the Supp. Table 1.

Expression of the new constructs was tested on small scale considering the selected conditions:

- Medium: LB, TB
- *E. coli* strains: BL21(DE3), Rosetta2(DE3), Lemo21, BL21(DE3)_GOLD
- Temperature: 37°C
- IPTG concentrations: 0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM

Protein expression was thus scaled up and proteins were expressed in BL21(DE3) strain, in TB medium supplemented with 50 μg/mL kanamycin, inducing at OD_{600}~2 with 0.2 mM or 0.1 mM IPTG for mature and precursor forms respectively. After 3.5 h at 37°C, cell cultures were harvested by centrifugation and pellets washed with 1x PBS+10% Glycerol, centrifuged again and stored at -80°C.

As a further optimization, ISG15 was co-expressed with the chaperonin GroE complex to obtain more soluble and folded protein. 6xHis-GST tagged ISG15 was co-expressed in TB medium supplemented with 50 μg/mL kanamycin and 34 μg/mL of chloramphenicol, inducing the GroE complex with 0.5 mg/mL of L-Arabinose when cells reached an OD_{600}~1. ISG15 was induced at OD_{600}~2 with 0.2 mM IPTG, for 3.5 h at 37°C.

Expression vector maps are reported in the Supplementary Material.

4.2 Protein purification

4.2.1 PLpro

PLpro and PLpro $C_{111}S$

Cell pellets were resuspended in lysis buffer (50 mM Hepes pH 8; 500 mM NaCl; 10 mM Imidazole pH 8; 5% Glycerol; 1 µM $ZnCl_2$; 10 mM β-mercaptoethanol; 5 mM MgCl2; 1mM PMSF + 4 µg/ml Leupeptin + 1 µg/ml Pepstatin A + 1 µg/ml Aprotinin, 5 µg/mL DNAse I). Cells were lysed by homogenization for 2 cycles at ca. 1000 bar (PandaPlus 2000, GEA), and the soluble fraction was isolated by centrifuging 60 min at 30 000 xg, 4°C. The supernatants were then incubated 1h at 4°C with 1.5 mL of Ni-NTA resin (Qiagen) per L of culture, pre-equilibrated in binding buffer (50 mM Hepes pH 8; 500 mM NaCl; 10 mM Imidazole pH 8; 5% Glycerol; 1 µM $ZnCl_2$; 10 mM b-mercaptoethanol). Fractions were collected by gravity and protein eluted with same buffer at 300 mM of Imidazole pH 8. 6xHis-tag was removed by TEV cleavage o/n in dialysis against binding buffer. Cleaved proteins were then loaded on Superdex 200 Hiload 26/600 (GE) equilibrated in binding buffer. Fractions containing the PLpro were collected and 2 mL of Ni-NTA resin was added to remove histidine-tag and TEV protease. Flow-through plus 3 volumes of wash containing cleaved PLpro were pooled together and dialyzed o/n against 20 mM Hepes pH 7.5, 150 mM NaCl, 1 µM $ZnCl_2$, 10 mM DTT. The dialyzed sample was finally concentrated up to 25/35 mg/mL (concentration measured by UV spectrophotometry at Nanodrop, Thermo Scientific), flash-frozen in liquid N_2 and stored at -80°C.

PLpro_NAB and PLpro_NAB $C_{111}S$

Cell pellet was resuspended in lysis buffer (20 mM Tris pH8.0, 500 mM NaCL, 10 mM Imidazole, 5 mM MgCl2, 1 mM DTT, protease inhibitors, 10 µg/mL DNAse I), homogenized and soluble fraction separated by centrifugation. The supernatant was loaded on a 5 ml HisTrap FF Crude column (GE Healthcare Life Sciences) equilibrated in binding buffer (20 mM Tris pH8.0, 500 mM NaCl, 10 mM Imidazole, 1 mM DTT) and the PLpro fraction was eluted by 0-100% gradient of the elution buffer (20 mM Tris pH 8.0, 500 mM NaCl, 300 mM Imidazole, 1 mM DTT). The Tobacco Etch Virus (TEV) protease treatment was done o/n in dialysis against 20 mM Tris pH 8.0, 150 mM NaCl, 5% Glycerol, 1mM DTT, 1 mM EDTA at 4°C, followed by a negative IMAC by gravity flow on Ni-NTA resin (Qiagen). In the first purification protocol we set up, the untagged protein was first loaded on a size-exclusion chromatography (SEC) column and followed by anion-exchange (AEX) chromatography. Using this protocol, truncated species were

detected as characterized by MS. A homogenous and integer sample was obtained by inverting the last two steps as follows: TEV-cleaved fractions were purified on 5ml HiTrap Q HP column (GE Healthcare Life Sciences) loading sample in buffer A: 20 mM Bicine pH 9.0, 2 mM DTT. PLpro eluted at 25-30% of buffer B: 20 mM Bicine pH 9.0, 1 M NaCl, 2 mM DTT. The fraction was concentrated and loaded on Superdex 200 Hiload26/600 (GE Healthcare Life Sciences) equilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 2 mM DTT. Eluted fractions were concentrated to 18-25 mg/ml (concentration measured by UV spectrophotometry at Nanodrop, Thermo Scientific), aliquoted and flash-frozen, and stored at -80°C till usage.

PLpro_NAB $S_{466}R$ and $T_{467}K$

The two δ-variant mutants were purified with the same protocol as the other PLpro_NAB constructs, except for the anion exchange chromatography, which was changed for Heparin chromatography to remove nucleic acid contaminants better. Untagged fractions were loaded on a HiTrap Heparin column (Cytiva) equilibrated in AEX buffer A and protein eluted at 20% buffer B. Collected fractions were then loaded on a SEC column as final step of purification.

4.2.2 ISG15 and proISG15

Cell pellets were resuspended in lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 10 mM Imidazole pH 8, 5% Glycerol, 1 mM DTT, protease inhibitors, 5 mM $MgCl_2$, 10 μg/mL DNAse I) and cells were lysed by homogenization for 2 cycles at ca. 1000 bar. The sample was clarified by centrifugation for 1h at 30'000 xg, 4°C and loaded on a HisTrap FF crude 5 mL column (GE Healthcare Life Science), equilibrated in binding buffer (50 mM Tris pH 7.5, 500 mM NaCl, 10 mM Imidazole pH 8, 5% Glycerol, 1 mM DTT). Proteins were eluted with a gradient 0-100% of elution buffer (50 mM Tris pH 7.5, 500 mM NaCl, 300 mM Imidazole pH 8, 5% Glycerol, 1 mM DTT). His-GST tag and His-ZB tag for ISG15 and proISG15 respectively were removed by TEV protease treatment o/n in dialysis against 20 mM Tris pH 7.5, 300 mM NaCl, 5% Glycerol, 5 mM Imidazole pH 8, 1 mM DTT. Cleaved proteins were separated by negative IMAC by gravity and untagged proteins were loaded on a Superdex 75 Hiload 16/600 column (GE Healthcare Life Science). Selected fractions were concentrated to 7-10 mg/mL and 25-30 mg/mL ISG15 and proISG15 respectively (concentration measured by UV spectrophotometry at Nanodrop, Thermo Scientific), flash-frozen and stored at -80°C until usage.

4.3 Protein-ligand analysis

4.3.1 Activity assay and screening of compounds

Activity assay and high-throughput screening were performed by the group of Dr Philip Gribbon at the Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP) in the Drug Discovery – ScreeningPort Department (Hamburg, Germany). Our collaborators performed the activity assay on recombinant proteins produced by our facility.

Assays were optimized regarding enzyme and substrate concentrations, incubation time, and volume and buffer composition. Four labelled substrates were tested: ISG15-AMC, Ubiquitin-AMC, a FRET pair-labelled long peptide fragment representing a sequence of viral polyprotein (FRLKGGAPIKGV-EVANS), and an AMC-labelled short peptide fragment Z-LRGG-AMC sequence reported to be necessary for PLpro recognition. Preliminary building of the assay brought to the choose of ISG15-AMC as principal substrate and the Z-LRGG-AMC peptide as a control-substrate (Figure 13).
In the primary screen, test compounds, positive (20 µM PR619) and negative (100 % DMSO) controls, were transferred to 384-well assay microplates (Corning® Low Volume 384-well Black Flat Bottom Polystyrene NBS Microplate, #3820) by acoustic dispensing (Echo, Labcyte). 5 µl of PLpro_NAB or PLpro mix was added to compound plates. Plates were sealed and incubated for 30 min at 25 °C. After addition of 5 µl ISG15- AMC (R&D Systems #UL-553) substrate, the final concentrations were: 0.15 µM substrate; 1 nM PLpro, 20 µM compound; and 0.2 v/v % DMSO in a total volume of 10 µL/well. The fluorescence signal was measured after 15 min of incubation with the substrate (Ex/Em340/460; Envision, PerkinElmer). Inhibition (%) was calculated relative to controls. Results were normalized to the 100 % inhibition (positive control, PR619 (Merck #662141)) 20µM and 0 % inhibition (negative control, DMSO) inhibition. Assay Buffer: 50 mM Tris, 150 mM NaCl, 1 mM DTT, 0.01 % Tween20, pH 7.5.
During hit confirmation and profiling, 1 mM DTT was exchanged with 1 mM L-Cysteine. For Hit Confirmation (HC), selected compounds from the primary screen were re-picked from the library and re-tested using the same primary assay conditions in triplicate at 20 µM compound. Confirmed hits were then profiled in triplicates in 8-point dose responses, starting from 30 µM top concentration with 1:3 dilution steps.

Library composition

The screened compounds were sourced from three collections. Firstly, the Dompe "Safe-In-Man" (SIM) proprietary collection contains ca 840 drug candidates which have

undergone at least successful Phase I studies. Secondly, the EU-OPENSCREEN collection of 2,464 compounds annotated in line with the Drugs & Probes database, which samples drugs and drug candidates in several development phases together, along with preclinical probes with high affinity for their primary targets. Finally, the Fraunhofer repurposing collection, assembled based on the design features of the Broad Repurposing collection [70]. The Fraunhofer Repurposing Library contains 5,632 compounds including 3,400 compounds that have reached clinical use across 600 indications and 1,582 preclinical compounds with varying degrees of validation. Overall, 8,702 compounds were available for screening. The three compound collections partially overlap in terms of identity. This overlap was useful in determining the consistency in compound response for material sourced from different collections. All test compounds were quality controlled by LC/MS for purity and identity (Purity >90 %) and were stored at -20 °C in 100 % DMSO prior to use.

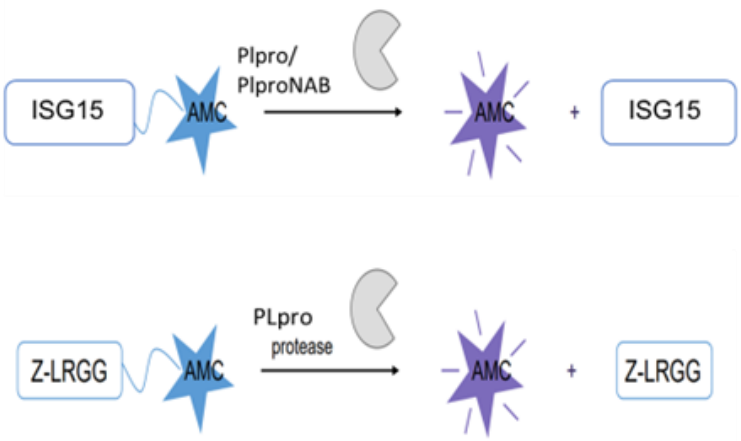


Figure 13: Schematic representation of the activity assay

Other targets

After hit identification, further investigation about selectivity toward other targets was conducted. Concerning the homology with human proteases, USP7, USP14 were chosen together with Cathepsin-L as representative of the Cys-protease family. Hit inhibitors were also tested against the Sars-Cov2 Mpro.

Cathepsin-L cysteine protease activity was measured using the fluorometric cathepsin-L Inhibitor Screening Kit (BPSBioscience #79591).

USP7/USP14 (BPS bioscience, #80364), human deubiquitinases with the highest homology to the PLpro, were tested with a similar assay set-up to the PLpro one (0.5 µM substrate; 100 nM USP7/USP14, in a total volume of 10 µL/well), using the Ub-AMC probe as labelled substrate.

The Mpro activity assay was described in Kuzikov M et al., 2021 [71]. Briefly, the Enzymatic activity of Mpro was measured as described previously. Briefly, the dual-labelled substrate, DABCYL-KTSAVLQ↓SGFRKM-EDANS (Bachem #4045664) containing a protease-specific cleavage site is hydrolyzed by Mpro generating a fluorescent product Incubation with compounds was performed at 37 °C, 25 °C were applied during for incubation with the substrate. Inhibition (%) was calculated relative to controls. Results were normalized to the 100 % inhibition positive (zinc pyrithione (medchemexpress, #HY-B0572) 10 mM in 100 % DMSO) and negative (100 % DMSO) controls.

Sars-CoV PLpro was also considered and used to test hit compounds to check the selectivity against the two isoforms which preserve high homology. Inhibition of SARS-CoV PLpro was measured according to the optimized assay for SARS-CoV-2 PLpro assay.

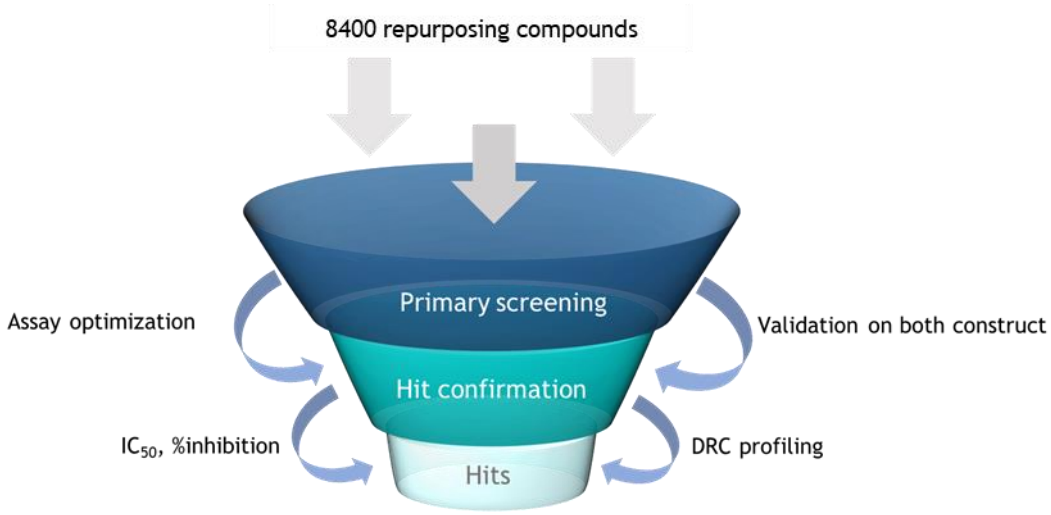


Figure 14: Scheme of the repurposing HT-screening workflow

4.3.2 Thermal shift assay (TSA)

Background

Protein stability *in vitro* is a fundamental parameter in protein biochemistry. Historically, Differential Scanning Calorimetry (DSC) has been the method of choice for characterizing thermal stability in vitro. However, due to its low throughput and expensive instrumentation dedicated to this study, a fluorescence-based thermal stability (or shift) assay has been replacing DSC [72].

The usage of hydrophobic fluorophore can be used to distinguish between folded and unfolded states of proteins. More precisely, in an ideal case, water quenches the dye's fluorescence at low temperatures, observing a basal fluorescence signal. Heating the system, the protein starts to melt exposing hydrophobic patches which could be bound by the fluorescent dye thus giving rise to a fluorescence signal. When finally proteins aggregate due to the denaturation induced by the increasing temperature, the dye dissociates from the protein giving a decrease in fluorescence signal [73]. The high-throughput and small-scale nature of the TSA makes it an excellent platform for screening small ligands, such as organic compounds.

The protein stability is related to Gibbs Free Energy (ΔG_u), which decreases to zero at the equilibrium between folded and unfolded state. The temperature measured at this point is considered the melting temperature (Tm). If a ligand binds to the protein, the free energy contribution of the binding in most cases results in an increase in ΔG_u which may be observed with an increase of the Tm [74]. The resulting ΔTm can give solid and useful information about the binding, considering also that the stabilizing effect of the binding is often proportional to the concentration and affinity of the ligand.

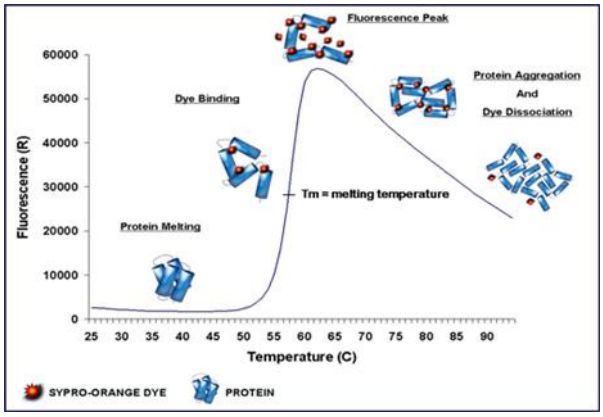


Figure 15: Scheme of the TSA technique explaining the folding transition states (from Wikipedia).

Experimental section

All TSA experiments were conducted in triplicates. The optimal concentration of protein and dye were tested from preliminary TSA for each protein construct tested, in order to establish the most advantageous setup.

Samples were prepared in white 96-multiwell plate (Biorad®) with a final volume of 20 μL in each well. A 5x stock of buffer and a 5x stock of protein solution were prepared hence 4 μL of both were added into the wells. Tested inhibitors were dissolved in 100% DMSO and a 40x stock of each concentration tested was prepared to keep a final concentration of DMSO at 2.5% in each condition tested.

Thermal shift assay on PLpro, PLpro_NAB, ISG15, and pro_ISG15 was performed in assay buffer containing 20 mM Tris pH 7.5, 150 mM NaCl as final concentrations, adding a final concentration of reducing agent (1 mM DTT or 1 mM L-Cys for PLpro and PLpro_NAB, 1 mM DTT for ISG and proISG15) or without any reductant, while PLpro constructs were kept at 5 or 7 μM, ISG15 and proISG15 at 5 mM. Compounds were tested at final concentrations of 0.5x, 1x, 3x, 5x, 10x, 15x, 20x 30x molar excess referring to the protein concentration. Once added buffer, protein and compounds stocks plus a proper volume of water to reach 20 μL, the multi-well plate was centrifugated at 100 xg, 4°C for 1 min to spin down and stir the components. Protein and inhibitors were then incubated at room temperature for 30 min. After incubation, SyproOrange dye (Protein Thermal shift dye, Thermo Fisher Scientific®) was added into each well to a final concentration of 0.5x or 0.7x from 1000x stock in 100% DMSO. The multiwell plate was centrifuged again and measurement was started. Measures were performed in a real-time PCR machine (CFX96, Biorad®), registering emission of the dye at 560-580 nm every 30 s, with a temperature gradient of 2°C/min. Each analysis was executed in comparison to a negative control represented by the only buffer or the compounds with SyproOrange, and a positive control consisted of the protein or protein with 2.5% of DMSO plus the fluorophore. Data were analyzed with the CFX software and elaborated with GraphPad Prism.

4.3.2 Limited proteolysis (LiP)

Limited proteolysis is a widely used technique in proteomics. Usually coupled with mass spectrometry experiments, LiP is a valuable method for protein identification, folding analysis, and protein-ligand interaction evaluation from cell extracts or solution samples [75]. During this PhD workflow, the LiP was applied as an ancillary protein folding study and a quick and easy method to evaluate the interaction between the PLpro and tested inhibitors.

4 μg of PLpro or PLpro $C_{111}S$ were incubated with α-chymotrypsin or trypsin in ratios 1:500 and 1:1000 respectively at 22°C on a Thermomixer, in the activity assay buffer. Different incubation times were considered, hence 5', 10', 30', 1h. After the chosen time, the sample was boiled for 5' at 95°C after the addition of SDS-loading buffer 4x in order to denature both PLpro and the protease, thus blocking the proteolysis. Each sample was then loaded on an SDS-PAGE with the non-incubated sample composed by the PLpro and the protease as control. The experiment was repeated on PLpro samples pre-incubated with selected compounds in a ratio 1:5 protein:compound. The resulting electrophoretic gels were compared to evaluate the stabilization effect of the compound on the PLpro.

4.3.4 Surface Transfer Difference – Nuclear Magnetic Resonance (STD NMR)

Background

Ligand-based NMR techniques represent a solid and valuable tool in the rational drug discovery process. Even if less informative than target-based methods, ligand-based NMR does not require isotope-labelled protein and much lower amounts of target to be performed, features that make these techniques very attractive [76].

Among the ligand-based techniques, Saturation or Surface Transfer Difference NMR (STD NMR) has become a popular method in ligand screening due to its robustness and structural knowledge it can provide. The STD-NMR spectroscopy is based on the intermolecular transfer of magnetization from a macromolecular target to a ligand at fast equilibrium between bound and unbound state through the Nuclear Overhauser Effect (NOE) [77]. This technique is generally suitable for ligands with a dissociation constant between mM and μM range, which makes it particularly appreciable in fragment screening. The STD experiment is performed by saturating a resonance that belongs to target protons, usually in the range from 0 to -1 ppm to avoid a direct saturation on the ligand. The saturation is propagated among target protons via spin diffusion and to the closest hydrogens of the ligand (≤ 5 Å) by cross-relaxation at the protein-ligand interface. The resulting spectrum ("*on-resonance*") will be then subtracted from the control spectrum ("*off-resonance*") obtained saturating in a region far from protein or ligand signals. The difference spectrum yields only the resonances of ligand protons which experienced the transferred saturation (Figure 16). Since ligand protons that are in closest contact with the target receive the most saturation, it is thus possible to map the "binding epitope" of the ligand by calculating the so-called "amplification factors" (A_{STD}) for each proton signal, represented by the ratio between the intensity in the difference spectrum and the intensity in the on-resonance one, multiplied by the ligand excess [78].

Another method based on NOE is the transferred NOE (trNOE), in which it is possible to study the nuclear Overhauser effect between adjacent spins. Small molecules experience small, positive cross-relaxation rates, instead of macromolecules that have larger magnitudes and cross-relaxation rates are negative. In presence of the target, the small ligand experiences negative NOE from the macromolecule to the unbound population. Therefore, in a 2D-NOESY spectrum is possible to appreciate positive cross-peaks of bounded ligands, while negative cross-peaks of the unbound population will be negative. From this experiment is possible to deduce structural information of the

small molecule in the binding pocket, especially the 3D conformation at the bounded state [79].

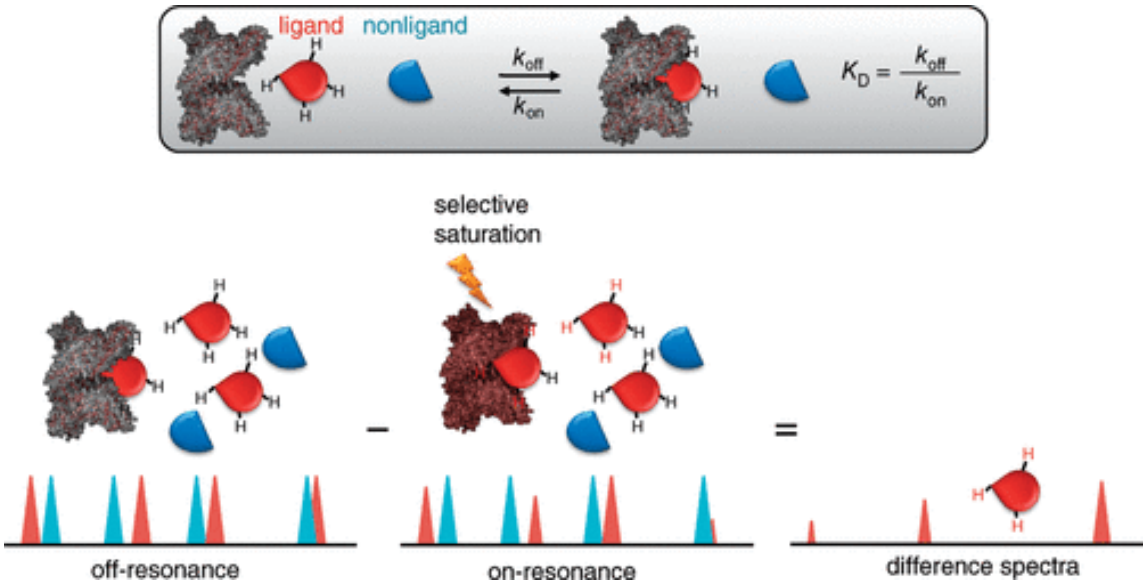


Figure 16: Schematic representation of the STD NMR experiment [76].

Experimental section

The 1D ^{1}H STD and 2D tr-NOESY experiments were recorded on a Bruker Avance Neo 600 MHz spectrometer with a cryoprobe at the Slovenian NMR Centre in the National Institute of Chemistry, Ljubljana (SLO). Spectra were recorded at 298 K using the pulse sequences included in the Bruker TopSpin library of pulse programs. PLpro_NAB buffer was exchanged against 20 mM phosphate buffer pH 8, 50 mM NaCl with 10% deuterated water; the tested compounds were dissolved in DMSO-d_6. Due to the different solubility of the compounds, STD experiments were performed with different concentrations while the protein:compound ratio was always 1:100. Accordingly, PR-619 was tested at 0.15 mM in 5% DMSO-d_6; Semapimod was tested at 0.2 mM in 5% DMSO-d_6; CPI-169 was tested at 0.5 mM in 6% DMSO-d_6; GRL-0617 at 0.3 mM in 6% DMSO-d_6. Proton chemical shifts were assigned according to the standard procedure using 2D-HSQC, 2D-TOCSY, 2D-HMBC and 2D-NOESY NMR experiments recorded without protein in the same buffer as STD and trNOESY experiments. The STD experiments for confirmation of binding of

compounds were performed with 32768 data points, a relaxation delay of 3 s, and 800 scans. The ^{1}H spectral width was 5883 Hz. The on-resonance selective saturation of the PLpro was applied for 2 s at −0.772 ppm with a transmitter offset referenced to 4.699 ppm. The off-resonance irradiation was applied at 30 ppm for the reference spectrum. The residual water signal was suppressed by excitation-sculpting with 2 ms selective pulse and a $T_1\rho$ filter of 100 ms was used to eliminate background protein resonances. Spectra were zero-filled twice and apodized with an exponential line-broadening function of 3 Hz.

The STD ligand epitope mapping experiment of CPI-169 [80] was performed with a larger number of data points and scans while the protein saturation time was shorter to improve quantitative comparison of saturation transfer within the molecule. 65.536 data points, 3520 scans, and a selective on-resonance saturation of PLpro of 1 s were used.

Quantification of the STD effect was obtained by calculating the STD amplification factors (A_{STD}):

$$A_{STD} = (I_0 - I_{STD})/I_0 * \text{inhibitor excess}$$

The ligand-binding epitope was represented with relative A_{STD} values normalized to the highest A_{STD} value, which was classified as 100%. The competitive STD experiment was performed with a lower, 0.15 mM high concentration of CPI -169 because the solubility of a mixture of compounds was limited. Selective protein saturation was prolonged to 2 s to achieve a higher signal-to-noise ratio of STD signals at lower concentration. First, the 1D ^{1}H STD spectrum was recorded at a PLpro_NAB:CPI-169 ratio of 100:1. To the same sample, GRL-0167 was added at a GRL-0617:CPI-169 ratio of 2:1, keeping the final DMSO-d_6 concentration always at 5.5%, and the second 1D ^{1}H STD experiment was performed. The A_{STD} of the methyl protons, which had a sufficient signal-to-noise ratio of the STD signals, were calculated and compared.

The trNOESY [81] spectra were acquired with a spectral width of 5882 Hz, 4096 data points in t_2, 64 scans, 128-182 complex points in t_1, a mixing time of 250 ms, and a relaxation delay of 1.5 s. Spectra were zero-filled twice and apodized with a squared sine bell function shifted by $\pi/2$ in both dimensions.

4.4 Protein-protein interaction characterization

4.4.1 HPLC-light scattering (OMNISEC)

The absolute molecular weight (MW) and oligomerization status in solution of recombinant proteins were measured using an OMNISEC (Malvern Panalytical) instrument. Such instrumentation includes an HPLC where an analytical size-exclusion Zenix150 column (Sepax) was mounted and connected to a triple detector: UV, refractive index (RI), right angle static light scattering (RALS), and low angle (7°) static light scattering.

OMNISEC was employed as a quality characterization system for recombinant products and to preliminarily characterize the PLpro or PLpro_NAB – ISG15 complex. Proteins were thawed and diluted to 1 or 2 mg/mL, centrifuged at 10'000 xg for 10' at 4°C, and transferred to a 96 multi-well plate kept at 8°C for auto-sampling. 50 μL of sample was injected for each run with a 1 mL/min flow rate. Different buffers were tested to minimize the unspecific interactions with resin: PBS 1x; PBS 1x, 1 mM DTT; Storage buffer; activity assay buffer; 20 mM Tris pH 7.5, 200 or 250 mM NaCl, 1 mM DTT; 50 mM Tris pH 7.5, 200 or 250 mM NaCl, 1 mM DTT. The sharpest peaks in the resulting chromatograms were observed in the last buffer, that was selected for the characterization.

PLpro or PLpro_NAB and ISG15 or proISG15 were diluted at 1 mg/mL for complex characterization to obtain a 1:1 ratio; 1:3; 1:5 and 1:7 ratios were also tested. The samples were incubated at RT 30', centrifuged and then injected. Data were analyzed using the OMNISEC software, in particular the measured concentration was calculated from the RI signal and the MW from the RALS.

Light scattering $= K_{LS} * MW * \left(\frac{dn}{dc}\right)^2 * conc$

RI $= K_{RI} * \frac{dn}{dc} * conc$

(dn/dc is the refractive index increment; K_{LS} is the constant for LALS)

4.4.2 PLpro-ISG15 pull-down

Protein pull-down is an *in vitro* technique used to identify interaction between one or more proteins of interest with potential partners, widely common in proteomics procedures. Despite the immunoprecipitation, which requires immobilised antibodies to capture the protein complex, the pull-down method uses a purified and tagged protein as bait for the partners. The same approach could be applied to purified proteins to determine a qualitative interaction and co-purify the complex.

His-tagged PLpro_NAB WT, PLpro WT, and PLpro $C_{111}S$ were purified following the established protocol, avoiding the TEV treatment. The pull-down protocol was adapted from the one reported by Louche A, Salcedo S and Bigot S [82]. 75 μg of His-tagged protein was mixed in 500 μL of binding buffer (50 mM Tris pH 7.5, 150 mM NaCl, 5 mM Imidazole, 2 mM DTT) with 75 μg of ISG15 or proISG15 and left for 2h at 4°C. 150 μL NiNTA resin (Qiagen) was added to each sample and left 30' at 4°C. The flow-through, a wash in binding buffer and 3 CV of elution (binding buffer at 300 mM Imidazole) were collected to be loaded on SDS-PAGE.

4.4.3 Grating-coupled interferometry (Creoptix)

Background

Grating coupled interferometry is one of the most sensitive biosensors commercially available. It is an SPR-like technique, based on a phase-shifting interferometer with an increased sensitivity with respect to SPR. The refractive index changes as time-dependent phase-shift signals, providing a more robust readout than classical Waveguide Interferometry or SPR (Supp. Figure 9).

Experimental section

6xHis tagged proteins were produced following the main purification protocol for each construct, avoiding the tag cleavage and the next negative affinity step. Grating-coupled interferometry (GCI) experiments were performed by the Creoptix WAVE system, in collaboration with Dr Sonia Covaceuszach (IC-CNR, Trieste). Borate buffer (100 mM sodium borate pH 9.0, 1 M NaCl) was used for chip conditioning.

Experiments with the WT proteins were performed reversibly capturing His-tagged PLpro/PLpro_NAB (10 μg/ml in running buffer, i.e., PBS, 0.05% Tween 20, 1 mM DTT) on 4PCP-NTA WAVE chips according to the manufacturer instructions at a density of 3000 pg/mm2. 6xHis-GST ISG15 and 6xHis-ZB proISG15 were used as well for capturing on the chip. In this case, capture and amine coupling succeeded, instead of the binding experiment which did not give any interaction signal.

In the case of the variants, His-tagged PLpro_NAB mutants (10 μg/ml) were captured and covalently immobilized by amine coupling on 4PCP-NTA WAVE chips according to the manufacturer's instructions.

Regeneration-free injections of a 1:2 dilution series of ISG15 and proISG15 in running buffer were performed at 25 °C, using a flow rate of 30 μl/min (20 s association, 45 s dissociation and 600s dissociation for the last injection).

Blank injections were used for double referencing and a dimethyl sulfoxide (DMSO) calibration curve for bulk correction. Analysis and correction of the obtained data were performed using the Creoptix WAVE control software (correction applied: X and Y offset; DMSO calibration; double referencing).

4.4.4 Cleaving activity test of PLpro_NAB on proISG15

To obtain a qualitative activity comparison of the δ-variant mutants and the WT, a cleaving test on the proISG15 was performed. A mother solution of PLpro_NAB and proISG15 in ratio 1:1 was prepared diluting proteins to 1 mg/mL in the activity assay buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM DTT). The reaction was conducted in ice bath to slow down the cleaving performance. 5 μL of mother solution were collected at times 0, 15'', 30'', 1', 3', 5' and the reaction was stopped adding SDS loading buffer, then boiling the sample at 95°C for 5'. The procedure was repeated for WT, $S_{466}R$ and $T_{467}K$ δ-variant mutants, $C_{111}S$ inactive mutant. Samples were loaded on a 16% polyacrylamide gel.

4.5 Crystallization and data collection

PLpro apo form

PLpro construct was crystallized following the published conditions by Osipiuk et al. [31]. Crystals of the apo form $C_{111}S$ mutant were obtained in hanging drop, with a protein concentration of 20 mg/mL. Crystals grew in 0.1 M Acetate buffer, 0.8 M NaH_2PO_4 / 1.2 M K_2HPO_4, pH 4.5, in asymmetric drops 2:1 protein/reservoir at 4°C after 3 days. The $C_{111}S$ crystals were used to produce seeds for cross-seeding procedure in order to obtain crystals of the WT at 15 mg/mL in the same conditions, using a seed dilution of 1:1000 from the mother. Other crystallization conditions were explored, testing several concentrations of protein, protein-reservoir ratios, temperature, and precipitant conditions using commercially available screens. WT and $C_{111}S$ mutant crystals were observed in sitting drop in conditions 35 and 36 of the SaltRX screen (Hampton), containing 0.1 M Bis-Tris propane pH 7 and 1.4 and 2.4 M sodium malonate pH 7, respectively. The shape of the crystals looked different from the previous condition, therefore was further optimized in handing drop using sodium malonate screen (Hampton). Larger crystals grew in the sodium malonate buffer at pH 6 and 7, at a buffer concentration between 1.5 and 2 M, 4°C. Crystals were flash-frozen using a cryo-protectant solution containing the mother-liquor additionated with 25% of glycerol.

In order to have crystals with a less tight packing of molecules in the cell unit, the condition reported by Zhao et al [51]. was reproduced. Only crystals of the $C_{111}S$ mutant grew at different protein concentrations (5, 10, and 15 mg/mL) in sitting drop against 80 µL of reservoir containing 0.1M Tris pH 8.0, 1.6/1.8 M Ammonium sulfate, pre-incubating proteins with 8% of No. 26 of Silver Bullets™ (Hampton Research) additive, containing 0.033% w/v Caffeine, 0.033% w/v Dithioerythritol, 0.033% w/v L-Methionine at 20°C. Crystals were frozen in a cryo-protectant solution of mother-liquor at 20% ethylene glycol.

X-ray diffraction was tested on XRD2 beamline in Elettra Sincrotrone Trieste and structure solved by molecular replacement with Phaser (CCP4 suite), using structure 6WRH as model (Supp. Table 2).

A large crystallization screening was also performed on the PLpro_NAB apo form, using commercially available crystal screens from Hampton and Molecular dimensions. Despite the number of conditions tested, no crystal grew in any of those.

Co-crystallization and soaking

In order to obtain co-crystals of the PLpro or PLpro_NAB in complex with candidate inhibitors, a large screening of commercially available conditions was tested, as well as the conditions optimized for the apo form. Only crystals of the small molecules tested were obtained.

Soaking trials were performed using the apo crystals grown in the condition reported by Zhao et al. or the crystals grown in the sodium malonate condition. Unfortunately, due to the low water solubility of the compounds and the high concentration of salts in the reservoir, all trials failed.

4.6 Small angle X-ray scattering (SAXS)

Batch mode

SAXS data for PLpro WT, $S_{466}R$ and $T_{467}K$ mutants were collected on the Austrian SAXS beamline in Elettra Sincrotrone Trieste with the help of Prof. Heinz Amenitsch, as 10 x 1 s exposures time using a 2D Pilatus3 1M detector and wavelength of 0.99 Å. Scattering profiles for the collected frames were compared to detect radiation damage. Measurements were carried out at five different concentrations (the ranges are reported in Table 18) in SEC buffer (20 mM Tris pH 8, 150 mM NaCl, 2 mM DTT), freshly changed in centrifugal concentrator (Thermo Fisher Scientific, MWCO 30 kDa) and then centrifuged at high speed (10'000 xg, 10', 4°C) to remove protein aggregates.

Collected scattering data were analysed by Dr Sonia Covaceuszach (IC-CNR, Trieste). In detail, frames were merged for each sample after normalization to the intensity of the transmitted beam. The buffer's contribution to the scattering and further processing steps were subtracted using PRIMUS [83] from the ATSAS 2.6.0 program package [84]. The forward scattering I(0) and the radius of gyration Rg were evaluated using the Guinier approximation [85], assuming that at very small angles (s < 1.3/Rg) the intensity is represented as:

$$I(s) = I(0) * e^{\frac{-(s*Rg)^2}{3}}$$

Pair distance distribution functions of the particles p(r) and the maximum sizes Dmax were computed using GNOM [86]. MM were estimated by comparison of the calculated forward scattering I(0) of the samples with that of the standard solution of bovine serum albumin (MM 66 kDa). Vp was calculated using the Porod approximation [87]:

$$V_p = \frac{2\pi^2 I(0)}{\int I_{exp}(s) s^2 ds}$$

The program DAMMIF [88] was employed to construct low-resolution ab initio beads models of PLpro WT, S466R and T567K that best fit the scattering data. It employs a simulated annealing procedure to build a compact beads configuration inside a sphere with the diameter Dmax that fits the experimental data Iexp(s) to minimize the discrepancy:

$$\chi^2 = \frac{1}{N-1} \sum_j \left[\frac{I_{exp}(s_j) - c I_{calc}(s_j)}{\sigma(s_j)} \right]^2$$

Twenty independent DAMMIF runs were performed for each scattering profile, using default parameters and no symmetry assumptions (P1 symmetry). The models resulting

from independent runs were superimposed using the program SUPCOMB [89] and aligned models were averaged using DAMAVER [90] to identify the most typical models representing the two protein variants in solution whose similarity was estimated by the normalized spatial discrepancy parameter (NSD) [89] obtained from DAMAVER. NSD values ≤1.0 are expected for similar models. The resolution of the obtained ab initio models was estimated by SASRES [91].

A simulated annealing protocol implemented in CORAL [92] was employed to find the optimal positions and orientations of the available high-resolution model of the PLpro (PDB 6XAA) and the model of NAB domain (PDB 7LGO) of PLpro WT, $S_{466}R$ and $T_{467}K$ mutants. In addition, the program *SWISS-MODEL* [93] generated approximate clash-free conformations of the missing portions of polypeptide chain (21 amino acids at the C-terminal and the missing linker between the catalytic and the NAB domains).

The inter-domain flexibility and size distribution of possible conformers, consistent with the measured scattering data for PLpro WT, ST and KR mutants, was analysed using the ensemble optimization method (EOM) [94]. This method selects an ensemble of possible conformations from a pool of 10,000 randomly generated models constructed from rigid domains linked by randomly generated flexible linkers. The program CRYSOL [95] calculates the theoretical scattering profiles of these models, and a genetic algorithm, GAJOE, is used to select an ensemble of conformations, whose combined scattering profiles best fit the experimental data. The crystal structures of the PLpro and the NAB domain were used as rigid bodies for the analysis of the scattering data, employing ensemble optimization.

SEC-SAXS

SEC-SAXS analysis was recently performed on the BioSAXS beamline in ESRF (BM-29). Samples were centrifuged at high speed before the injection in the HLPC (Shimadzu). 10 μL of each sample were injected into a Bio-SEC 300 column (Agilent), provided by the beamline facility, equilibrated in running buffer (20 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT). In order to obtain scattering signal of the PLpro or PLpro_NAB in complex with the ISG15, a mixture of the two proteins was loaded in the HPLC system, testing two different protein ratios: 1:5 (2 mg/mL PLpro or PLpro_NAB, 10 mg/mL ISG15) and 1:3 (3 mg/mL PLpro or PLpro_NAB, 9 mg/mL ISG15).

Data analysis is in progress.

5. Results and discussion

5.1 PLpro

5.1.1 PLpro constructs production

PLpro and PLpro $C_{111}S$

The expression and purification protocol used for the PLpro construct was optimized basing on the published information from Osipiuk et al. [31]. Despite the many conditions of expression tested, the results were not satisfactory for the poor solubility of the product. PLpro was indeed well expressed in the tested conditions, but only a small amount was extracted in the soluble fraction and almost lost in the first steps of purification (Figure 17 left). To assist the protein folding and consequent solubility, the co-expression of the PLpro with GroE complex led to an increment in the total yield of protein produced and an enhanced solubility, well observable in the preliminary tests on small scale (Figure 17 right).

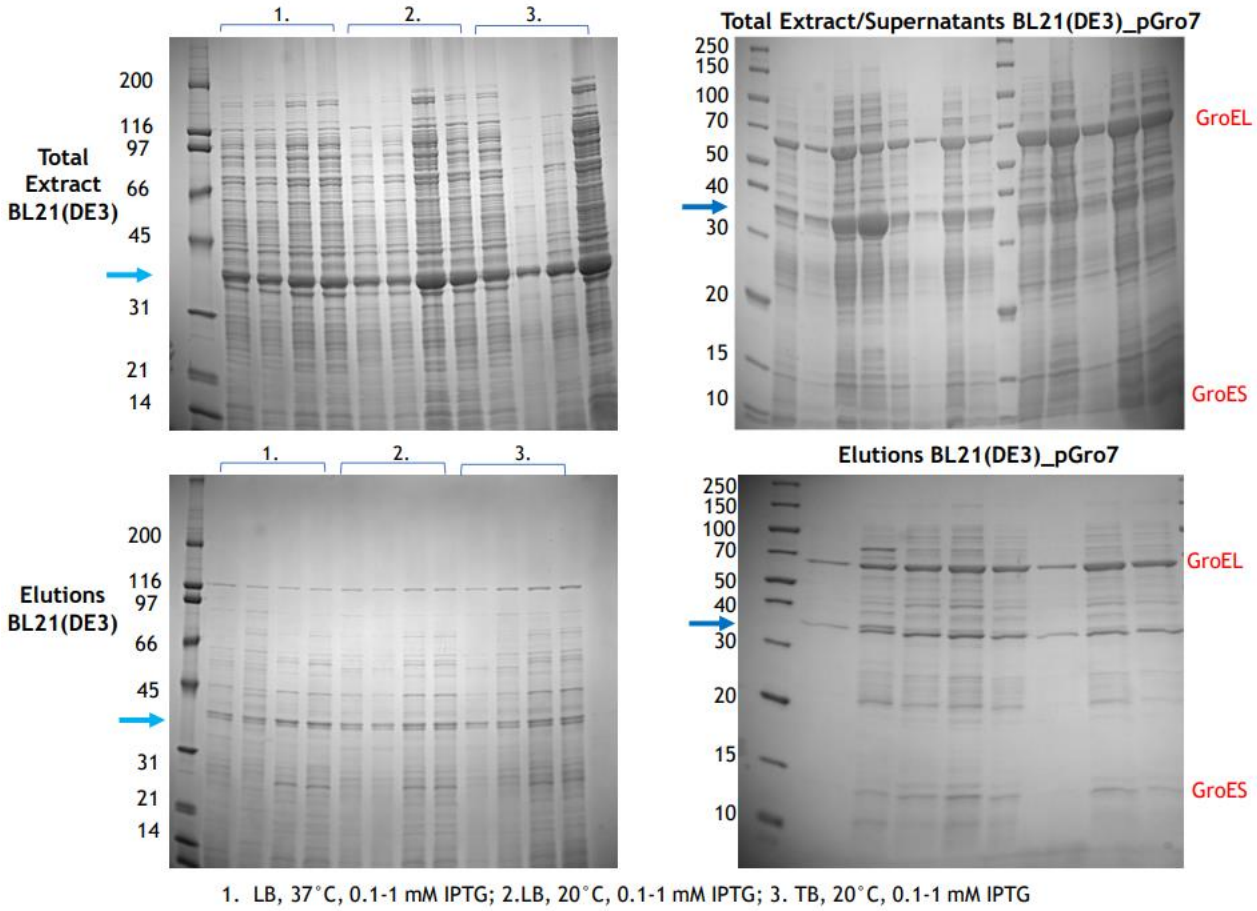


Figure 17: Small scale expression of PLpro $C_{111}S$. Left panel: SDS-PAGE of total extracts (top) and eluted (bottom) fractions of test expression in BL21(DE3); Right panels: SDS-PAGE of total extracts (top) and eluted (bottom) fractions of test expression in BL21(DE3)_pGro7 on the right. The blue pointers show the band corresponding to the PLpro.

The scale up expression was thus performed using BL21(DE3)_pGro7 strain in LB medium, inducing the chaperonin complex with 0.5 mg/mL of L-Arabinose.

The purification followed the protocol published in the paper [31]. The overexpressed chaperonins, in particular GroEL, were washed out during the first affinity, adding a second wash at 20 mM Imidazole before the elution. The remaining chaperonins as well as other contaminants were finely eliminated during the size-exclusion chromatography step, leading to a highly pure final product.

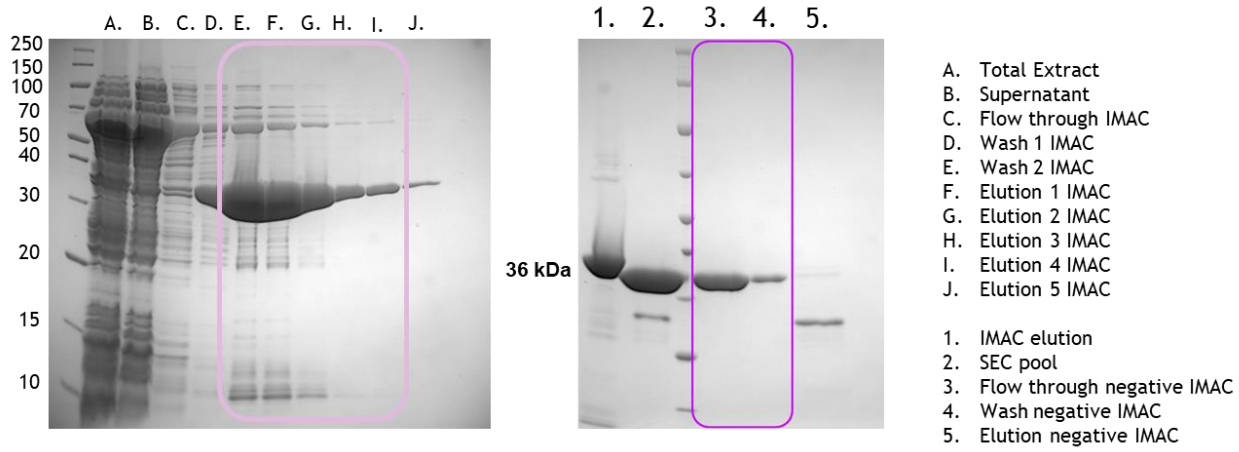


Figure 18: SDS-PAGE (16%) of PLpro first step of purification (left); SDS-PAGE (16%) of PLpro final step of purification.

PLpro_NAB WT and mutants

The PLpro_NAB wild type (WT) was expressed in BL21(DE3) in a good yield. After the first IMAC step of purification and the consequent His-tag removal via TEV cleavage, the sample was loaded on a size-exclusion column during the first trials of purification. A further purification step by anion exchange chromatography was added to remove protein impurities and nucleic acid contaminants revealed by UV-spectrometry during the quantification. The first resulting batch of protein was then characterized by Mass Spectrometry (MS) at the CEINGE institute in Naples, by the group of Prof. Maria Monti. Besides the major component corresponding to the full-length PLpro_NAB, two other species were detected and identified by MALDI-MS and LC-MS/MS: a larger component corresponding to the fragment 1-361 (40.91 kDa) and fragment of 15.12 kDa. In order to obtain a more homogeneous sample to be used in our assays and studies, the two final steps of purification were reversed and the major contaminant which corresponds to the fragment 1-361 was separated from the full-length (Figure 19). Notably, the PLpro_NAB protein elutes in SEC at a lower retention volume than expected. Based on

column calibration, it elutes as a 70/75 kDa protein. In the next chapter, more analytical deepening will explain this phenomenon.

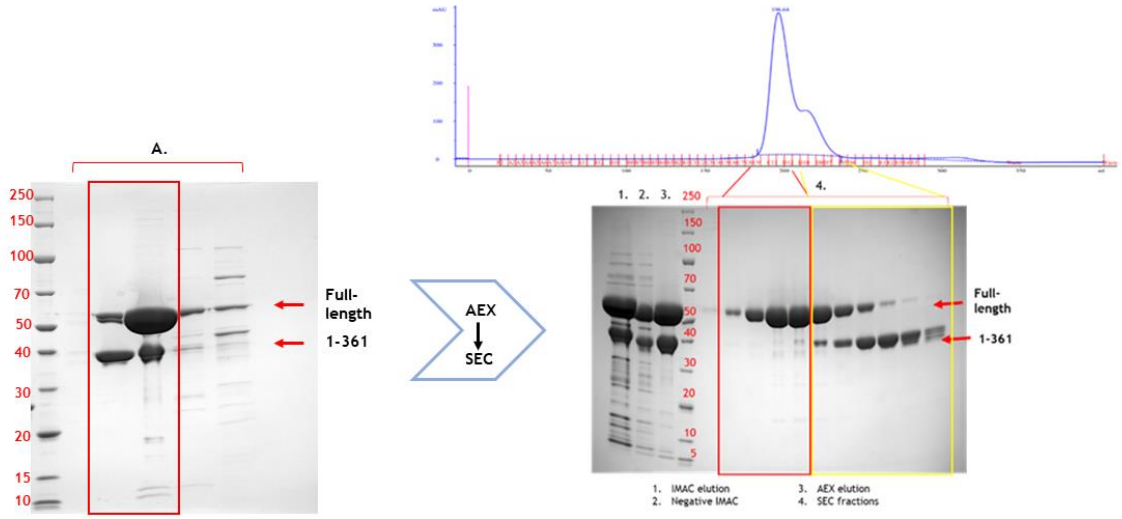


Figure 19: Final SEC step of purification of PLpro_NAB improvement. SDS-PAGE of samples (A) of thefirst batch of PLpro_NAB produced (final step was AEX). Right: PLpro_NAB final samples after optimized purification protocol: SEC chromatogram showing a peak corresponding to the PLpro_NAB full-length with a shoulder that corresponds to inhomogeneous sample containing the fragment 1-361. SDS-PAGE below shows the fractions loaded with the same fractionation order as chromatography.

The PLpro_NAB inactive mutant $C_{111}S$ was obtained by site-direct mutagenesis. This mutant was produced with the same protocol as the WT. Notably, no relevant degradation of the protein is visible during the purification, suggesting that the truncation site at the residue Asp361 could be a non-specific auto-cleavage site.

The δ-variant mutants $S_{466}R$ and $T_{467}K$ were also obtained by site-direct mutagenesis. Applying the same protocol as the WT, a slight nucleic acid contamination was detected in the final product. The NAB domain's mutagenesis introduces supplementary positive charge, resulting in increased bound nucleic acids and stronger interaction with putative RNAs.

Since the AEX was not sufficient to clean out the nucleic acids, this step was switched to Heparin chromatography, which brilliantly improved the purity of the protein by effectively removing the nucleic acid contamination.

Interestingly, even though the catalytic Cys111 is not mutated, the partial degradation after Asp361 is not detected during δ-variant mutants' purification (Figure 20 right).

Table 1 PLpro recombinant constructs produced with respective yields per L of culture.

	PLpro WT	PLpro $C_{111}S$	PLpro_NAB WT	PLpro_NAB $C_{111}S$	PLpro_NAB $S_{466}R$	PLpro_NAB $T_{467}K$
Yield/L of culture	15-19 mg	19-23 mg	14-32 mg	18-20 mg	18 mg	15 mg

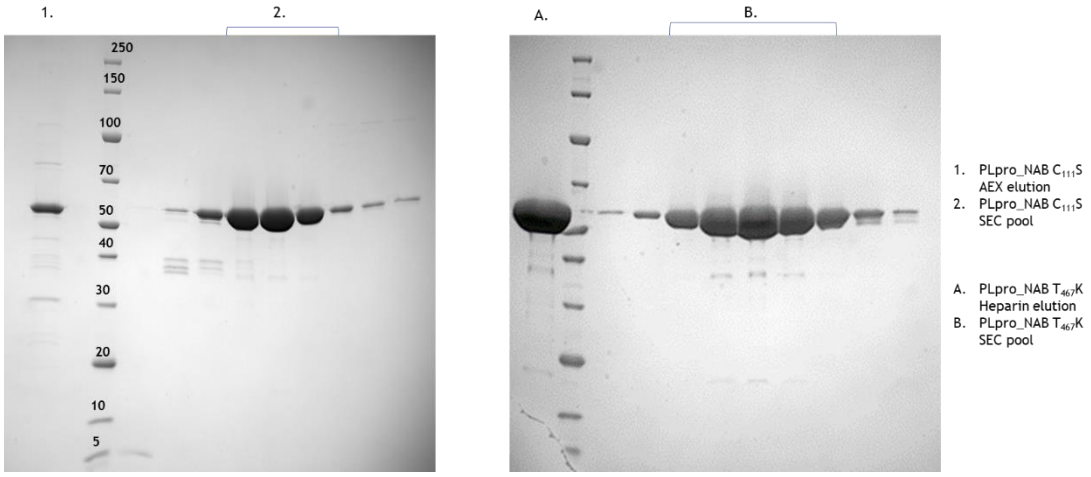


Figure 20: SDS-PAGE 4-20% of PLpro_NAB $C_{111}S$ (left) and $T_{467}K$ (right). The truncated form 1-361 is not present as in the WT.

5.1.2 PLpro characterization and crystallization

Recombinant PLpro and PLpro_NAB were employed to perform all of the subsequent studies presented in this thesis. The stability and oligomerisation in solution were characterized to ensure the quality of the protein products necessary to achieve our goals. Thermal stability was measured via TSA, testing different batches produced and different conditions used in the other experiment. In particular, the $C_{111}S$ mutant showed a higher and more constant melting temperature (Tm) than the WT in both constructs. PLpro and PLpro_NAB WT measured Tm presented a variability of 2/3°C from one assay and the other. Notably, the WT constructs were slightly but significantly more stable in the storage buffer at pH 8, than in the assay buffer at pH 7.5, while this difference in buffers' pH does not affect the $C_{111}S$ mutant. The presence of reducing agents also improves the stability, due likely to the reactivity of the catalytic Cys and the presence of superficial ones.

The oligomerization analysis was performed via Static Light Scattering (SLS) with an Omnisec instrument. As mentioned, PLpro_NAB eluted from the SEC as a larger protein, as it shows an elution profile corresponding to a predicted molecular weight (MW) of 70-75 kDa, that is in the middle between the MWs of the monomer and the dimer. The Omnisec detector calculates the experimental absolute MW from the light scattering signal and thereby demonstrates that both WT and mutants' samples were constituted by a well-pure monodisperse monomer. The PLpro construct is also a monodisperse monomer, although the SEC elution profile is more coherent with the real MW, probably due to a more globular behaviour than the PLpro_NAB.

Table 2: Average Tm of recombinant PLpro produced, referring to various TSA performed.

Construct	Assay buffer +DTT (50 mM Tris pH 7.5, 150 mM NaCl, 1mM DTT)	Assay buffer - -DTT (20 mM Tris pH 7.5, 150 mM NaCl)	Storage buffer	NMR buffer (Phosphate buffer pH 8, 50 mM NaCl)
PLpro_NAB WT	48° ± 2°C	46° ± 2°C	47° ± 0.5°C	47° ± 0.5°C
PLpro_NAB $C_{111}S$	51° ± 1°C	50° ± 1°C	50° ± 0.5°C	48° ± 0.5°C
PLpro_NAB $S_{466}R$	47° ± 1°C	Not Tested	47° ± 1°C	Not Tested
PLpro_NAB $T_{467}K$	48° ± 0.5°C	Not Tested	48° ± 0.5°C	Not Tested
PLpro WT	49° ± 2°C	48° ± 1°C	48° ± 0.5°C	Not Tested
PLpro $C_{111}S$	51° ± 1°C	50° ± 0.5°C	51° ± 0.5°C	Not Tested

Table 3: Parameters measured by HPLC-SLS of protein produced. Measured MWs deviate from the theoretical ones within 1 kDa.

Construct	MW (kDa)	MW/Mn	RV (mL)	[measured]/[injected]
PLpro_NAB WT	55.02	1.005	6.81	0.93
PLpro_NAB $C_{111}S$	56.37	1.001	6.81	0.99
PLpro_NAB $S_{466}R$	55.71	1.007	6.85	1.00
PLpro_NAB $T_{467}K$	55.63	1.008	7.02	0.91
PLpro WT	35.56	1.001	7.79	1.00
PLpro $C_{111}S$	36.04	1.002	7.98	0.87

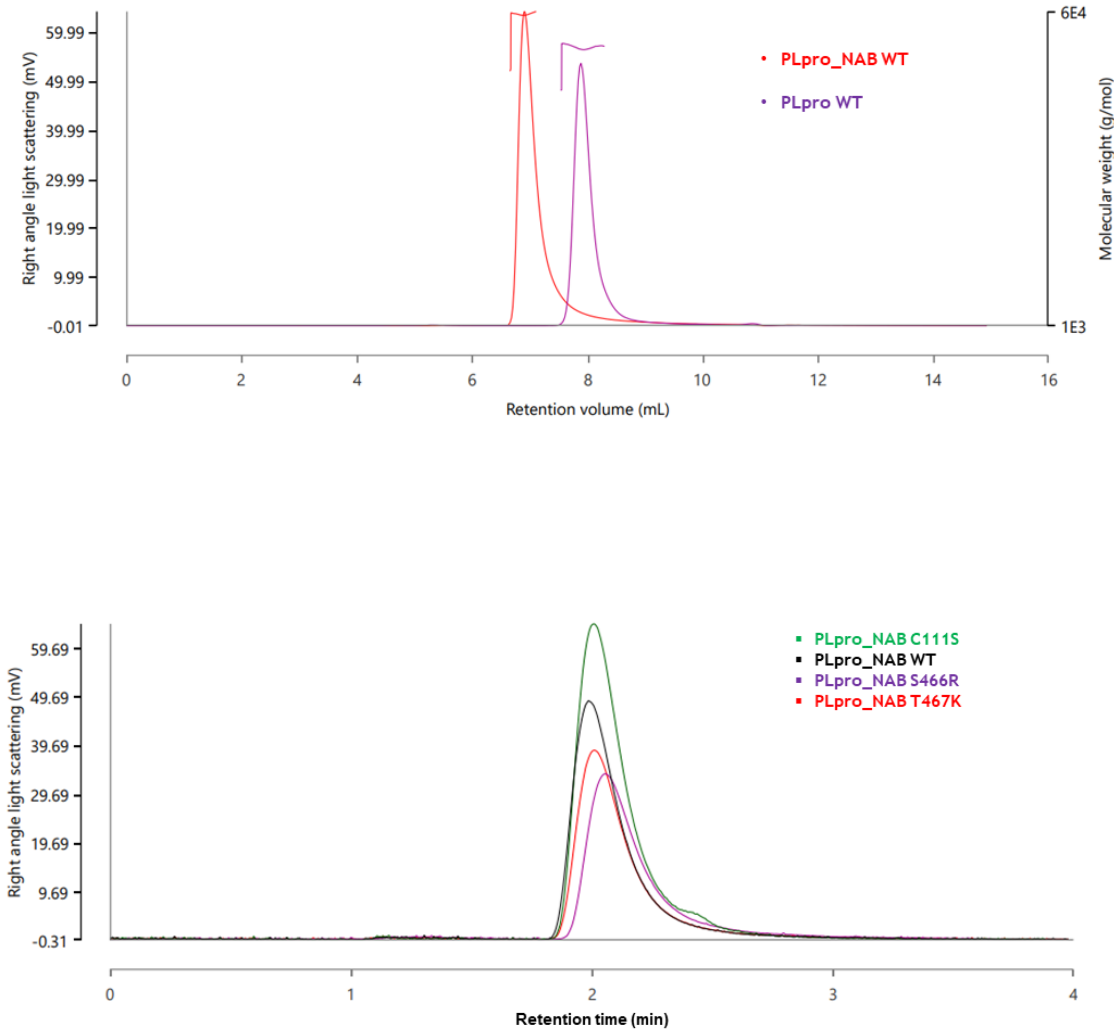


Figure 21: HPLC-SLS (OMNISEC) characterization of recombinant PLpro constructs. Above, comparison between SEC elution profiles of PLpro_NAB and PLpro WT; below, overlap of the elution peaks of PLpro_NAB WT and mutants.

Apo form crystals of the PLpro construct grew in three different conditions. The reported crystallization conditions by Osipiuk et al. [31] led to well-diffracting trigonal crystals of the PLpro $C_{111}S$ mutant (space group $P3_2$ 2 1) and trigonal crystals of the WT which grew less frequently by cross-seeding with $C_{111}S$ mutant seeds (Figure 22A). The apo form crystallization was optimized by changing the precipitant conditions. The sodium malonate condition led to the growth of well-diffracting and resistant crystals, sharing the space group of the crystals grown in the published condition (Figure 22B). In the perspective of owning crystals with less tight packing and thus more suitable for soaking candidate ligands, the reported condition using the Hampton additive containing caffein was reproduced [51]. Crystals in the third condition were obtained only with the PLpro $C_{111}S$, not the WT (Figure 22C). Even if crystals diffracted at a good resolution range, unfortunately, the crystallization conditions were not convenient for the co-crystallization or soaking of water-insoluble compounds, because of the high concentration of salts and the absence of organic co-solvents and polymers (i.e. PEG). Large crystallization screening of the PLpro_NAB was conducted as well, although no crystal grew in any of the conditions tested. We assume that the construct could be too flexible to crystallize, at least in apo form.

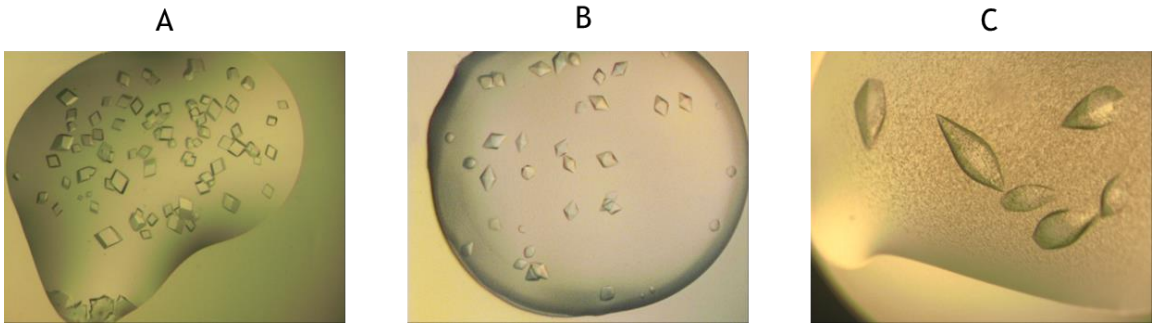


Figure 22 A: PLpro $C_{111}S$ apo form crystals grown in conditions reported by Osipiuk et al.; B: PLpro WT apo form crystals grown in 1.5 M Na malonate pH 6; C: PLpro $C_{111}S$ apo form crystals grown in reported conditions with the caffeine additive.

Crystals were tested at the XRD2 beamline in Elettra Sincrotrone. From the automatized data processing (Autoproc), no improvement in resolution or in quality of the diffraction data was obtained with respect to the deposited structures, the complete refinement of the collected data was not concluded.

5.1.3 Repurposing screening

Activity assay and high-throughput screening were performed at the Fraunhofer Institute for Translational Medicine and Pharmacology ITMP, in the Drug Discovery – ScreeningPort Department headed by Dr Philip Gribbon (Hamburg, Germany).

PLpro and PLpro_NAB were incubated with increasing concentrations of substrate. ISG15_AMC was selected as principal probe for the assay with a Km of 0.37 μM, while with Ub-AMP and Z-LRGG-AMC similar Km values were calculated (41.6 and 45.2 μM), thus Z-LRGG-AMC was selected as control-substrate. However, the elongated peptide substrate had a higher Km at 174 μM, therefore was discarded for further experiments. With ISG15-AMC substrate, no difference in terms of Km was determined between the PLpro and the PLpro_NAB construct. As a result of this evidence, the screening was conducted on the PLpro_NAB and PLpro was used as a control in order to ensure that inhibitors bound either the PLpro domain or the NAB. According to TSA analysis, the presence of DTT was also tested in the activity assay. 1 mM DTT increased indeed the protease's activity (ca. five-fold higher) and stabilized the protein during incubation and assay time. Stabilization and enhanced activity of the PLpro in the presence of DTT were also confirmed by mass mapping (Table 4). The presence of the reducing agent prevents the slight but significant oxidation of the catalytic Cys, leading to a loss of activity. According to screening conditions reported in literature, 1 mM DTT was used in the optimized setup.

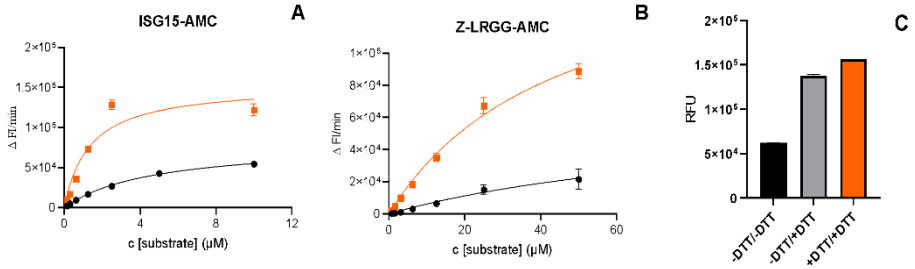


Figure 23: Activity curves of PLpro_NAB on the two substrates (A. ISG15-AMC; B. Z-LRGG_AMC) in presence and absence of 1 mM DTT. Histogram (C) shows the importance of reductant both during incubation and assay.

Table 4: cysteine oxidation state in presence and absence of 1 mM DTT detected by LC-MS/MS

Oxidation state	DTT	
	+	-
No-modified	99%	98%
Di-oxidation	1%	2%
Tri-oxidation	0%	0%
Total modified	1%	2%

Previously reported PLpro inhibitors PR619 and GRL-0617 were evaluated in dose response under optimized screening conditions. The IC_{50} for PR619 was calculated with 1.1 µM and 2.1 µM for GRL-0617, confirming published data [50,96]. To exclude any effect of DMSO which is used as solvent for compounds in screening, we tested the tolerance of PLpro to DMSO. No significant effect on enzymatic activity or thermal stability was observed for up to 2.5 % v/v DMSO.

The three repurposing libraries (8702 compounds) were tested on the PLpro_NAB using the previously optimized conditions. According to the results of the primary screen, all assay plates showed a Z' >0.5 (Mean 0.73) with a calculated S/B ratio value of ~ 2.1, whereby S refers to the DMSO control and B refers to values observed using 20 µM PR-619 as inhibitor. Outliers in the control area were excluded using the three-sigma method. Fifty-four compounds, which showed > 50 % inhibition, were selected for hit-confirmation (HC). HC was performed under primary screening conditions with 50 out of 54 hits confirming >50% inhibition. To eliminate compounds that interfere with the assay by quenching the generated AMC signal and causing false positive hits, the assay was first run to completion to generate free AMC. Test compounds were then added and interference with the fluorescence signal was determined. 49 compounds showed an interference with free AMC of less than 25% compared to the DMSO control and therefore were selected for further hit profiling in dose response, including the inhibitory activity also on the PLpro.

Table 5: Nominal hits with calculated IC50 δ 20 µM. Inhibition values are normalized to the activity of 20 µM PR619 (red) set as positive control for PLpro inhibition with 100% inhibition. GRL-0617 (green) was considered as positive control. PS- Primary Screen; HC-Hit Confirmation; HP- Hit Profiling. EOS - EU OpenScreen Collection; FhG – Fraunhofer Repurposing Library

Compound Name	PS PLpro_NAB Inhibition [%] 20 µM single point	HC PLpro_NAB Inhibition [%] 20 µM triplicates	HP PLpro_NAB IC50 [µM] Triplicates
PR-619	99,24	101,31	1,39
GRL-0617	Not Tested	105,7	2,12
Walrycin B (FhG)	129,2	160,54	0,040
PD119507	165,79	186,4	0,065
Walrycin B (EOS)	173,58	181,14	0,069
3-Methyl Toxoflavin	172,34	181,38	0,074
NSC-663284 (FhG)	124,8	150,9	0,220
Propidium-Iodide	116,44	135,55	0,240
SF1670	174,02	183,62	0,250
Ryuvidine	118,63	159,9	0,26
BVT-948	125,38	154,88	0,39
Ro-08-2750	105,62	119,29	0,40
9,10-Phenanthrenequinone	165,68	180,94	0,50
PD081125	165,23	186,1	0,58
PD086277	164,6	180,4	0,81
Dihydrotanshinone I	53,29	80,74	0,92
Semapimod (EOS)	117,98	140,16	1,21
NSC 663284 (EOS)	151,83	168,59	2,15

Cyanocobalamin	119,59	141,27	2,90
ML120	57,6	82,78	3,14
Alpha Lapachone	123	146,26	4,43
Sodium-Tanshinone-ii-A-Sulfonate	114,01	124,17	4,67
Beta-Lapachone	169,8	175,4	4,71
Semapimod (FhG)	62	118,8	5,23
TAS-103 (dihydrochloride)	90,83	154,91	5,94
BYK-204165	91,74	137,58	6,17
Sennoside A	123	159,68	7,01
Evans Blue	104,01	126,18	8,68
PT 1	84,33	87,07	9,32
Bacitracin (Zinc)	53,75	67,28	10,25
Menadione	65,41	115,37	11,50
Acriflavinium Hydrochloride	51,12	27,57	12,44
Homidium Bromide	91,06	149,01	13,69
CPI-169 (FhG)	69,85	99,81	14,17
YM-155	72,8	138,51	17,86
CPI-169 (EOS)	95,38	96,05	19,80
Pyrithione Zinc	50,6	80,42	20,41

5.1.4 Hit compounds (un)validation: an oxidative reaction leads to protein unfolding

A selection of ten among the most promising repurposing hit compounds was chosen for biophysical and structural evaluation, considering potency, structural properties, and commercial availability: Walrycin B, PD119507, NSC-663284, Ryuvidine, BVT-948, β-Lapachone, SF1670, PD081125, Semapimod, Dihydrotanshinone I (Figure 24).

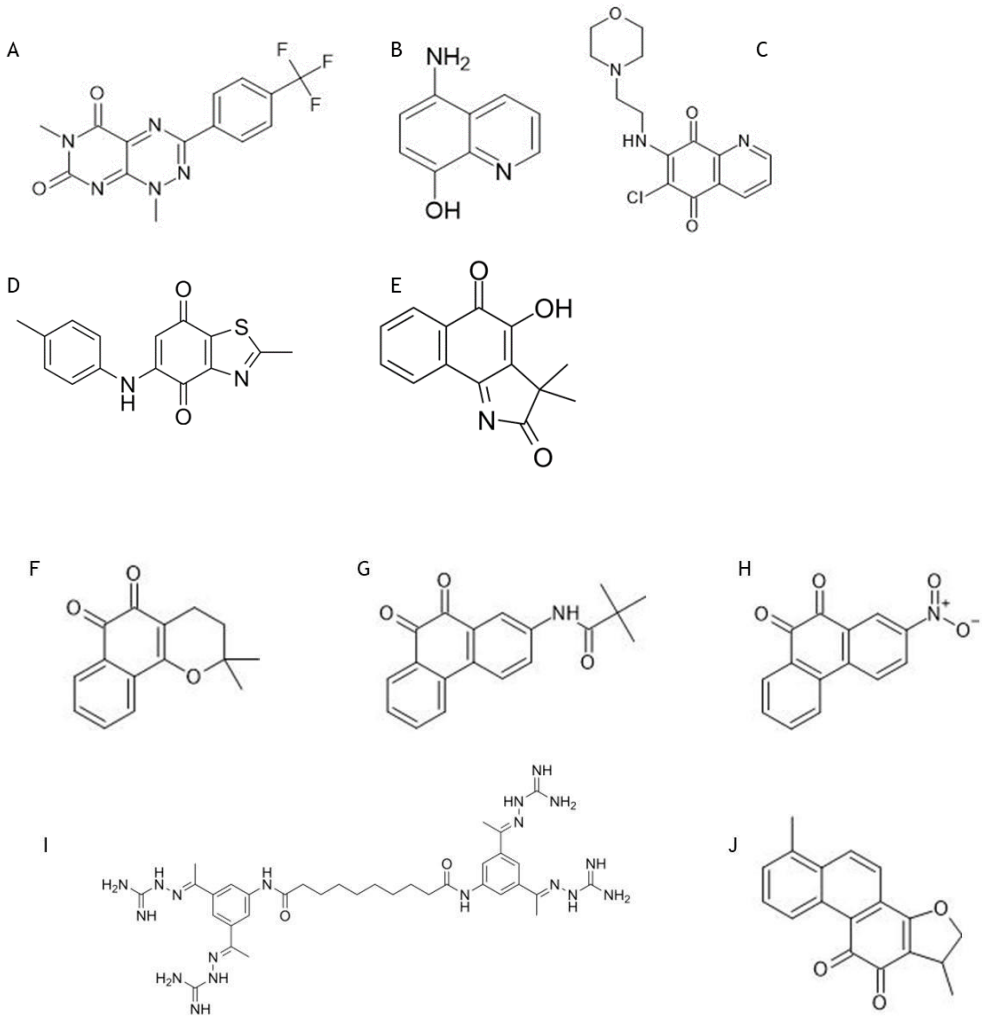


Figure 24: Chemical structures of selected hit-compounds. A. ***Walrycin B*** *(1,6-dimethyl-3-(4-(trifluoromethyl)phenyl)pyrimido[5,4-e][1,2,4]triazine-5,7(1H,6H)-dione); B.* ***PD119507*** *(5-Amino-8-hydroxyquinoline); C.* ***NSC-663284*** *(6-Chloro-7-(2-morpholin-4-yl-ethylamino)quinoline-5,8-dione); D.* ***Ryuvidine*** *(2-Methyl-5-[(4-methylphenyl)amino]benzothiazole-4,7-dione); E.* ***BVT-948*** *(4-Hydroxy-3,3-dimethyl-2H-benz[g]indole-2,5(3H)-dione); F.* ***β-Lapachone*** *(2,2-Dimethyl-3,4-dihydro-2H-benzo[h]chromene-5,6-dione); G.* ***SF1670*** *(N-(9,10-Dihydro-9,10-dioxo-2-phenanthrenyl)-2,2-dimethyl-propanamide); H.* ***PD081125*** *(9,10-Phenanthrenedione, 2-nitro); I.* ***Semapimod***

(N,N'-bis[3,5-bis[(E)-N-(diaminomethylideneamino)-C-methylcarbonimidoyl]phenyl]decanediamide); J. ***Dihydrotanshinone I*** *(1,6-dimethyl-1,2-dihydronaphtho[1,2-g][1]benzofuran-10,11-dione)*

Table 6: Repurposing hit-compounds with known targets and biological activity.

Inhibitor	**Known targets**	**Biological effect**
Walrycin B	WalR response regulator factor (RR)	Bactericidal effect on *S. aureus* and *B. subtilis* [97]
PD119507	Human Indolamine-2,3-dehydrogenase I	Pro-apoptosis in breast cancer cells [98]
NSC 663284	Cyclin-dependent kinases	Inhibition of cell cycle [99]
Ryuvidine	CDK4, SETD-8	Cytotoxic activity on cancer cell lines. [100]
BVT-948	Tyrosine-phosphatase	Enhanced insulin tolerance *in vivo* model [101]
β-Lapachone	Topoisomerase I	Pro-apoptotic [102]
SF1670	PTEN	Survival of transplanted granulocytes *in vivo* [103]
PD081125	carboxydiesterase I, AChE, S110-A4, PTPRC, coagulation factor XII	
Semapimod	IL-1β, TNF-α, IL-6, p38 MAPK	Inhibition of pro-inflammatory cytokines. In phase II for Chron's Disease [104]
Dihydrotanshinone I	LOX-1, NOX4, NF-κB,	Inhibits atherosclerosis, inhibits MERS-CoV entrance in model cells [105,106]

Except for Semapimod and PD119507, the compounds present structural similarities: Walrycin B, NSC 663284, Ryuvidine, and BVT-948 have a quinone-like scaffold, while the others have a dihydrophenanthrene-like scaffold.

A wide number of co-crystallization trials were performed to obtain co-structures of the PLpro with the most potent inhibitors, using commercially available crystallization screens and reported conditions in the PDB with other compounds. Due to the failure of the trials, apo form crystals gained in the sodium-malonate condition and in the condition with the caffeine additive were used to soak the compounds dissolved in 100% DMSO into the crystals. In the first case, even though apo trigonal crystals proved to be quite resistant, in the presence of the repurposing compounds crystals dissolved in the soaking drop or did not present any ligand bound. In the second case, crystals grew in highly concentrated $(NH_4)_2SO_4$ which impeded the solubilization of the compounds even at low concentrations in the soaking drop.

Observed that adding the dissolved compounds to the protein solution, the PLpro tended to precipitate immediately, a Thermal Shift Assay was performed to elucidate the nature of the interaction with the compounds. Compounds were tested at increasing concentrations and Tm values were compared with the Tm of the protein in the same DMSO concentration as in the presence of compounds. DMSO at low concentration (below 5%) slightly stabilizes the protein. GRL-0617 and PR-619 were used as controls also in this assay. While GRL-0617 stabilized the PLpro and PLpro_NAB WT and $C_{111}S$ in a similar way, the PR-619 at higher concentrations decreased the WT Tm of 7°C, effect not visible on the $C_{111}S$ mutant, suggesting that the catalytic Cys111 was involved in the interaction (Figure 25). On the other hand, the compounds showed a harsh destabilizing effect on both the PLpro and PLpro_NAB with a concentration-dependent relation. The same experiments were applied to the catalytic mutant $C_{111}S$. In this case, the stability of the protein was not remarkably affected (Figure 26).

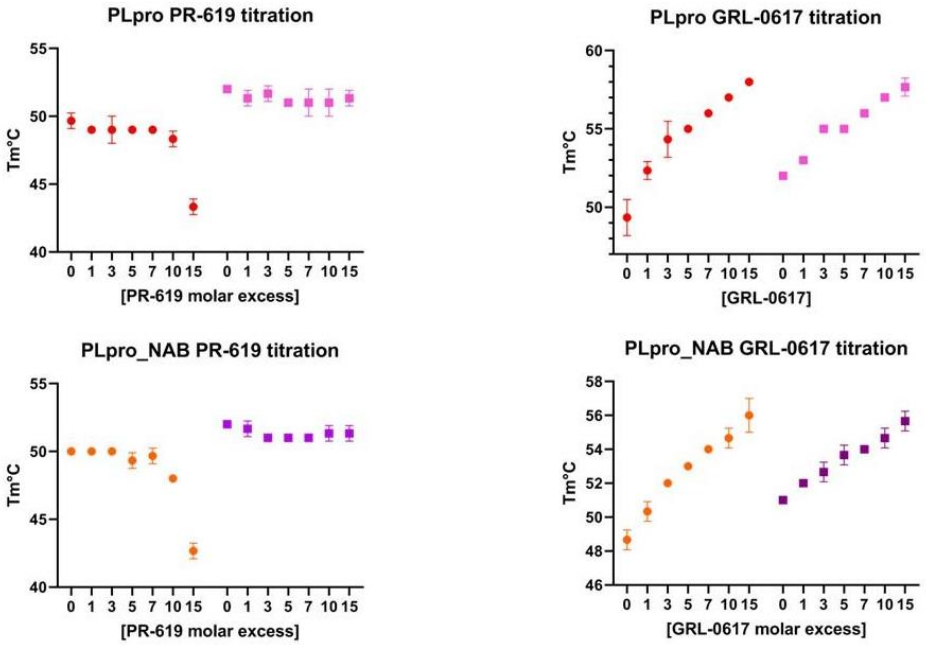


Figure 25: TSA titration of PR-619 and GRL-0617 with PLpro and PLpro_NAB WT (red and orange dots, respectively), PLpro and PLpro_NAB $C_{111}S$ (pink and purple, respectively).

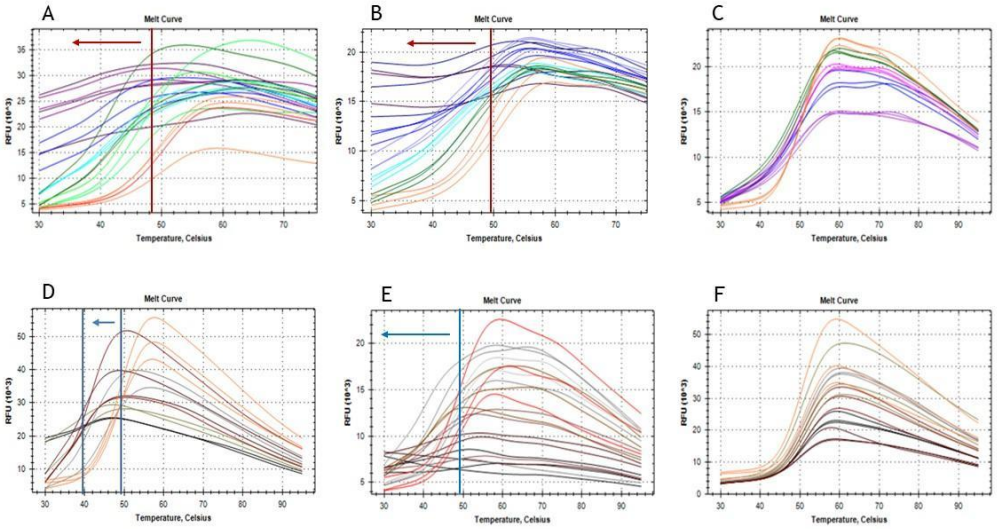


Figure 26: TSA graphics show a negative shift or a complete unfolding of the WTs in presence of compounds. The $C_{111}S$ mutant does not show a significant negative shift. Walrycin B titration on PLpro_NAB WT (A), PLpro WT (B), and PLpro_$C_{111}S$ (C); NSC 663284 titration on PLpro_NAB WT (D), PLpro WT (E), PLpro_$C_{111}S$ (F). The other compounds showed a similar behaviour, except for Semapimod of which no shift was detected.

A second proof of the destabilizing mechanism of action was provided by the limited proteolysis. PLpro WT and $C_{111}S$ were treated with the chosen proteases, at the same time the procedure was conducted in the presence of Walrycin B, PD119507, or Dihydrotanshinone I as representative compounds in protein/compound ratio 1:5. SDS-PAGE displayed that the presence of the compounds clearly triggers the unfolding of the WT in a resulting more evident digestion by proteases (Figure 27.2), while do not cause the same effect on the mutant (Figure 27.ii).

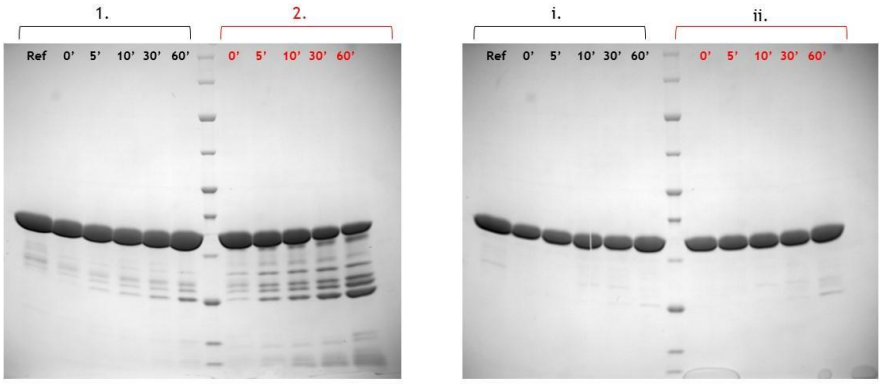


Figure 27: Limited proteolysis conducted with trypsin in protease/PLpro ratio 1:1000 at 22°C. 1. PLpro WT; 2. PLpro WT in the presence of 5x PD119507; i. PLpro $C_{111}S$; ii. PLpro $C_{111}S$ in the presence of 5x PD119507.

The majority of the hit compounds resulting from the screening have a quinone-like or orthoquinone-like scaffold. As reported in the literature, such scaffolds may be affected by the presence of strong reducing agents in the assay conditions. Hyun Lee et al. demonstrated that different reducing agents in the activity assay affect the truthfulness of screening readout, selecting viral Cys-proteases, such as 3C-like and PLpro from SARS-CoV, as a model. Strong reducing agents like tris(2-carboxyethyl)phosphine (TCEP), DTT, or β-mercaptoethanol may lead to poor hits (false positives) by altering the inhibitory activity of screened compounds [107]. Deubiquitinases (DUBs) are susceptible to oxidation by superoxide species. In particular, catalytic Cys of DUBs, which is more reactive, undergoes different levels of oxidation starting from sulfenic acid (-SOH), sulfinic acid ($-SO_2H$) and sulfonic acid ($-SO_3H$). The last two are not reversible, even in the intracellular environment [108]. The oxidative reaction caused by orthoquinone compounds in presence of strong reducing conditions was also pointed out on other enzymes presenting a catalytic Cys.

Yu-bo Zhou et al. proposed a mechanism of reaction catalysed by DTT or β-mercaptoethanol through which the orthoquinone compound is reduced to its semiquinone anion radical ($RQ^{\cdot}$), generating superoxide species that rapidly oxidize the catalytic Cys of the target [109].

Notably, β-Lapachone was described to generate irreversible sulfonic acid in reducing conditions on the catalytic cysteine of USP family DUBs like USP2, USP1, and USP7 [110]. The reaction scheme is illustrated in the following figure.

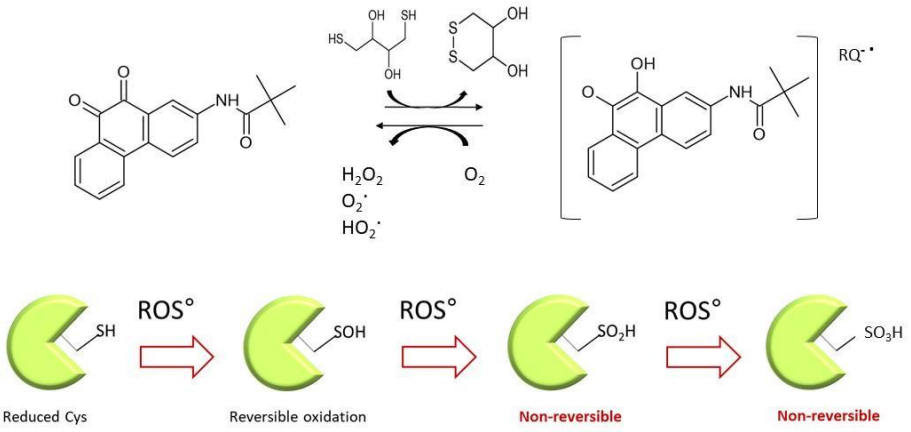


Figure 28: Proposed oxidative reaction summarized according to published data. The orthoquinone compound (SF1670 taken as a model) is reduced to a semiquinone radical anion that reacts in the water environment to generate superoxide species, which in turn oxidize protein cysteine.

Taking into consideration the previously published information and the results from TSA and LiP experiments, the mechanism of hit-compound inhibition was further explored to determine if oxidative reactions also occurred with these compounds. Additionally, some orthoquinone compounds had already been invalidated by Chunlog Ma and Jun Wang, especially the tanshinone family whose activity towards the PLpro was proven to be related only to the presence of reducing agents [67].

The repurposing hits were tested again in activity assay with the main setup, comparing the inhibition in the presence or absence of 1 mM DTT. Consistently with the information acquired from the literature, in the absence of reducing agents, the

compounds completely lost the inhibitory activity. The two positive controls represented by the GRL-0617 and PR-619 retained their activity on the PLpro, even though the PR-619 inhibition slightly decreased. The presence of DTT during the incubation of the PLpro with compounds also affects the inhibitory readout, even if the assay was then conducted in absence of DTT (figure 29).

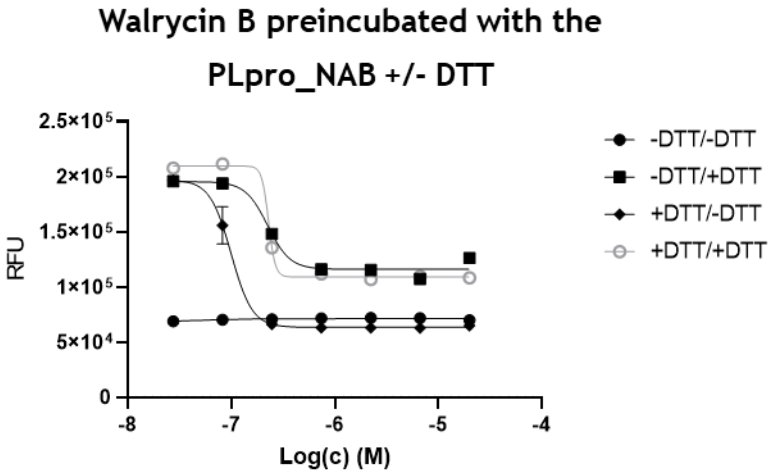


Figure 29: Activity assay performed with Walrycin B in the presence of the PLpro_NAB with or without DTT conducted at the Fraunhofer ITMP in Hamburg. The fluorescence signal clearly reveals that DTT alters the inhibitory activity even if present only in the incubation conditions.

A new TSA was performed without reducing agents to confirm the second activity test evidence. As a consequence of the activity assay, the storage buffer was exchanged overnight in dialysis against the assay buffer without reducing agents. The TSA was then performed in the same buffer conditions. As expected, no negative shift or evidence of the strong unfold-inducing effect was detected. Like the new activity results, TSA did not show any proof of binding, since it was not possible to reveal appreciable positive shifts. In this assay, only GRL-0617 and PR-619 kept the same behaviour on the PLpro and PLpro_NAB (figure 30).

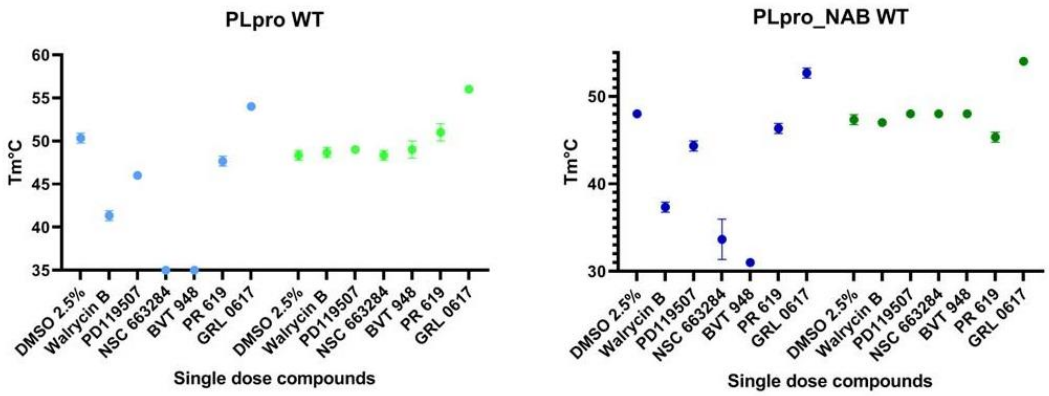


Figure 30: TSA on PLpro and PLpro_NAB WT with single doses of repurposing hit and control inhibitors, +/- 1 mM DTT. Data highlight the destabilizing mechanism owed to the presence of DTT, without DTT there is no shift hence no proof of interaction with the PLpro.

To definitely demonstrate the oxidative mechanism of the compounds on the PLpro, a mass mapping was performed at CEINGE in Naples by the group of Prof. Maria Monti. PLpro_NAB was incubated with/without DTT in the presence of Walrycin B or PD119507 with a molar ratio of 1:5, then hydrolysed with pepsin and analysed by LC-MS/MS. In order to appreciate the modifications of the Cys111, peptides were identified and quantified with an extracted ion current approach: the percentage of modification species was calculated as the ratio of the total area of all species containing the specific modification and the total area of all the species (modified and not modified), with results expressed as a percentage (Table 7).

Table 7: Percentage of di- and tri-modification of the catalytic Cys111 in the presence of compounds, with or without DTT. The LC-MS/MS analysis perfectly justifies the oxidative hypothesis.

	Walrycin B		**PD119507**	
DTT 1mM	**+**	**-**	**+**	**-**
No-modified	19%	90%	12%	97%
Di-oxidation	67%	10%	59%	3%
Tri-oxidation	13%	0%	30%	0%
Total modified	81%	10%	88%	3%

The Cys111 is highly di- or tri-oxidized by the presence of the compounds and the DTT, implying that there is the formation of sulfinic or sulfonic acid on the thiol group of the Cys which was previously proven to be irreversible, while is not observable in the absence of one or both the components.

5.1.5 New hit compounds

In light of the previous results which led to the invalidation of the hit inhibitors, the HT screening was revised. Different reducing agents were tested at the Fraunhofer ITMP in the presence of the compounds and compared to the 1 mM DTT conditions or the absence of reductants, in order to find a reducing condition suitable for the screening of the library of compounds and the activity/stability of the PLpro. 1 mM L-Cys was chosen as mild (but necessary) reducing agent and the activity assay was repeated on the 54 confirmed hit compounds. Only 7 of those retained the inhibitory activity on the PLpro in the L-Cys buffer: CPI-169, Semapimod, Sennoside A, Purpurogallin, DOM_SIM 710 (3',4',5',5,6,7-hexahydroxyflavone), and the positive controls PR-619 and GRL-0617. SRT 1720 was not confirmed since a second batch of compound did not show any activity.

Table 8: Results of the revised H screening. Most of the repurposing compounds lost their inhibitory activity in the mild reducing condition. Inhibitors active in L-Cys buffer are in bold.

Compound Name	PLpro_NAB inhibition [%] triplicates DTT BUFFER	PLpro_NAB IC50 [µM] DTT BUFFER	PLpro_NAB inhibition [%] triplicates L-CYS BUFFER	PLpro_NAB IC50 [µM] L-CYS BUFFER
PR-619	**101,31**	**1,39**	**106,6**	**3,7**
GRL-0617	**105,7**	**2,12**	**117**	**4,36**
SRT 1720	**109,6**	**>20**	**169,11**	**0,82**
Purpurogallin	**62,76**	**>20**	**100,79**	**1,03**
DOM_SIM_710	**65**	**>20**	**154,95**	**1,49**
Semapimod (FhG)	**118,8**	**5,23**	**51,28**	**14,4**
CPI-169 (FhG)	**99,81**	**14,17**	**70,67**	**17,47**
Sennoside A	**159,68**	**7,01**	**149,39**	**18,2**
Walrycin B (FhG)	160,54	0,04	7,9	>20
PD119507	186,4	0,065	-16,42	>20
Walrycin B (EOS)	181,14	0,069	-35,9	>20
3-Methyltoxoflavin	181,38	0,074	-26,71	>20
NSC-663284 (FhG)	150,9	0,22	7,56	>20

Propidium-Iodide	135,55	0,24	14,9	>20
SF1670	183,62	0,25	-32,73	>20
Ryuvidine	159,9	0,26	-33,69	>20
BVT-948	154,88	0,39	-16,22	>20
Ro-08-2750	119,29	0,4	-33,99	>20
9,10-Phenanthrenequinone	180,94	0,504	-40,79	>20
PD081125	186,1	0,58	1,14	>20
PD086277	180,4	0,807	-30,11	>20
Dihydrotanshinone I	80,74	0,92	-176,06	>20
Semapimod (EOS)	140,16	1,21	50,46	>20
NSC 663284 (EOS)	168,59	2,15	-38,39	>20
Cyanocobalamin	141,27	2,9	32,29	>20
ML120	82,78	3,14	-56,07	>20
Alpha Lapachone	146,26	4,43	-6,97	>20
Sodium-Tanshinone-ii-A-Sulfonate	124,17	4,67	-51,93	>20
Beta-Lapachone	175,4	4,71	-37,39	>20
TAS-103 (dihydrochloride)	154,91	5,94	17,75	>20
BYK-204165	137,58	6,17	7,1	>20
Evans Blue	126,18	8,68	-67,82	>20
PT 1	87,07	9,32	-36,09	>20
Bacitracin (Zinc)	67,28	10,25	6,99	>20
Menadione	115,37	11,5	-45,68	>20
Acriflavinium Hydrochloride	27,57	12,44	-99,55	>20
Homidium Bromide	149,01	13,69	18,32	>20
YM-155	138,51	17,86	-5,61	>20
CPI-169 (EOS)	96,05	19,8	49,41	>20
Pyrothione Zinc	80,42	20,41	16,01	>20

New hit compounds could be divided into 3 groups: compounds which retain a similar activity in both buffers (CPI-169, PR-619, GRL-0617); compounds which have decreased activity in the L-Cys buffer (Semapimod, Sennoside A); compounds with increased activity in the L-Cys buffer, thus being putatively reactive with Cys (DTT > L-Cys): Purpurogallin, DOM_SIM_710. In addition, these two last compounds showed a different activity on the PLpro construct. In particular, DOM_SIM_710 and Purpurogallin were more potent against PLpro compared to PLpro_NAB, whereas Semapimod (FhG) was only active against PLpro_NAB, although this result was not confirmed using Semapimod from the EU-OpenScreen collection. All other compounds showed comparable potencies against both constructs.

Selectivity on other viral and human targets was also tested. The inhibition on the PLpro of SARS-CoV2 was compared to the measured inhibition on the PLpro of SARS-CoV and the SARS-CoV2 Mpro as viral targets, to the inhibition on USP7 and USP14 as representative of DUB family, and Cathepsin-L as representative human Cys-protease. CPI-169, Semapimod, Purpurogallin, DOM-SIM_710 and GRL-0617 showed preferred inhibition of PLpro over Mpro, similar activity on the SARS-CoV PLpro instead. CPI-169 (and GRL-0617) results selective for PLpro as they do show no inhibition against the selected human proteases. The remaining compounds showed inhibition against at least one of the human targets.

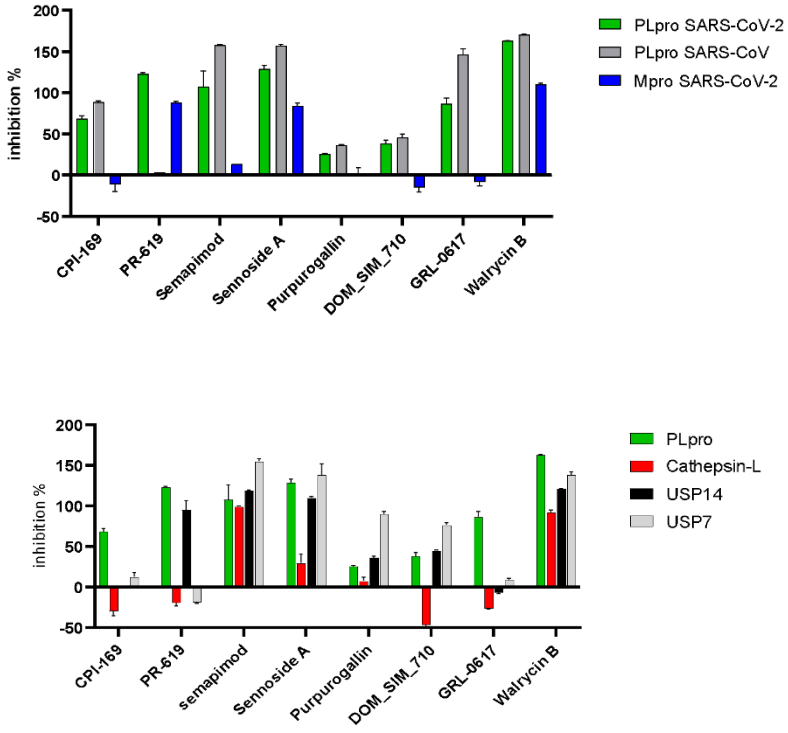


Figure 31: Inhibition of the new hit compounds on other targets. Walrycin B was selected as representative of the DTT-dependent mechanism, tested in presence of 1 mM DTT.

5.1.6 Computational and experimental studies on CPI-169

CPI-169 became the main hit repurposing compound of this project. CPI-169 showed promising results at a pre-clinical level against lymphoma in an *in vivo* model. The compound potently targets the histone-lysine N-methyltransferase enzyme coded by the *EZH2* gene (Enhancer of Zeste Homolog 2), the Polycomb repressive complex 2 catalytic component [111].

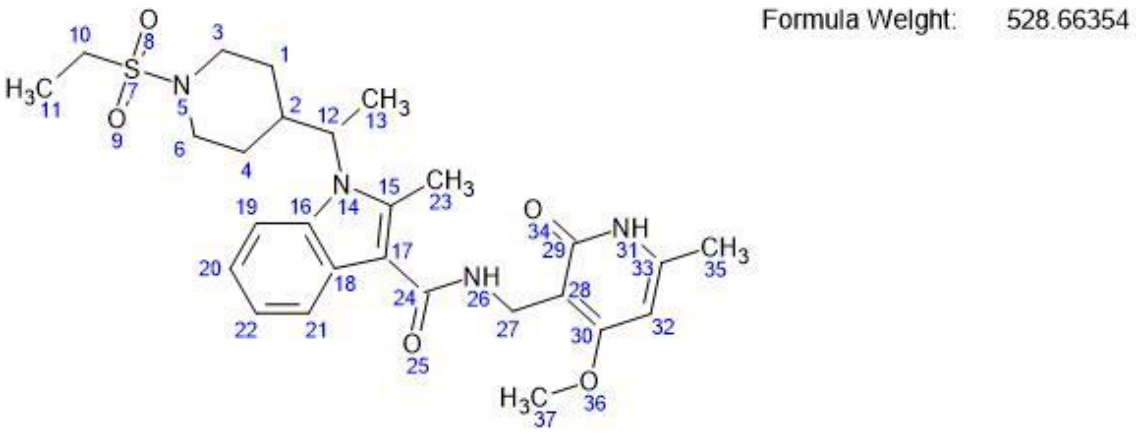


Figure 32: Structure of CPI-169 and molecular weight.

CPI-169 binding to the PLpro was tested by TSA as preliminary validation. Similarly to the GRL-0617, the CPI-169 increases the Tm of both constructs PLpro and PLpro_NAB in a dose-dependent way. Interestingly, the same behaviour was detected for the inactive mutant $C_{111}S$, suggesting the catalytic Cys was not involved in the binding (Figure 33).

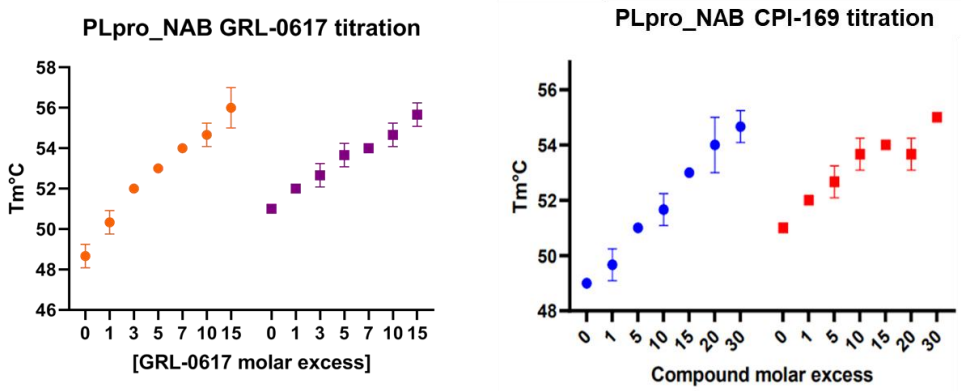


Figure 33: TSA titration of CPI-619 on PLpro_NAB WT (blue dots) and PLpro_NAB $C_{111}S$ (red squares), in comparison with titration of GRL-0617 on PLpro_NAB WT (orange dots) and $C_{111}S$ (purple squares). The two compounds present a similar trend at the TSA, even though CPI-169 confirms to be less potent than the GRL-0617.

Due to the biochemical and biophysical analogies of the two compounds, a similar binding mode of the two was suspected. A first confirmation of the hypothesis was given by a docking experiment performed by our collaborators in Hamburg, conducted with the PLpro/GRL-0617 co-crystal structure as reference (PDB code: 7JRN). Indeed, CPI-169 occupied the same allosteric cavity of GRL-0617, with a PLP scoring function comparable to the reference compound, even if CPI-169 is less potent. The CPI-169 overlaps the reference compound in the docking pose, except for the pyridine ring interacting with the PLpro out from the pocket (Figure 34d).

The interaction with the PLpro residues was explored using LigPlot software [113], highlighting that CPI-619 interacts with more residues than the GRL-0617. Although, CPI-169 does not make any hydrogen bonds with the residues D164 and Q269, only weaker interactions, consistently with the lower potency of the compound with respect to the reference one (Figure 35).

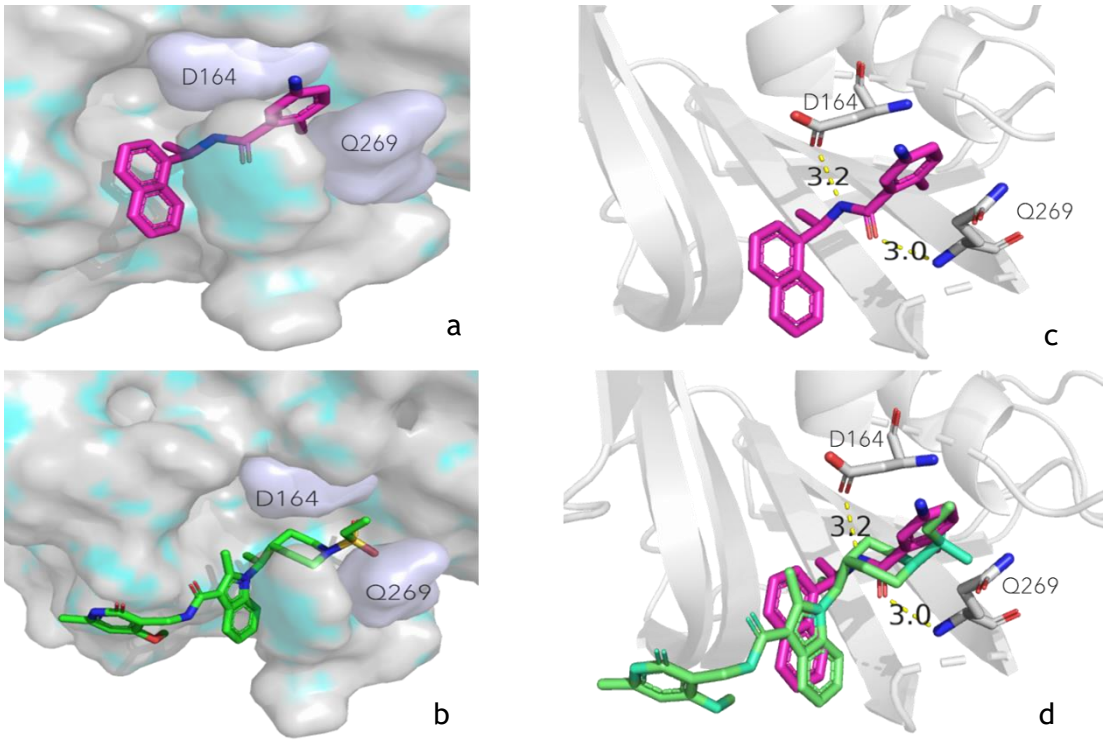


Figure 34: [a] Pose of the GRL-0617 in the binding pocket from structure 7JRN; [b] Docking pose of CPI-169 in the same binding pocket; [c] GRL-0617 with distances to Q269 and D164 showing the hydrogen bonds; [d] Overlap of the two compounds in the binding pocket.

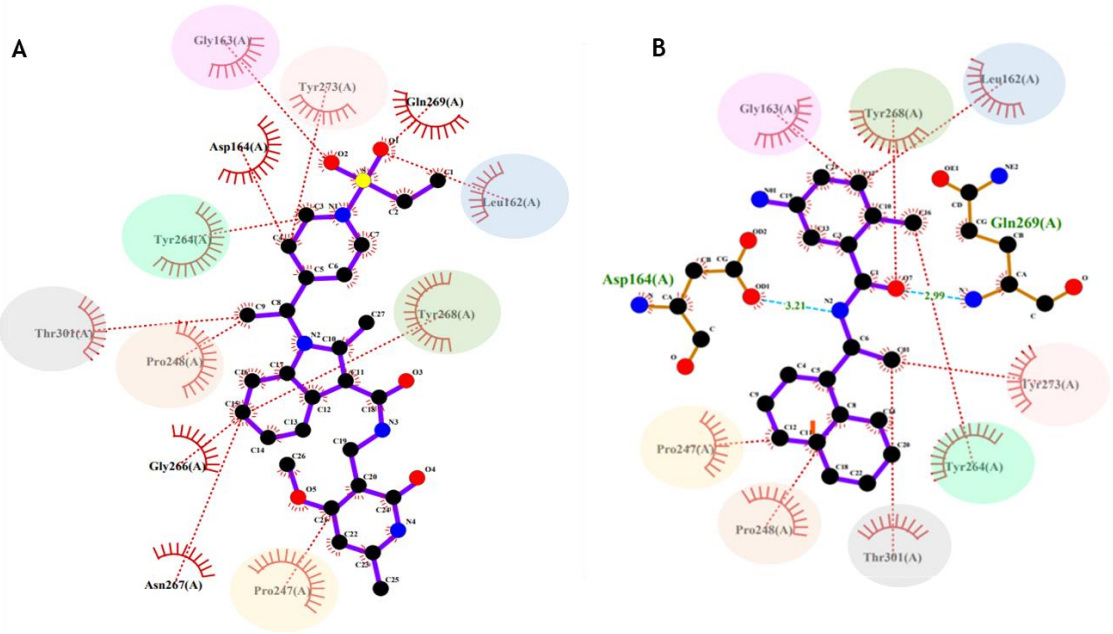


Figure 35: Comparison of the residues of PLpro interacting with the CPI-169 (A) and GRL-0617 (B), analysed with LigPlot. Residues are highlighted with the same colour in the two plots.

Co-crystallization trials with PLpro $C_{111}S$ and CPI-169 were conducted without obtaining any crystals. Soaking trials were also performed using formed crystals in the condition with the additive containing caffeine. Due to the fragility of crystals and probably the low solubility of the compound in the cryoprotection solution, crystals did not survive the soaking attempts.

5.1.7 Ligand-based NMR experiments

Experimental setup and binding studies on PR-169 and Semapimod

Surface Transfer Difference NMR (STD NMR), performed at the Slovenian NMR Centre in Ljubljana, was chosen to confirm the binding and get structural information about the ligand in the binding pocket to be compared to the docking prediction.

A first ^{1}H spectrum of the PLpro_NAB in STD assay buffer (20 mM Pi buffer, 50 mM NaCl, 10% D_2O, pH 8) was recorded. A signal at -0.772 ppm was detected and selected for STD experiments (Figure 36). The applicability of the STD NMR to this target and inhibitors was tested with one of the two positive controls, the PR-619. A protein/ligand ratio of 1:100 was selected and the concentration of DMSO of 5% which led to keeping the protein stable during the measurement time. 1D STD NMR experiment shows an evident saturation transfer from the PLpro_NAB to the ligand. The H3 signal is appreciable in the difference spectrum confirming the binding, while the primary amine proton signal is not present in the difference spectrum, probably due to the proton exchange in water solution (Figure 37).

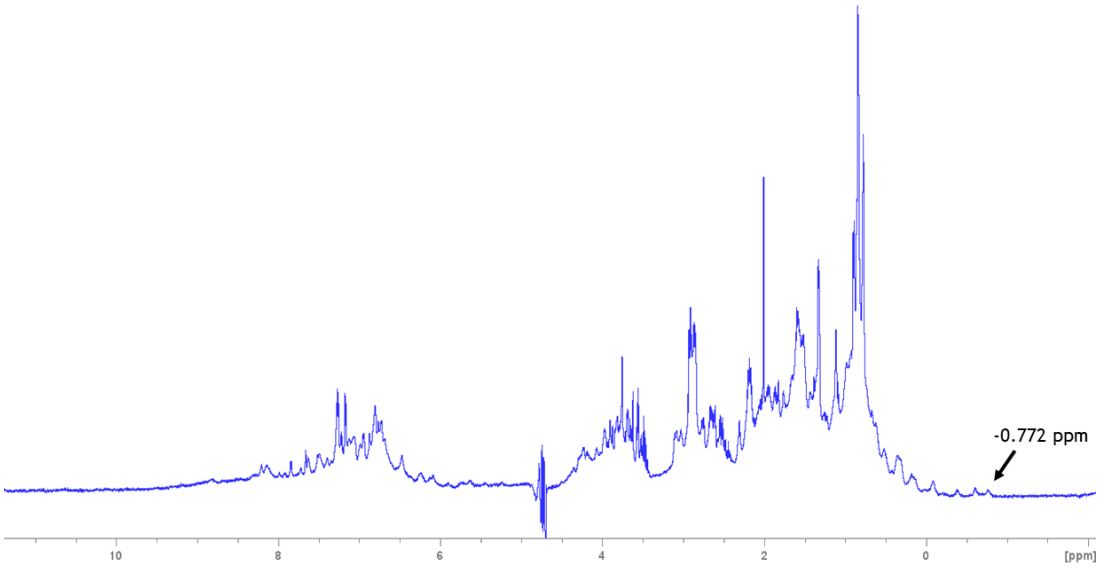


Figure 36: 1D ^{1}H spectrum of 50 µM PLpro_NAB, with pointed the ppm of signal at which the selective saturation was applied in further experiments.

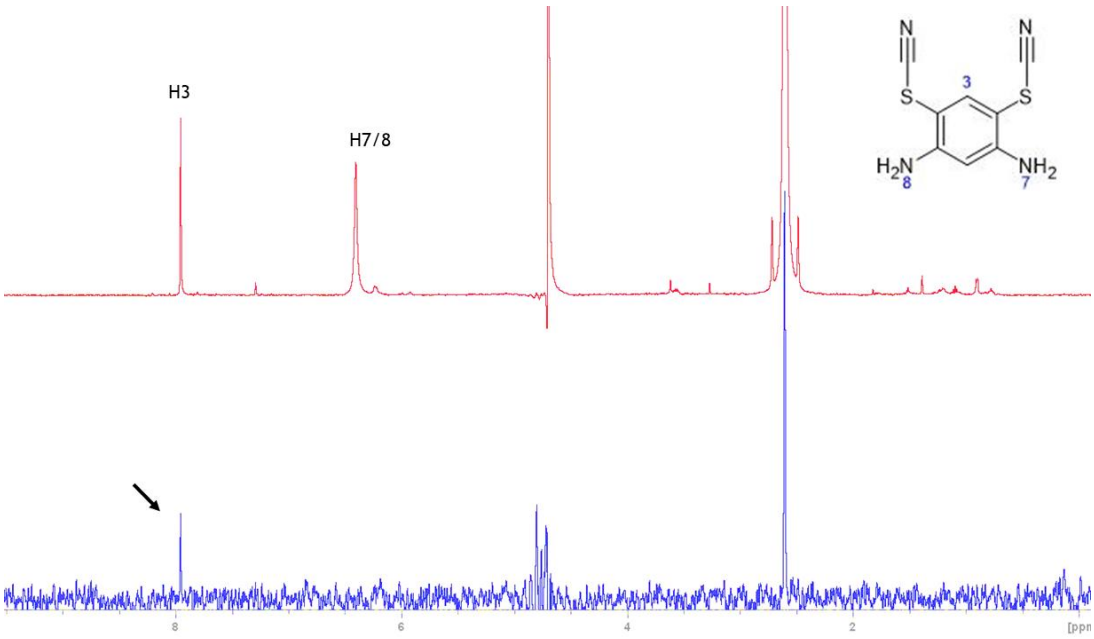


Figure 37: 1D STD NMR of 0.15 mM PR-619 in the presence of 1.5 µM PLpro_NAB WT. In red, the reference spectrum (off-resonance) with the proton assignment; in blue, the difference spectrum showing a saturation transfer to the H3 of the PR-619.

The same setup was used for the binding confirmation of Semapimod, which exhibited different behaviours in the activity assay depending on the batch of compound.

A preliminary TSA titration of the Semapimod from EOS library, given by our colleagues from the Fraunhofer did not show any shift, supporting the activity results. 1D and 2D NMR experiments were also conducted to obtain binding validation or invalidation and to ensure the integrity and pureness of the molecule. While the Semapimod characterization affirmed its health, the 1D STD NMR finally invalidated the interaction with the PLpro_NAB, where no signal was evidenced in the difference spectrum (Figure 38).

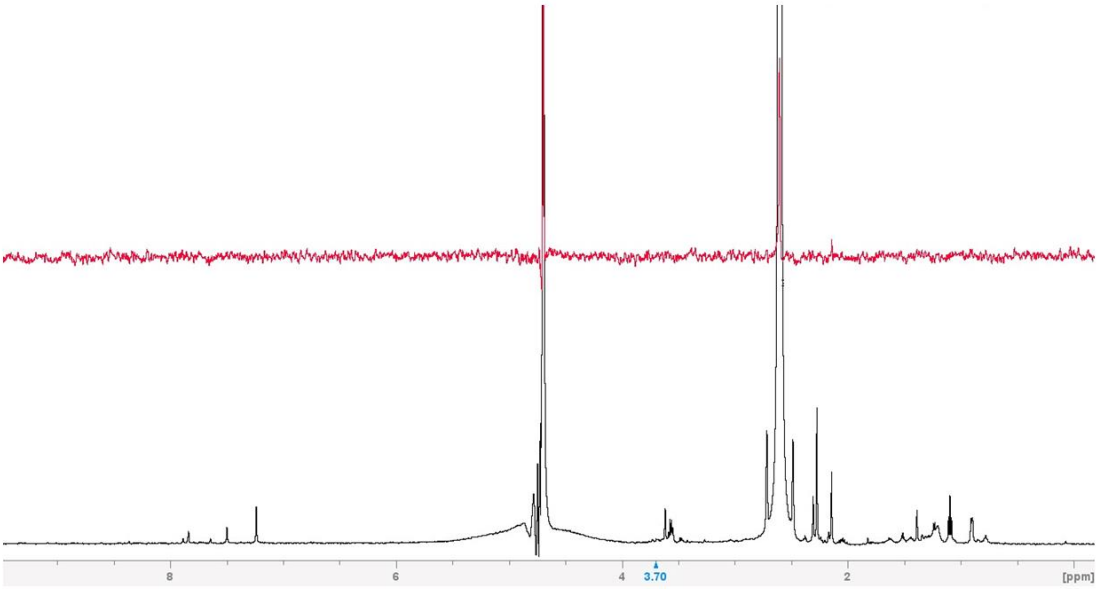


Figure 38: 1D STD NMR of 0.1 mM Semapimod in the presence of 1 μM PLpro_NAB. Below, the reference spectrum (off-resonance); above, the difference spectrum shows no signals except for water and DMSO.

Characterization of CPI-169 binding to PLpro_NAB

CPI-169 proton full assignment was performed considering 1D ^{1}H, 2D-HSQC, 2D-TOCSY and 2D-NOESY spectra in the STD buffer. The interaction with the PLpro_NAB was then demonstrated by the established experimental setup by 1D STD NMR. Since several proton signals were revealed in the difference spectrum, a second 1D STD experiment was conducted with a shorter protein saturation time (1s) to map the binding epitope of CPI-169 from the calculated amplification factors (Table 9). The binding is indeed distributed along the whole molecule, especially appreciable in the two extremes, the pyridine ring and the ethyl-sulfone (Figure 39).

Table 9: Amplification factors (A_{STD}) and % of relative A_{STD} calculated at 0.5 mM CPI-169 in the presence of 5 µM PLpro_NAB. Rel. A_{STD} highlights the involvement in the interaction with the target.

Proton	A_{STD}	Rel.A_{STD}
H32	1.93111	100%
H37	1.510432	78%
H10	1.706214	88%
H23	1.47266	76%
H35	1.604915	83%
H13	1.722235	89%
H11	1.880406	97%

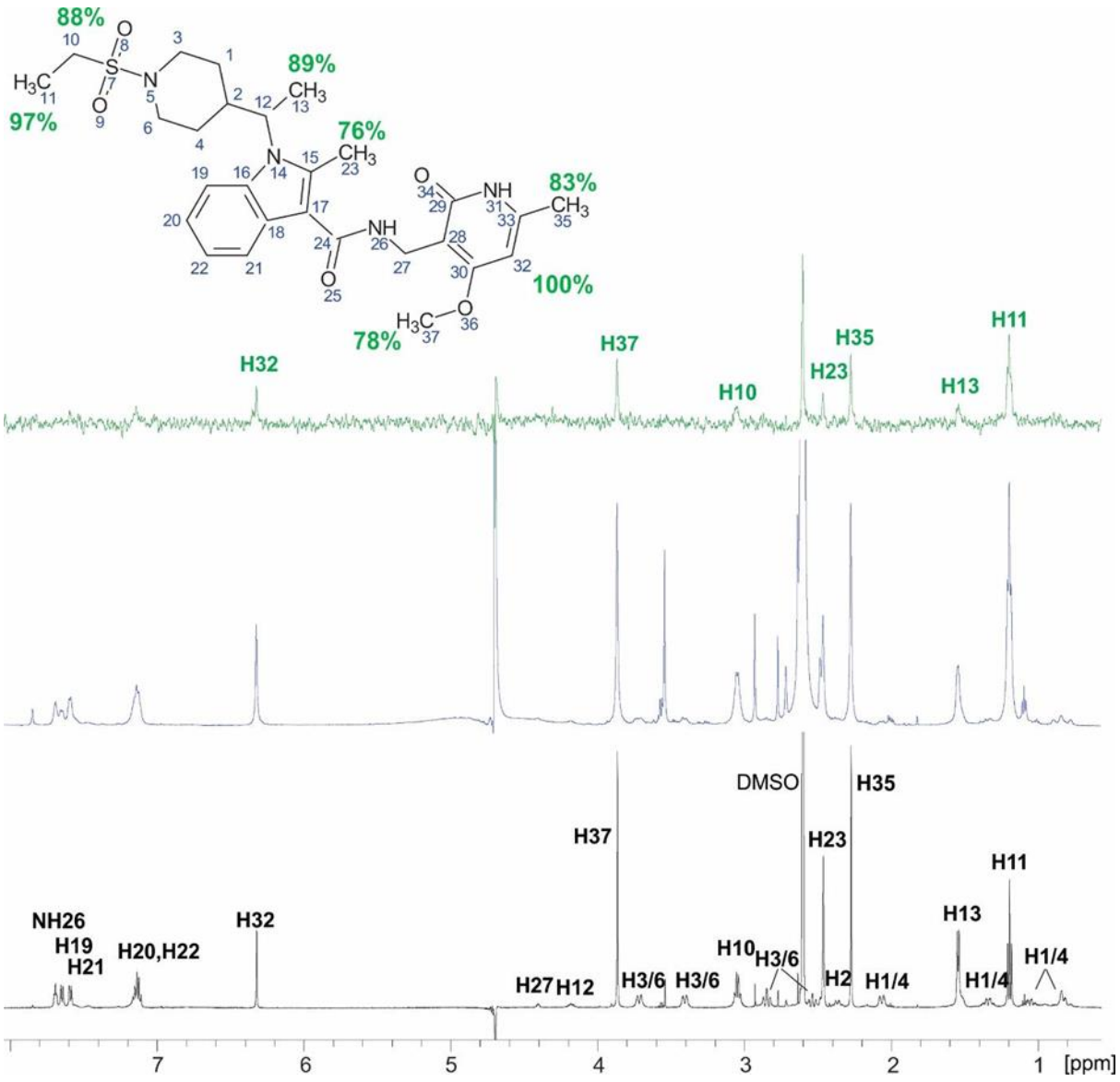


Figure 39: (In black) 1D 1H spectrum of CPI-169 showing assignment of proton chemical shifts. (In green) 1D 1H difference STD spectrum of 0.5 mM CPI-169 at a PLPro:ligand ratio of 1:100. The STD amplification factors were calculated for the signals with sufficient signal-to-noise ratio marked in green and normalized to the intensity of the signal with the largest STD effect. Above, the molecular structure illustrates the proton nomenclature and the relative degrees of saturation of the individual protons. (In blue) 1D 1H reference (off-resonance) STD spectrum.

The same sample was used for the 2D tr-NOESY experiment. The negative NOEs of CPI-169 delineated the molecule's conformation in the binding pocket compared to the distances measured in the docking pose. The intensities of NOE cross-peaks are consistent with the docking pose in the GRL-0617 binding site, except for the methoxide substituent of the pyridine group (H37) which has a different orientation, away from H21, H22 and H26 (Table 10).

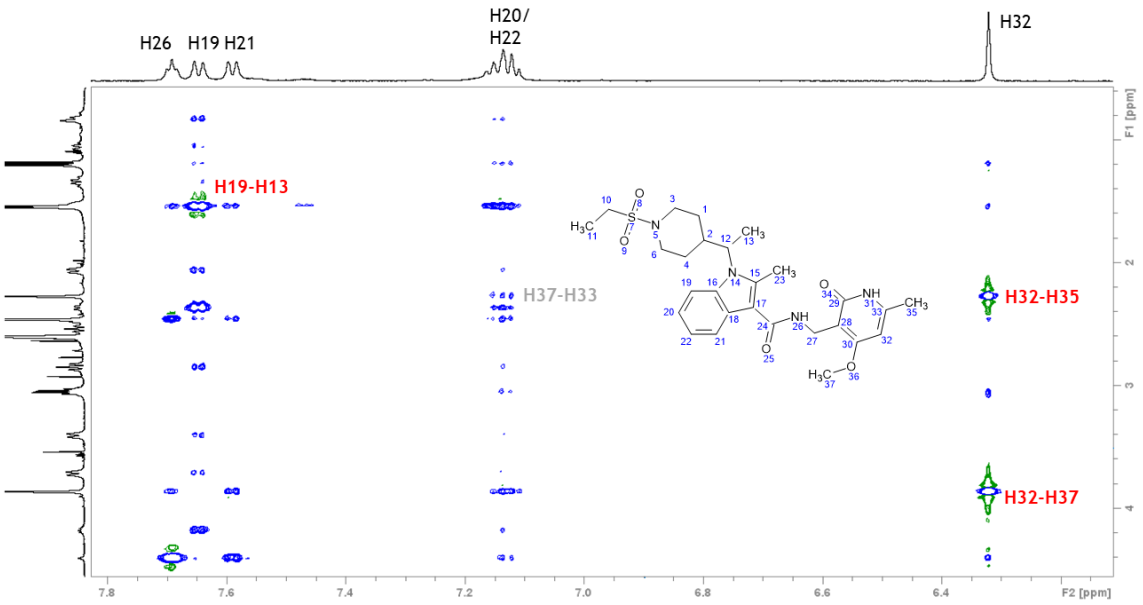


Figure 40: Expanded region of trNOESY of 0.5 mM CPI-169 at ligand:PLpro_NAB ratio of 100:1, showing the cross-peaks of aromatics with methyls. Weak intensity of the cross-peak between H37 and H22 is in grey.

Table 10: The non-trivial NOEs of 0.5 mM CPI-169 observed in the trNOESY spectrum at a ligand-PLpro ratio of 1:100 and corresponding distances from the model structure of CPI-169 in complex with PLpro_NAB. In red, strong NOEs showed in the figureXX; in grey, weak NOEs of the methoxide group that do not match with docking pose distances.

NOE cross-peak	Intensity of NOE cross-peak	Distance from the docking pose of CPI-169 in complex with PLpro (Å)
H13-H19	strong	3.3 (C13-H19)
H2-H19	strong	2.3
H23-NH26	weak	4.8 (C23-NH26)
H37-H32	strong	3.0 (C37-H32)
H12-H19	weak	5.0
H27-NH26	medium	2.6, 2.9
H27-H21	weak	3.6, 4.3
H21-H37	very weak	3.1 (H21-C37)
H22-H37	very weak	4.2 (H22-C37)
H26-H37	very weak	2.3 (H26-C37)
H3/6-H11	weak	4.0, 5.2 (C11-H3/6)
H23-H13	strong	3.9 (C23-C13)
H23-H1/4	medium	2.9, 3.7 (C23-H1/4)

CPI-169 vs GRL-0617 competition profiling

In the docking prediction, CPI-169 overlaps GRL-0617 in the binding pocket. Considering that the GRL-0167 has a 4/7 folds higher IC_{50} measured by activity assay (depending on buffer conditions), competition between the two compounds was suspected. The hypothesis was demonstrated by a competition assay performed by STD NMR.

The GRL-0617 was tested in the set conditions of the assay to ensure the feasibility of the STD assay with this compound. Methyl and aromatic proton signals appeared in the difference spectrum, in the same protein-ligand ratio of the CPI-169 (Figure 41).

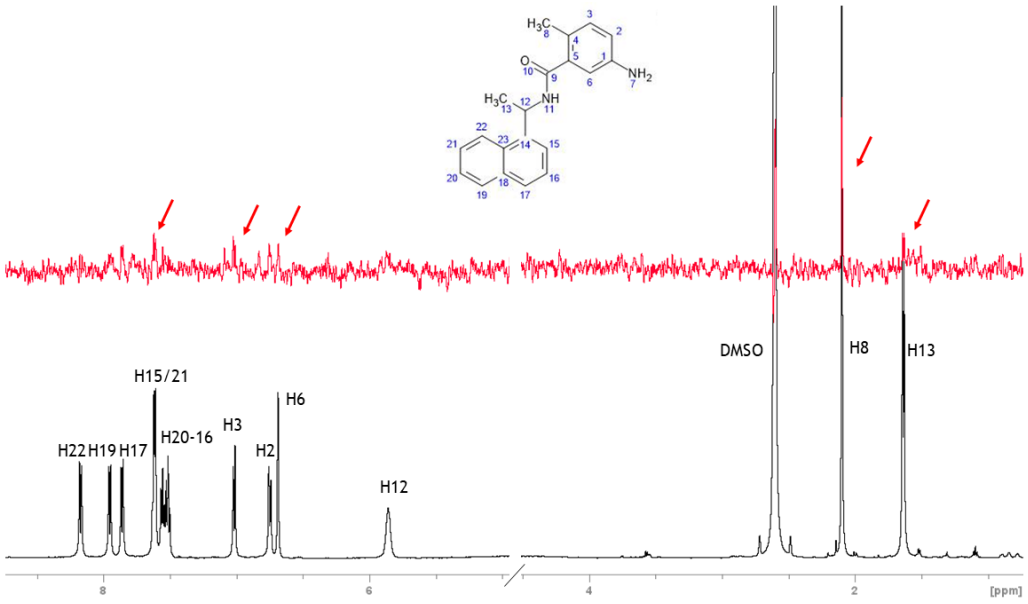


Figure 41: 1D STD NMR of 0.3 mM GRL-0617. In black, the reference spectrum with proton signal assignment; in red, the difference spectrum with pointed the most appreciable signal/noise ratio peaks.

The 1D STD NMR with CPI-169 was performed again at a lower concentration of ligand, 0.15 mM, taking into account the later addition of a double concentration GRL-0617 and keeping the concentration of DMSO at 5.5%, which was optimal for the stability of the protein. A_{STD} of CPI-169 were calculated for methyl protons as monitor signals in the assay, in the absence and presence of GRL-0617 in ratio 1:2, then compared. In the presence of the more potent compound, most of the A_{STD} of CPI-169 significantly decreased, thus verifying the competitive behaviour of the two compounds which bind the same allosteric pocket. Interestingly, the A_{STD} of CPI-169 H37 increased in the

presence of the GRL-0617, an occurrence coherent with the docking pose, where the pyridine moiety of CPI-169 does not overlap with the other known inhibitor but interacts with the outer portion of the binding pocket (Figure 42, Table 11).

A competitive trNOESY attempt was conducted using the same sample, the CPI-169 NOE connectivities were too low though. A second sample at different protein-ligand ratios was tested: PLpro_NAB at 3 μM, CPI-169 at 0.15 mM, and GRL-0617 at 0.3 mM (1:50:100). The recorded trNOESY was then compared to the trNOESY spectrum acquired at 0.3 mM of CPI-169 in the presence of 6 ⌈M PLpro_NAB (1:50). Also in this experiment, GRL-0617 distinctively displaces the CPI-169, since several negative NOEs of the weaker compound disappear in the presence of the GRL-0617, while the NOE connectivities corresponding to the pyridine moiety are still visible, in concert with the 1D competitive STD experiment (Figure 43). However, it was not possible to reproduce the trNOE data keeping the same concentration of CPI-169 in the trNOESY alone and the presence of the competitive compound, due to solubility problems in the assay buffer.

Table 11: A_{STD} of CPI-169 methyl protons in presence and absence of GRL-0617. In green, the CPI-169 H37 A_{STD} that increases in presence of the second compound.

CPI-169 ^{1}H	A_{STD}	A_{STD}' with GRL-0167	A_{STD}'/A_{STD}
H11	0,74	0.44	0.59
H13	1.19	0.81	0.68
H23	0.43	0.27	0.63
H35	0.89	0.55	0.62
H37	**0.10**	**0.21**	**2.1**

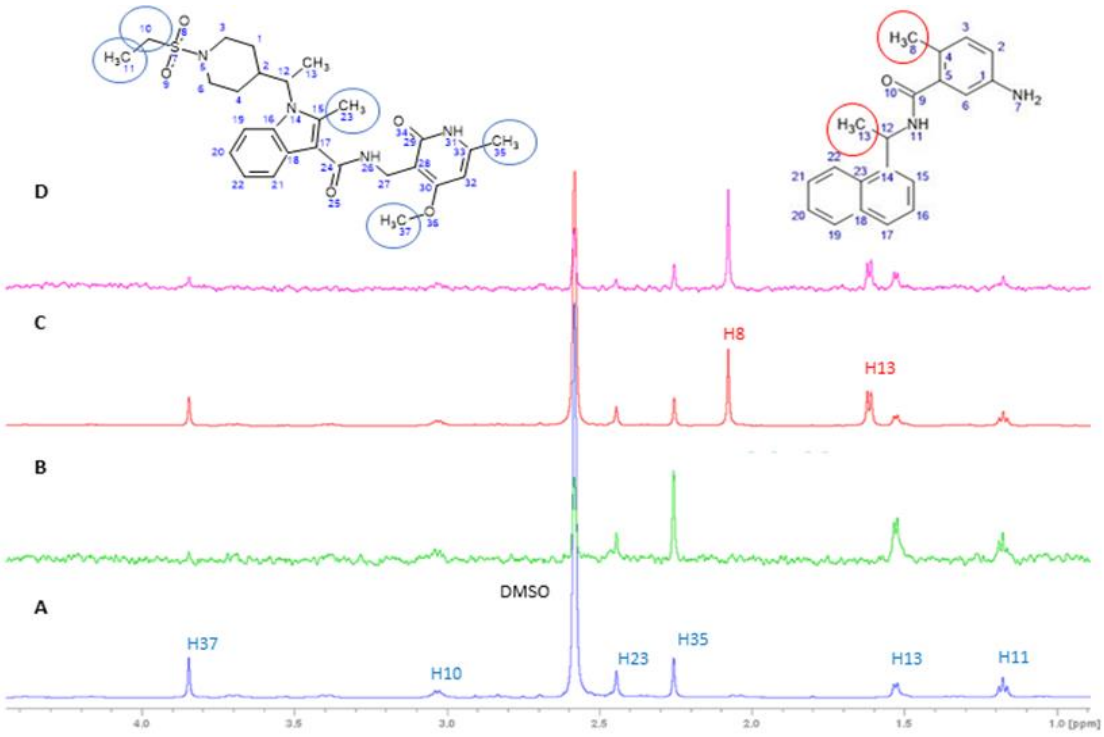


Figure 42: Expanded regions of 1D 1H STD spectra showing signals from methyl protons used to monitor competition between CPI-169 and GRL-0617. The proton signals from CPI -169 (structure on the left) and GRL-0617 (structure on the right) are labelled in blue and red, respectively. A: 1D 1H reference STD spectrum of 0.15 mM CPI-169 recorded at a PLpro_NAB: compound ratio of 1:100. B: 1D 1H STD difference spectrum of 0.15 mM CPI-169 recorded at a PLpro_NAB:CPI-169 ratio of 1:100. C: 1D 1H reference STD spectrum of 0.15 mM CPI-169 recorded at a PLpro_NAB: CPI-169 ratio of 1:100, in the presence of 0.3 mM GRL-0617. D: 1D 1H STD difference spectrum of 0.15 mM CPI-169 recorded at a PLpro_NAB:CPI-169 ratio of 1:100, in the presence of 0.3 mM GRL-0617.

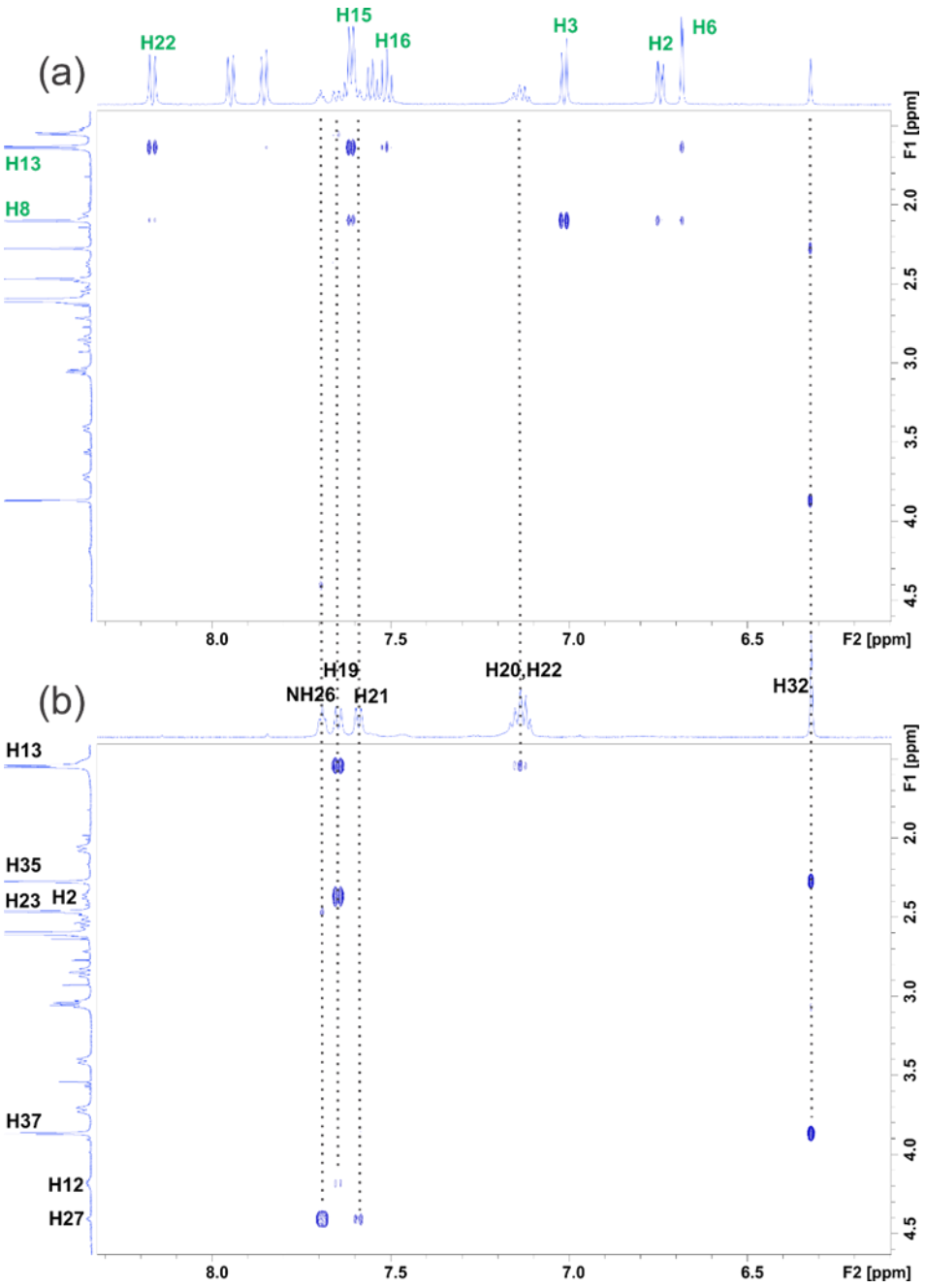


Figure 43: Expanded regions of trNOESY spectra indicating a displacement of CPI-169 from the GRL-0617 binding site. (a) Expanded region of the trNOESY spectrum of 0.15 mM CPI-169 at a ligand:PLPro ratio of 50:1 in the presence of 0.3 mM GRL-0617. The proton signals of GRL-0617 showing NOE connectivities are marked in green. The dotted lines indicate the positions of aromatic protons from CPI-169. Most NOEs from CPI-169 disappear. (b) Expanded region of the trNOESY spectrum of 0.3 mM CPI-169 at a ligand:PLPro ratio of 100:1. The proton signals of CPI-169 that exhibit NOE connectivities are marked in black.

5.1.8 Discussion

The PLpro branch of Exscalate4Cov required an intense involvement of our group, both from an experimental and networking point of view. PLpro and PLpro_NAB constructs were produced at fair yields and quality levels and provided to our partners for all the assays necessary for the drug repurposing objective. From the beginning of this study, a keen sensitivity of the protein towards oxidative conditions was observed, revealing this target particularly challenging in the development and optimization of the assays. The susceptibility of PLpro to redox agents was also observed by other groups, demonstrating that buffer conditions and the chemical nature of candidate inhibitors affect indeed the oxidative state of PLpro cysteines, concerning not only the catalytic one but also more superficial ones (i.e. Cys270 in the B2 loop) [46]. Evaluating the intrinsic activity and stability of the protein, the usage of reducing agents in the assay conditions appeared necessary, confirmed by the MS mapping of the Cys111 which showed a slight di- and tri-oxidation in the absence of reducing agents such as DTT. Unexpectedly, our observation generated false positives in the first round of the repurposing screening, where false hit-compounds seemed as potent as promising. Emerging published remarks [50,107], together with the disheartening results of TSA and crystallization trials, suggested an oxidation issue with the hits coming out from the first screening. The mechanism of action was then elucidated using a multi-technique approach: the hit-compounds, which all shared a quinone-like structure, were generating superoxide species catalysed by the strong reducing agents present in the assay, irreversibly oxidating in the end the reactive cysteines of the PLpro, above all the catalytic one (Cys111). Diketons, nitartes and hydroquinones are indeed considered Pan-Assay Interference Compounds (PAINs), showing an inhibitory activity in biochemical assays due to a non-specific interaction, often driven by oxidation. Their behaviour, depending also on the chemical environment, makes their confirmation demanding and, consequently, becomes difficult to define their benefit as scaffolds for compound optimization [115]. The establishment of such compounds as real inhibitors or false positives in a fluorescence-based activity assay is crucial, therefore the biophysical tools resulted to be essential to discriminate the real inhibitors. A critical revision of the screening was then conducted and new hits were discovered. The selection of the most promising one, considering selectivity and activity independence from the reducing agent in the assay, led to the revelation of a new interesting inhibitor, the CPI-169. The validation and mechanism of action of this compound were elucidated through a combination of computational and biophysical tools, which harmoniously showed that the CPI-169 shares the same binding site as the known repurposing compound GRL-0617, the pocket recognizing the P2 site of the substrate. Even though a co-structure was not achieved, the ligand-based NMR

techniques allowed the verification of the docking predictions, revealing a novel repurposing compound with more interactions with PLpro residues than the known one. CPI-169 showed a calculated IC_{50} around 15 μM in the *in vitro* activity assay. Considering that the GRL-0617 resulted to be active in cell-based assays around 50 μM and that CPI-169 is ca. 7 times less potent than GRL in our assay, we expect that the effective concentration of CPI-169 in an antiviral assay could be quite high. Even though CPI-169 could not be potent enough for its development, we firmly believe it could be considered a promising scaffold for a rationally improved designed molecule, like its competitor, the GRL-0617.

5.2 PLpro-ISG15 complex

5.2.1 ISG15 constructs design, cloning and expression

Small-scale test expression of constructs from GenScript were performed. Despite all the conditions tested, the expression rate was low. An appreciable expression was obtained inducing at 37°C for 3h, in Rosetta2(DE3), confirmed by a Western blot using anti-His antibody (data not shown). However, there was no condition where the proteins could be extracted in decent amounts (Figure 44).

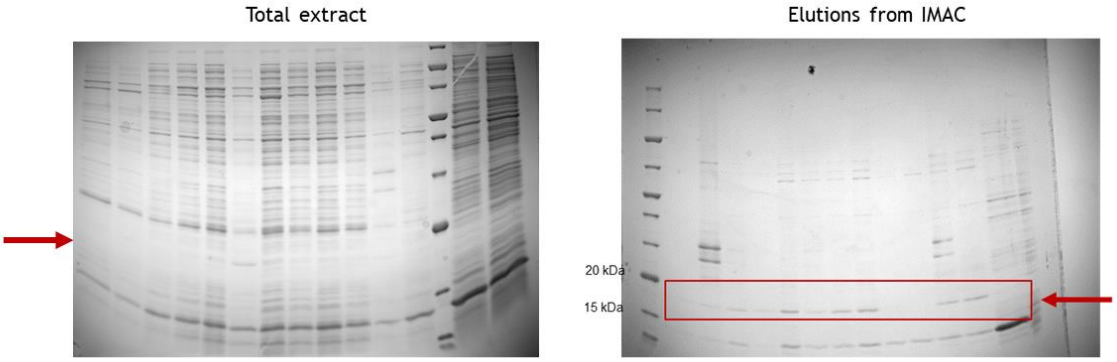


Figure 44: Test expression of (6x)His-proISG15 in pET21d. On the left, the SDS-PAGE of total extracts from cultures in TB, 37°C, at [IPTG] of 0.1, 0.3, 0.5, and 1 mM, testing bacterial strains BL21(DE3), Rosetta2(DE3), LEMO21(DE3). On the right, the SDS-PAGE with the corresponding elutions. Red pointers indicate the bands of proISG15.

Since the expression yield was too low, the genes coding for the ISG15 and proISG15 were cloned in other expression vectors to explore different tags: N-terminal 6xHis-TrxT, 6xHis-ZB and 6xHis-GST were selected. The Ligation-Independent cloning led to gaining the following constructs, verified by sequencing:

N-terminal tag	$ISG15^{C78S}$	$proISG15^{C78S}$
6xHis-TrxT	positive	positive
6xHis-ZB	positive	positive
6xHis-GST	positive	**negative**

A new small-scale test expression of the resulting construct was conducted, showing a remarkable improvement in the expression yield for any of those (Figure 45A). Despite the good expression rate, the 6xHis-TrxT constructs were less stable in the soluble fraction than those with other tags. 6xHis-ZB_proISG15 and 6xHis-GST_ISG15 were then selected to be produced on larger scale, in BL21(DE3) strain grown in TB medium. Temperature of induction proved to be decisive in the expression, proteins were indeed produced only at 37°C with a better yield. The expression of 6xHis-GST_ISG15 was further improved by co-expression of GroE chaperon complex, keeping the same expression conditions as the previous optimization.

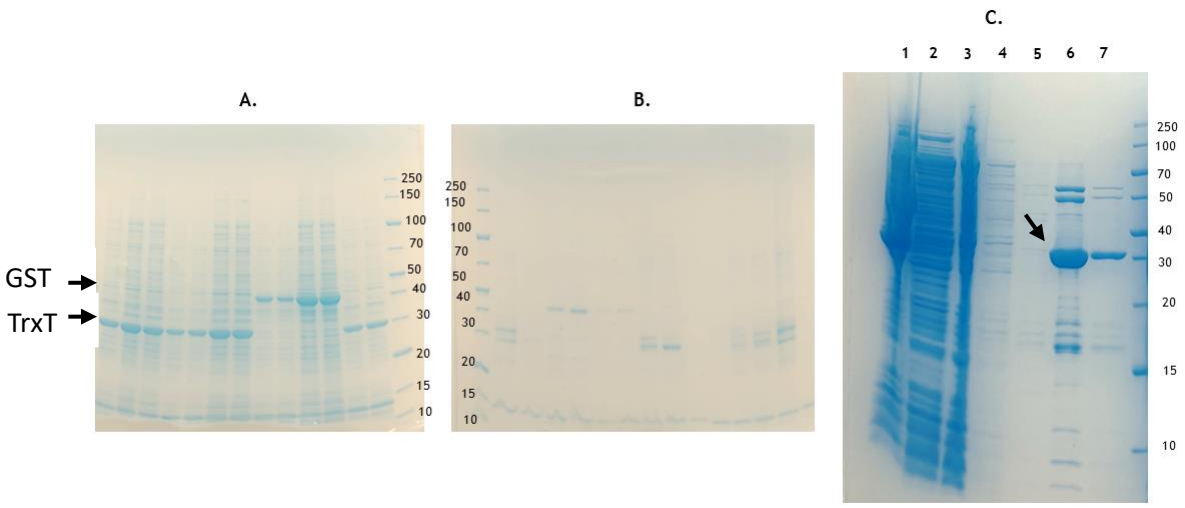


Figure 45: Expression optimization of 6xHis-GST_ISG15. A. Supernatants of small-scale test expression of 6xHis-TrxT_ISG15 and 6xHis-GST_ISG15, in TB or LB medium with IPTG concentration of 0.1 or 0.5 mM; B. IMAC elutions from small-scale test expression; C. 1L co-expression of 6xHis-GST_ISG15 in BL21(DE3) with GroE complex: 1-total extract, 2-supernatant, 3-Flowthrough IMAC, 4-Wash1 IMAC, 5-Wash2 IMAC, 6-Elution1 IMAC, 7-Elution2 IMAC.

The scale-up expression was therefore performed in BL21(DE3) strain in TB medium, co-expressing the mature form with chaperons GroEL and GroES. The first affinity chromatography step was performed in an FPLC system (Äkta pure, Cytiva) loading on a 5 mL HisTrap FF column, since the elution in gradient of the imidazole-concentrated buffer led to obtaining a better separation of tagged ISG15 and proISG15 from the bacterial chaperons. After the tag removal, a SEC step was applied to eliminate impurities present in the samples. The final product reached a good purity grade and was then subjected to characterization and quality control. Despite the computational tools' indication of the recombinant product stability (ProtParam, ExPASy), which predicted the protein as “unstable” from the primary sequence, the TSA showed quite good stability of the products in all the buffers tested (Table 13). The HPLC-SLS analysis confirmed the good yield of purity, showing well-shaped peaks with a measured MW very close to the theoretical one, proving a monodisperse monomeric sample (Figure 47, Table 12).

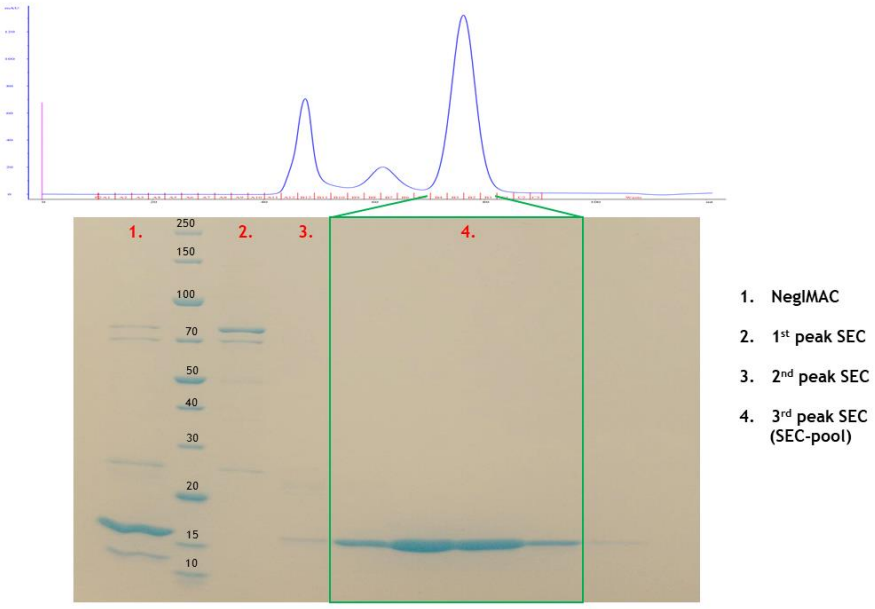


Figure 46: Last step of purification of proISG15 with SEC chromatogram

Table 12: Parameters measured by HPLC-SLS of produced proteins. Measured MWs deviate from the theoretical ones <1 kDa.

Construct	MW (kDa)	MW/Mn	RV (mL)	[measured]/[injected]
ISG15 $C_{78}S$	16.67	1.002	8.77	0.95
proISG15 $C_{78}S$	17.97	1.006	9.83	0.91

Table 13: Average Tm of recombinant ISG15 and proISG15 produced

Construct	**Storage buffer +DTT 1mM**	**Storage buffer -DTT**	**PBS 1x +DTT 1 mM**
ISG15 $C_{78}S$	53.7 ± 0.6	54.0 ± 0.0	54.0 ± 0.0
proISG15 $C_{78}S$	54.5 ± 0.5	55.0 ± 0.0	54.0 ± 0.0

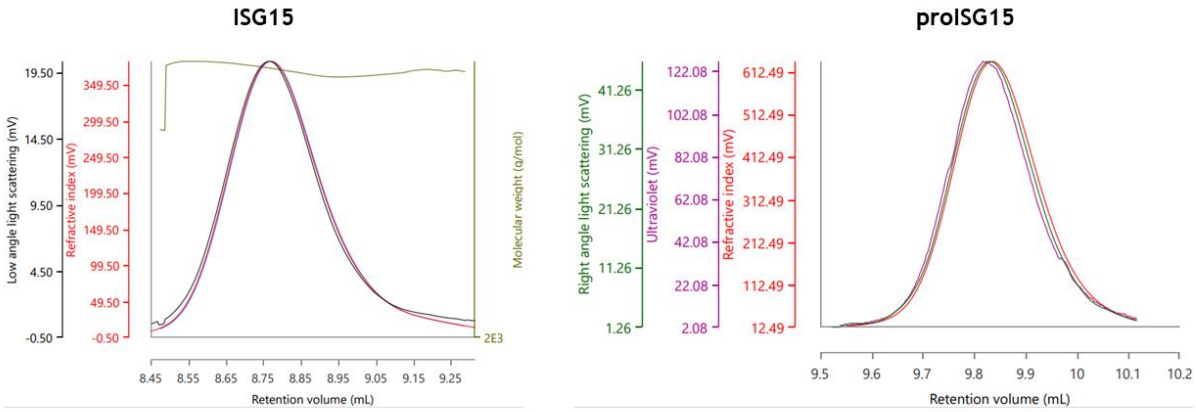


Figure 47: Derived data from HPLC-SLS experiments of ISG15 constructs.

5.2.2.1 Preliminary characterization of the complex

In order to gain preliminary information about the PLpro-ISG15 complex formation, pull-down and HPLC coupled with SLS were performed.

The protein pull-down, performed at the 6xHis-PLpro:ISG15 ratio 1:1, suggested a feeble interaction between the PLpro and ISG15. The majority of ISG15 was not present in the eluate of the IMAC with the tagged PLpro but was eluted in the flow-through or the wash in binding buffer (Figure 48). The proISG15 was clearly cleaved by the two WT constructs but not by the inactive mutant $C_{111}S$, indicating a recognition with the substrate. Interestingly, a slight portion of the uncleaved proISG15 is visible in the elution with the tagged PLpro_NAB $C_{111}S$ (Figure 48.3).

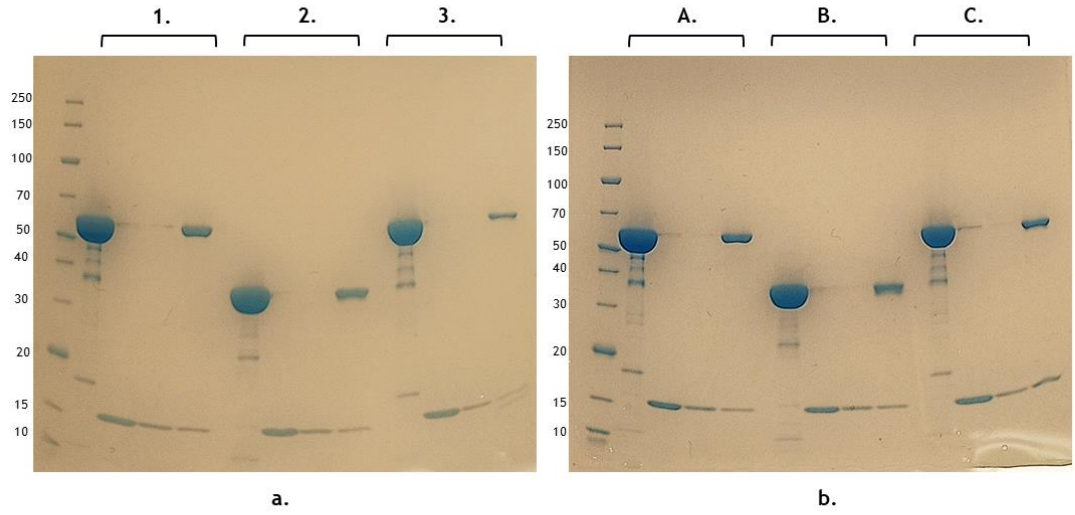
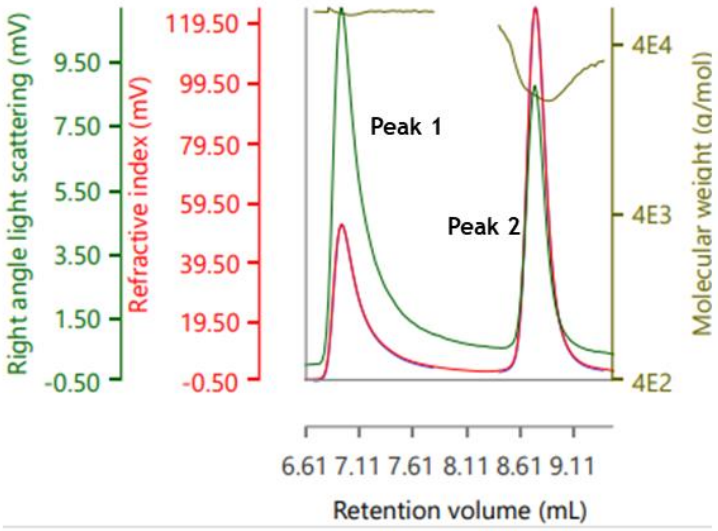


Figure 48: SDS-PAGE 16% of the protein pull-down; in each group, 5 μg of tagged PLpro precedes the flow-through, wash in binding buffer and elution from IMAC: 1. PLpro_NAB WT + proISG15; 2. PLpro WT + proISG15; 3. PLpro_NAB $C_{111}S$ + pro ISG15; A. PLpro_NAB WT + ISG15; B. PLpro WT + ISG15; C. PLpro_NAB $C_{111}S$ + ISG15

A further investigation was conducted in HPLC-SLS to isolate a peak of the complex and have a measure of the MW. Different ratios of the PLpro constructs and substrate were tested, increasing the concentration of the ISG15 or proISG15 (1:1, 1:3, 1:5, 1:7). Different running buffer conditions were also explored, changing the pH of the solution, salts concentration or the chemistry of the buffer. In any case, it was not possible to observe a peak in the chromatogram corresponding to the complex, regardless of the construct of the PLpro or of the substrate used for the experiment.



Peak	MW (kDa)	MW/Mn	RV
Peak 1	55.10	1.000	6.96
Peak 2	18.75	1.022	8.76

Figure 49: Derived data from HPLC-SLS analysis of the complex PLpro_NAB/ISG15 1:3. Chromatogram and measured MW from light scattering show only the separated proteins' peaks and masses.

5.2.2.2 Affinity and Kinetics Analysis

For the investigation of the binding affinity and kinetic parameters of the PLpro/ISG15 complex, GCI was performed in collaboration with Dr Sonia Covaceuszach (IC-CNR, Trieste). 6xHis-PLpro constructs were captured on the NiNTA functionalized chip using the tag. First measurements were conducted on the WT constructs, injecting a series of dilutions 1:2 of the analytes (ISG15 and proISG15) from 10 μM. In both cases, the affinity was quite low for protein-protein interaction, with a mild major affinity for the precursor of the ISG15 (Table 14). The results agree with the reported cellular data, showing that ISG15 is the preferred substrate of the PLpro, but is not sequestered by the protease [59].

Table 14: Kinetic and affinity constants obtained for PLpro and PLpro_NAB WT by GCI experiments.

Ligand	Analyte	Kon ($M^{-1}s^{-1}$)	Koff (s^{-1})	K_D (M)
PLpro_NAB	ISG15	8.25±0.51 E^{3}	9.45±0.14 E^{-2}	11.52±0.92 E^{-6}
PLpro_NAB	proISG15	3.19 ±0.09 E^{4}	9.68±0.22 E^{-2}	3.03± 0.24 E^{-6}
PLpro	ISG15	1.16±0.11 E^{4}	8.14±0.81 E^{-2}	7.02±0.82 E^{-6}
PLpro	proISG15	1.07±0.08 E^{3}	5.33±0.86 E^{-2}	4.5±0.86 E^{-6}

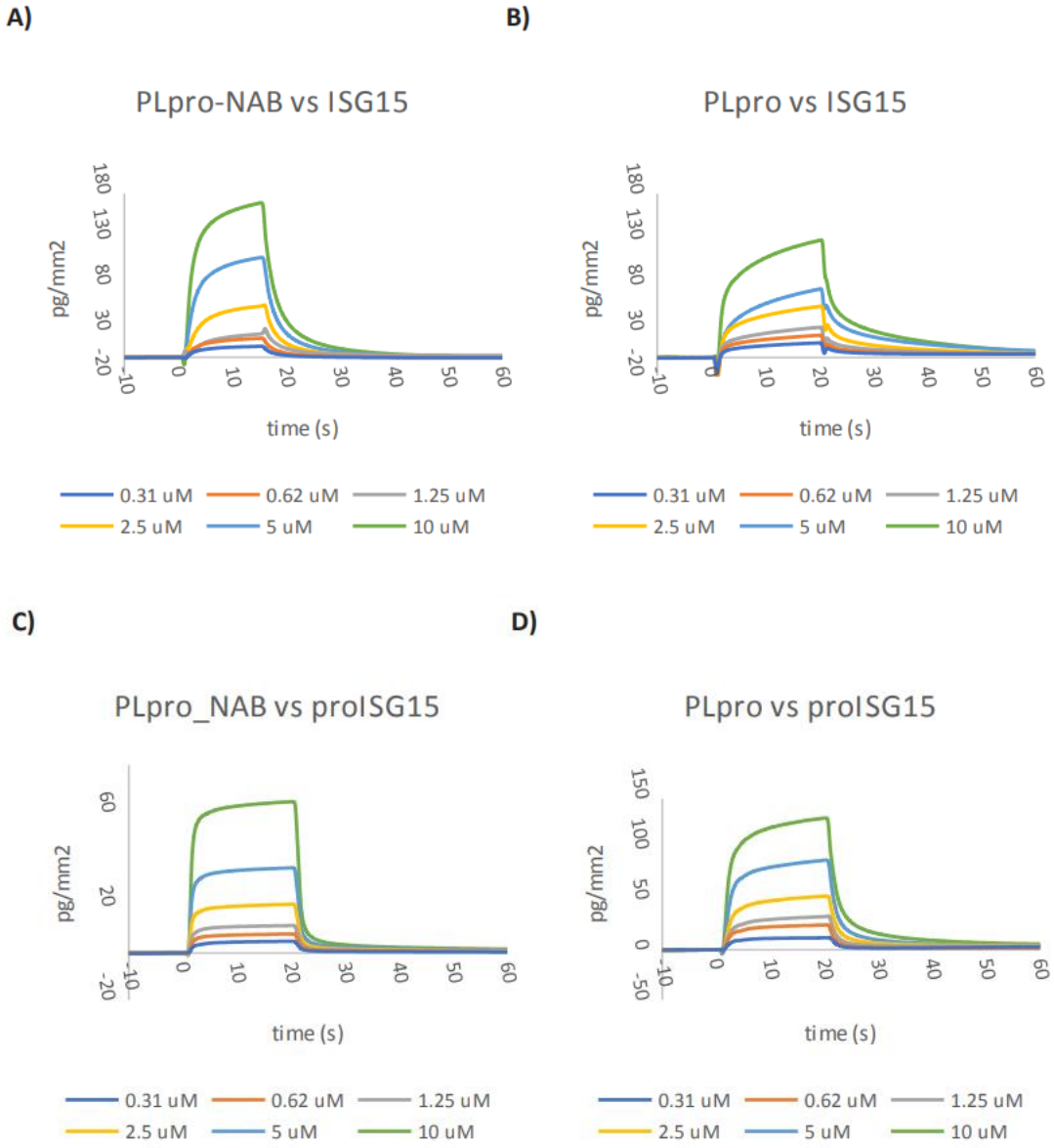


Figure 50: Quantitative binding kinetics for PLpro_NAB and PLpro by GCI: A) PLpro_NAB versus ISG15.Response curves of 1:2 serial dilutions of ISG15 (10uM-0.31uM range) on a cell coated with 3000 pg/mm² of PLpro_NAB. B) PLpro versus ISG15. Response curves of 1:2 serial dilutions of ISG15 (10uM-0.31uMrange) on a cell coated with 3000 pg/mm² of PLpro. C) PLpro_NAB versus proISG15. Response curves of 1:2 serial dilutions of proISG15 (10uM-0.31uM range) on a cell coated with 8000 pg/mm2 of PLpro_NAB. D) PLpro versus proISG15. Response curves of 1:2 serial dilutions of proISG15 (10uM-0.31uM range) on a cell coated with 3000 pg/mm² of PLpro. All curves were blank subtracted. Experiments performed in triplicate.

The experiments were replicated using the mutants of the PLpro and PLpro_NAB. This time, the reversible capture of the His-tag was not sufficient to perform the experiments in triplicate, so a subsequent amine coupling resulted to be necessary to immobilize the protein onto the chip. An attempt of immobilization of the tagged ISG15 was tried. Even though the immobilization succeeded, no interaction with the PLpro was detected. This failure confirmed the importance of the ISG15 N-terminus in the binding, in concert with the reported evidence of binding analysis [32,66]. The mutation of the catalytic cysteine into serine seems to decrease the affinity, even if the recognition of the substrate does not occur in the catalytic pocket (Table 15). An intriguing result was obtained with the two δ-variant mutants, especially with the $T_{467}K$ mutant. While the $S_{466}R$ mutant shows parameters similar to the WT, the other mutant presents a dramatic drop in affinity with the proISG15 (Table 16).

Table 15: Kinetic and affinity constants obtained for PLpro and PLpro_NAB $C_{111}S$ by GCI experiments. Curves of PLpro $C_{111}S$ with ISG15 were too noisy to calculate parameters.

Ligand	Analyte	K_{on} ($M^{-1}s^{-1}$)	K_{off} (s^{-1})	K_D (M)
PLpro_NAB	ISG15	1.36±0.38 E^3	4.61±0.08 E^{-2}	33.77±0.39 E^{-6}
PLpro_NAB	proISG15	2.14±0.71 E^3	9.52±1.13 E^{-2}	44.40±3.17 E^{-6}
PLpro	proISG15	2.56±0.73 E^3	4.69±0.33 E^{-2}	18.35±0.8 E^{-6}

Table 16: Kinetic and affinity constants obtained for PLpro_NAB $S_{466}R$ and $T_{467}K$ by GCI experiments. The PLpro_NAB $T_{467}K$ affinity was too low to be correctly calculated by the software.

Ligand	Analyte	Kon ($M^{-1}s^{-1}$)	Koff (s^{-1})	K_D (M)
$S_{466}R$	ISG15	1.55±0.38 E^4	6.53±1.14 E^{-2}	4.20±1.2 E^{-6}
$S_{466}R$	proISG15	8.33 ±0.91 E^3	1.03±0.06 E^{-1}	12.06± 0.91 E^{-6}
$T_{467}K$	ISG15	8.35±0.51 E^3	9.03±0.33 E^{-2}	10.80±0.60 E^{-6}
$T_{467}K$	proISG15	ND	ND	ND

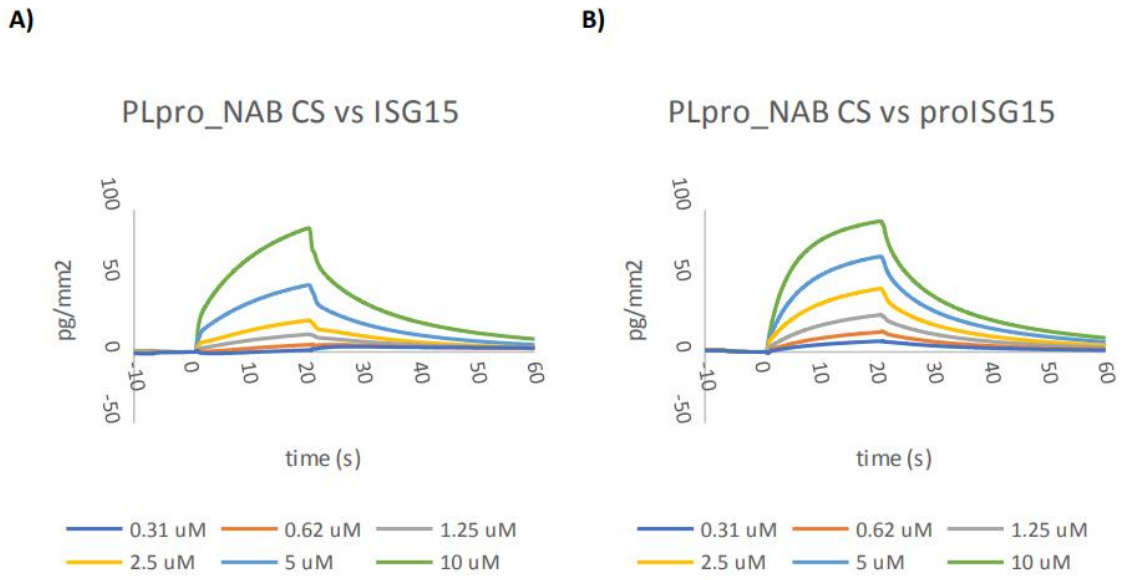
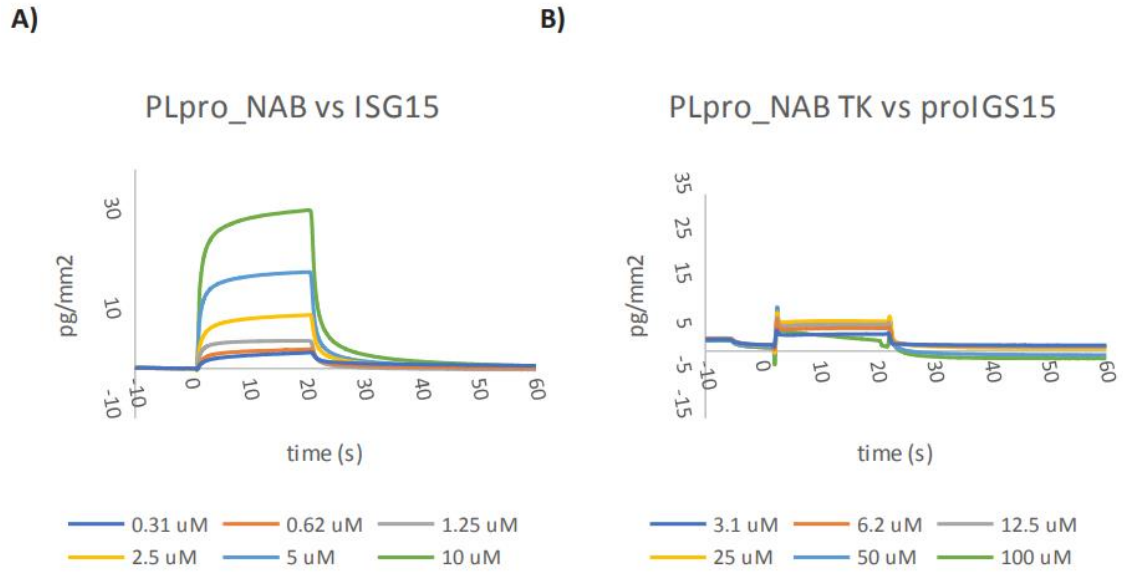


Figure 51: Quantitative binding kinetics for PLpro_NAB CS by GCI: A) Response curves of 1:2 serial dilutions of ISG15 (10uM-0.31uM range) on a cell coated with 5600 pg/mm² of PLpro_NAB $C_{111}S$. B) Response curves of 1:2 serial dilutions of proISG15 (10uM-0.31uM range) in a cell coated with 5600 pg/mm² of PLpro_NAB CS.



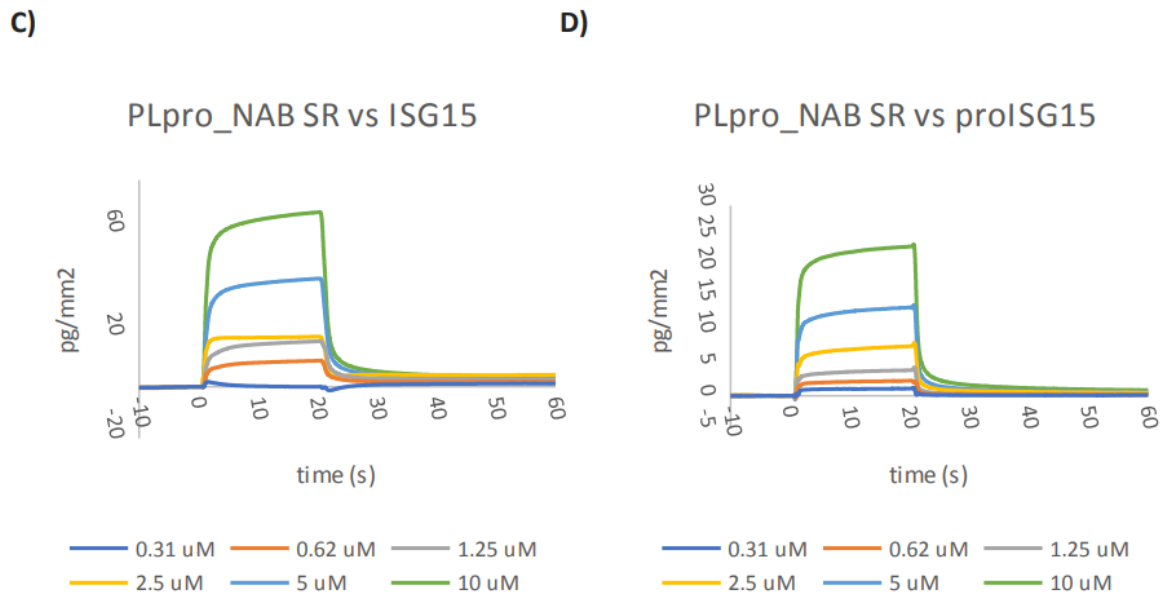


Figure 52: Quantitative binding kinetics for PLpro_NAB mutants by GCI: A) PLpro_NAB $T_{467}K$ versus ISG15. Response curves of 1:2 serial dilutions of ISG15 (10uM-0.31uM range) on a cell coated with 4000 pg/mm² of PLpro_NAB $T_{467}K$. B) PLpro_NAB $T_{467}K$ versus proISG15. Response curves of 1:2 serial dilutions of proISG15 (100uM-3.1uM range) on a cell coated with 4000 pg/mm² of PLpro_NAB $T_{467}K$. C) PLpro_NAB $S_{466}R$ versus ISG15. Response curves of 1:2 serial dilutions of ISG15 (10uM-0.31uM range) on a cell coated with 7600 pg/mm² of PLpro_NAB SR. D) PLpro_NAB $S_{466}R$ versus proISG15. Response curves of 1:2 serial dilutions of proISG15 (25uM-0.31uM range) on a cell coated with 2900 pg/mm² of PLpro_NAB $S_{466}R$. All curves were blank subtracted.

A probable reason why the $T_{467}K$ mutant binds the proISG15 with such a lower affinity could be related to a minor activity of the enzyme. A qualitative cleaving activity test was performed. The test, conducted in ice bath to slow the reaction as much as possible, showed an evident difference between PLpro_NAB $T_{467}K$ and the WT only at the shortest time (5 s). The band of the proISG15 is indeed much more evident for the mutant than the WT, no visible difference is deductible for the other δ-variant mutant (Figure 53).

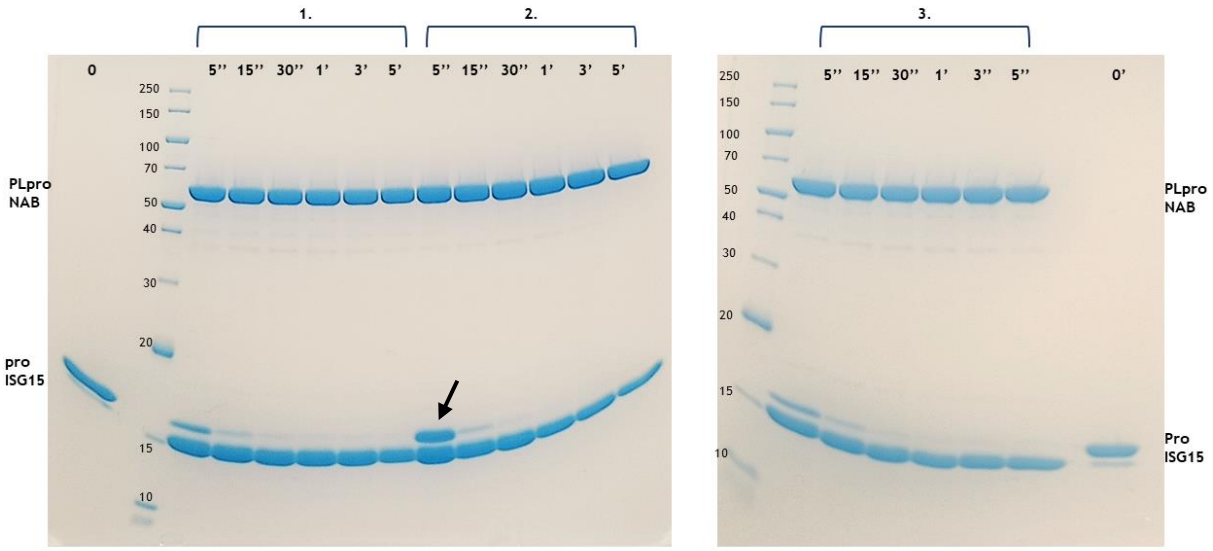


Figure 53: SDS-PAGE (4-20%) of the cleaving activity test. 1. PLpro_NAB WT + pro ISG15 1:1; 2. PLpro_NAB $T_{467}K$ + proISG15 1:1; 3. PLpro_NAB $S_{466}R$ + proISG15 1:1. The pointer highlights the cleavage of the mutant at 5 s.

Activity kinetic was studied in the enzymatic assay at the Fraunhofer ITMP in Hamburg. To reach an appreciable saturation curve, our colleagues needed to increase the concentration of the enzyme from 1 nM to 10 nM, keeping the same range of substrate (ISG15-AMC) concentration. The results show an increased Km for δ-variant mutants compared to the WT, which is even higher for the $T_{467}K$ mutant, matching the observations deriving from the qualitative experiment and binding analysis at GCI (Table 17).

Table 17: Kinetic parameters from the enzymatic activity assay.

	WT 1 nM	**$S_{466}R$ 10 nM**	**$T_{467}K$ 10 nM**
Vmax	34186	410003	390874
Km µM	0.391	1.908	1.985

5.2.2.3 Small angle X-ray scattering

In order to study the PLpro_NAB structure at low resolution considering the failure of crystallization trials and to assess whether the single point mutations impact the conformation of PLpro in solution, SAXS experiments were performed on the WT and δ-variant mutants at different concentrations (Table 18) without observing systematic changes due to solute concentrations, then analyzed by Dr Sonia Covaceuszach (IC-CNR, Trieste). Figure 54-A displays the processed scattering data collected for the highest concentration of each protein, and Table 18 compares the resulting overall size.

The Vp and the MM calculated for all mutants (Table 18) were consistent with values expected for a monomeric species even at relatively high (up to 5 mg/ml) concentrations and are in agreement with the MM estimated and measured by MS.

The computed distance distribution functions p(r) are compared to the profile obtained for the WT protein (Figure 54-B), showing a single asymmetric peak with a tail, a pattern indicative of proteins having elongated structures and a significant change in the maximum dimensions of the $S_{466}R$ mutant (from 130±6Å for WT to 153±8Å for $S_{466}R$), combined with a variation in the radius of gyration Rg (from 41.5±2.0 Å for the WT to 43.2±1.3 Å for $S_{466}R$), suggesting a more extended shape of the mutant compared to the WT protein. On the contrary, the single mutant $T_{467}K$ is quite similar to the WT both in Dmax (130±6Å) and in the Rg value (41.4±1.1 Å).

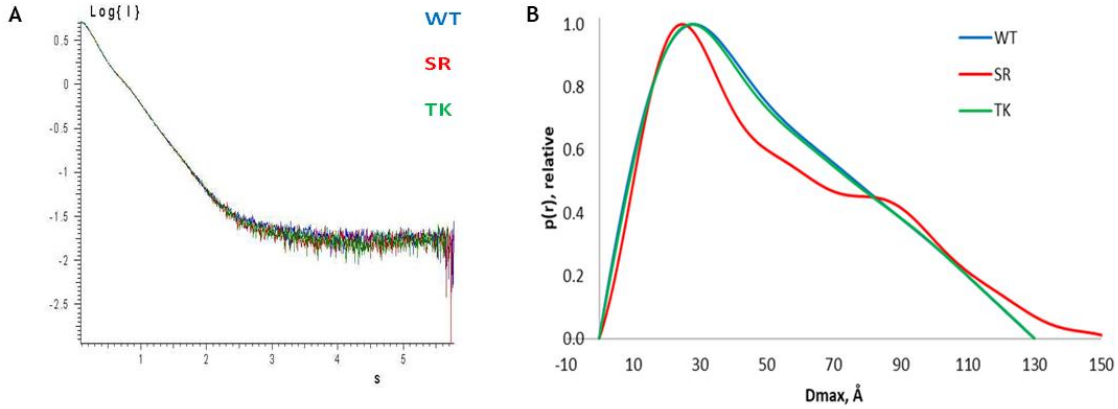


Figure 54: A) Comparison of the experimental X-ray scattering data of the PLpro_NAB WT, $S_{466}R$ and $T_{467}K$ mutants. The plots display the logarithm of the scattering intensity as a function of momentum transfer. B) Comparison of the distance distribution functions obtained for the PLpro_NAB WT and mutants.

The macromolecular shapes of the WT and two mutants in solution were reconstructed in parallel by *ab initio* modelling and by rigid-body modelling.

All the 20 independent solutions for each mutant, reconstructed from the X-ray scattering data using DAMMIF [88], showed a good fit to the experimental curves (χ^2= 0.507±0.003, χ^2= 0.0236±0.0005 and χ^2= 0.506±0.005, for WT, SR and TK mutants, respectively), with a resolution of 33 Å and averaged to obtain the final low-resolution models of the three proteins with quite low NSD values (0.704±0.027, 0.690±0.040 and 0.845±0.089 for WT, SR and TK respectively). The comparison of the resulting *ab initio* models with that of the WT suggests that small but significant rearrangements (at low resolution) in the orientation of the two domains occur, leading to a more extended structure for the $S_{466}R$ mutant (Figure 55).

In order to get more detailed information exploiting the respective high-resolution crystal structure of the PLpro (PDB 6XAA) and the NAB domain (PDB 7LGO), with both domains treated as rigid bodies, the program CORAL [92] was used to optimize the relative orientations of the two domains and to reconstruct the missing regions. The resulting models obtained with P1 symmetry (in Fig XW, with χ^2= 0.18±0.03 and NSD = 0.949±0.083, χ^2= 0.14±0.02 and NSD = 0.912±0.076, χ^2= 0.14±0.02 and NSD = 0.888±0.072 for WT, SR and TK mutants, respectively) provide a representation of the average conformations. It is interesting to note that, comparing the CORAL models of the two mutants to that of the WT one (Figure 57), the PLpro appears to be rotated to

different extents with respect to the NAB domain. Indeed, while in the $T_{467}K$ mutant this domain is slightly rotated towards the NAB domain, in the case of the $S_{466}R$ mutant the rotation is around 180°.

The inter-domain flexibility was investigated using two different approaches. The Kratky plot allows to identify interdomain flexibility at a qualitative level. In such a plot, folded compact globular proteins provide a bell-shaped curve at low angles, while deviations from this behaviour point to particle flexibility as in the case of the PLpro constructs (Figure 57-A). A second more quantitative approach, the ensemble optimization method (EOM) [94] was employed in order to analyze inter-domain flexibility and size distribution of possible multiple configurations in solution and to obtain typical optimized ensembles with a good fitting to the experimental scattering data (χ^2= 0.040, χ^2=0.033 and χ^2= 0.035, for WT, SR and TK mutants, respectively). The EOM analysis plots the Rg of the structures forming the initial random pool and the selected ensembles. The Rg distributions of these ensembles (Figure 57-B solid lines) are nearly as broad as the distribution of randomly generated models (Figure 57-B dashed lines) supporting the hypothesis of considerable inter-domain flexibility. Notably, the mutants display alterations in the roughly bimodal profile that characterizes the distribution of size observed for the WT protein. Indeed, these proteins are characterized by a high degree of conformational plasticity considering that an almost continuous distribution of the selected ensembles is observed that covers a slightly narrower range with respect to the distribution of randomly generated models (Figure 57-B).

Table 18: SAXS structural parameters: radius of gyration (Rg), maximum dimension (Dmax) obtained from the p(r) distribution using GNOM; Porod volume (Vp) and molecular mass (MM) in Da.

Data collection parameters	**PLpro_NAB WT**	**PLpro_NAB $S_{466}R$**	**PLpro_NAB $T_{467}K$**
Concentration (mg/mL)	5.0-0.3 mg/mL	5.0-0.3 mg/mL	5.0-0.3 mg/mL
Temperature (K)	283	283	283
Structural parameters			
I(0) (A.U.) [from p(r)]	5.19±0.16	4.68±0.11	5.05±0.10
Rg (Å) [from p(r)]	41.7±2.0	43.5±1.3	41.5±1.1
I(0) (A.U.) [from Guinier]	5.19±0.16	4.68±0.11	5.05±0.10
Rg (A) [from Guinier]	41.5±2.0	43.2±1.3	41.4±1.1
Dmax (Å)	130±6	153±8	130±6
V_p estimate ($Å^3$)	100986±5000	84053±4200	84744±4200
MM [from V_p] (Da)	59400±6000	49445±5000	49850±5000
Calculated MM (Da)	55125	54554	54512
Software employed			
Primary data reduction	PRIMUS	PRIMUS	PRIMUS
Data processing	GNOM	GNOM	GNOM
Ab initio modelling	DAMMIF	DAMMIF	DAMMIF
Rigid body modelling	CORAL/EOM	CORAL/EOM	CORAL/EOM
Validation and averaging	DAMAVER	DAMAVER	DAMAVER
3D graphic representations	PYMOL	PYMOL	PYMOL

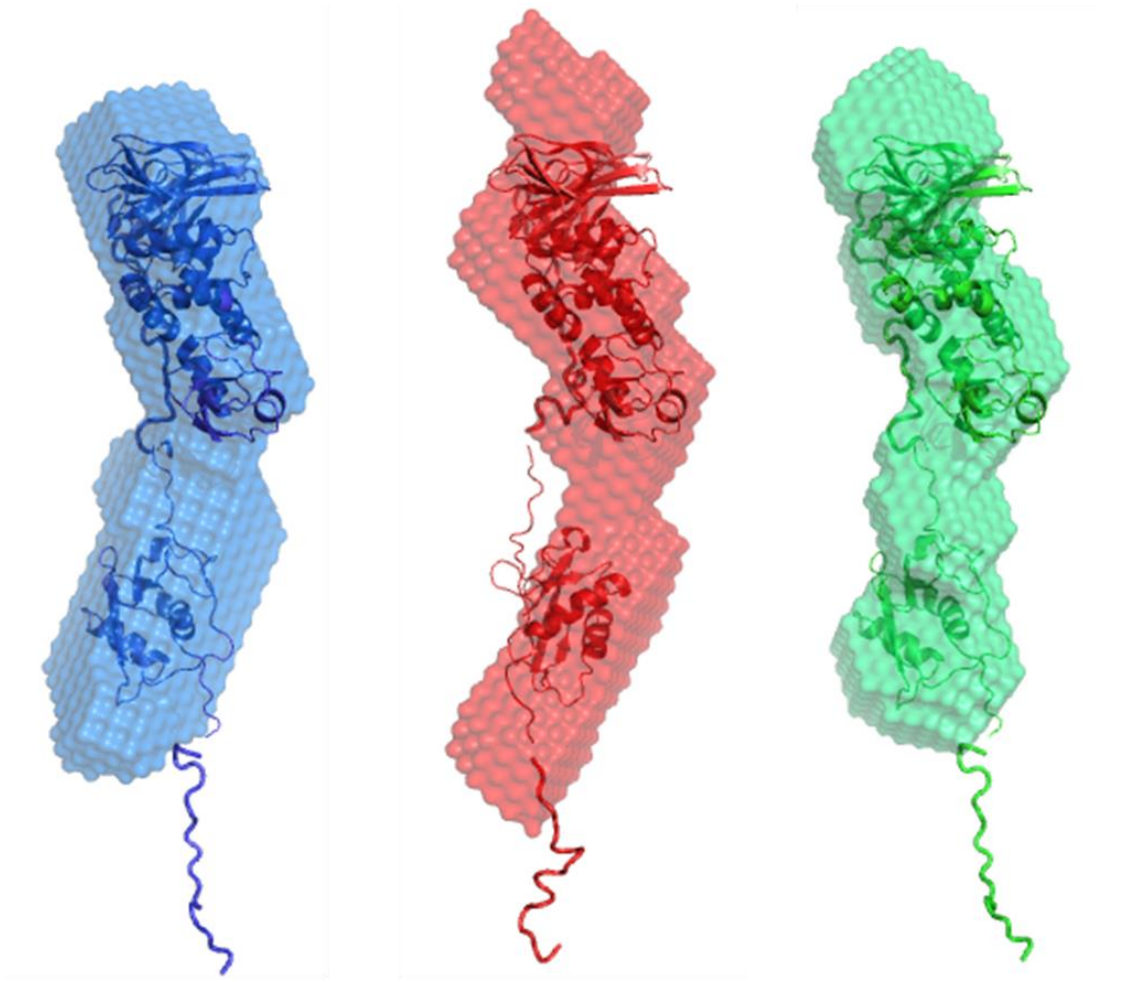


Figure 55: Comparison of the structural models of PLpro WT, $S_{466}R$ and TK mutants. Averaged and filtered ab initio bead models obtained with DAMMIF (semi-transparent surfaces) overlaid to the typical CORAL models (blue, red and green for WT, $S_{466}R$ and $T_{467}K$, cartoon representations for the PLpro and the NAB domain, ribbon representations for the reconstructed missing regions.

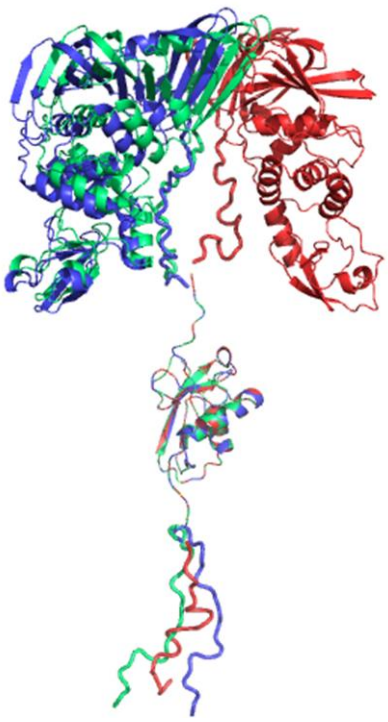


Figure 56: Superposition of the rigid body structural models of PLpro_NAB WT, $S_{466}R$ and $T_{467}K$ mutants.

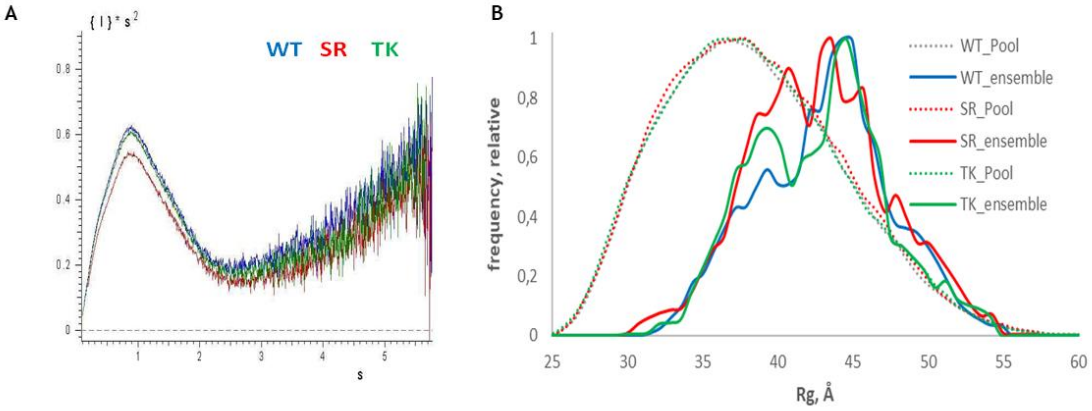


Figure 57: Comparison of PLpro WT, $S_{466}R$ and $T_{467}K$ mutants' plasticity in solution. A) Comparison of Kratky plots. B) Comparison of Rg distributions. The distributions for the initial random pools of models are shown by dot lines and the solid lines correspond to the selected ensembles.

In order to study in solution the shape and the low-resolution structure of the complex PLpro/ISG15, SEC-SAXS experiments were conducted at the BM29 BioSAXS beamline at the ESRF (Grenoble). Different PLpro and PLpro_NAB WT ratios were tested with the ISG15 due to the higher affinity measured, thus at ratios of 1:2 and 1:3. As expected from preliminary experiments conducted by HPLC-SLS, a peak of the complex was not achieved. Unfortunately, due to the available amount and concentration of ISG15, it was not possible to try higher ratios of the two proteins.

5.2.3 Discussion

The ISG15 is the preferred substrate of the SARS-CoV2 PLpro, binding at its N-terminus and C-terminus in two different sites on the protease [66]. The lack of kinetic information about the binding of the ISG15 and our double-domain PLpro_NAB, plus the complete absence of data about the δ-variant mutants, encouraged us to work on the ISG15. The PLpro is able to cleave the C-terminal tail of the precursor (proISG15), thus two constructs of the mature and immature forms were designed. The expression was optimized by exploring different tags and co-expression systems, obtaining in the end the ISG15 and proISG15 with a good grade of purity and reasonable yield. Working with the PLpro/ISG15 complex proved to be challenging, considering the low affinity and the low stability of the protein-protein interaction. Despite the daunting preliminary results, a further biochemical and biophysical characterization of the interaction led to an explanation of the previous evidence. The main result we can appreciate is the distinction in the activity of the δ-variant mutants. Even though the point mutations are located at the C-terminus of the NAB domain, they clearly affect the neighbour domain, the PLpro. The plasticity analysis of the PLpro_NAB WT and mutants by SAXS gave additional information on a possible reason why the δ-variant mutants present such different behaviour, even if an exhaustive explanation has not been reached yet. The enhanced mobility of the mutants could suggest a hamper in the binding with the substrate, but the binding analysis by GCI demonstrated an evident difference only for the pro-ISG15. The result hence suggests that the real discrepancy could reside in the catalytic activity, so deeper investigation is needed. From a structural point of view, a conformational study of the complex in solution could reveal additional differences between the PLpro/ISG15 and the PLpro_NAB/ISG15 complexes and, perhaps, with the complex formed with the mutants, but due to its instability, a cross-linking is necessary to perform the experiments. The structure of the PLpro/ISG15 complex has already been solved, however an in-depth structural study of the complex with the proISG15 could supply further information about the catalytic mechanism.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

This project focused on the SARS-CoV2 PLpro, a target as attractive as the more explored Mpro for a COVID-19 treatment development. Despite the large number of *in silico* and *in vitro* screenings published on this protease, only a few compounds have been validated as effective inhibitors, proving that PLpro represents a demanding protein for drug discovery [17,67]. In the present work, we applied an integrative biochemical and biophysical approach to overcome the hurdles found during the process. We thus started from the construct design and consequent protein production to perform all the screening steps, focusing on repurposing compounds. Repurposing known molecules provides various advantages, especially in drug development, but often hides challenges like false-positive results and selectivity issues, which can hamper drug development. The multi-technique application enabled us to discriminate which results were effectively promising. Indeed, a new repurposing inhibitor of PLpro, the CPI-169, was identified and characterized, showing similarities with another well-known repurposing inhibitor, the GRL-0617. Combining computational modelling and ligand-based NMR, we demonstrated that the two compounds share the same binding pocket, at the level of P2 site of PLpro, near the B2l. At this point, an effectiveness evaluation on a cell-based model could complete the inhibitory profile of the CPI-169. Likely GRL-0617, the CPI-169 is a non-covalent reversible inhibitor. Instead of the GRL compound, its pyridine portion interacts with the outer residues of the P2 pocket, providing new possibilities in the design of a more potent drug. A peculiarity of our work is the handling of a longer construct in addition to the PLpro, the PLpro_NAB which includes the nsp3 following domain at the C-terminus of the PLpro, the NAB domain. Up to now, all the structural and pharmaceutical investigations on this target have been performed on the PLpro domain. However, PLpro is a subdomain of the large multidomain protein nsp3 [15]. Because of the impossibility of handling such a complex protein *in vitro*, very little is known about the influence of a subdomain on another in terms of activity and structural properties. The construct we produced aimed to highlight such differences from the well-studied PLpro domain. Of course, the construct could not be considered representative of the whole nsp3, but in this work we achieved some interesting evidence. The interaction with the preferred human substrate of the PLpro, the ISG15, was investigated. In particular, we demonstrated that single-point mutations on the NAB domain can affect the catalytic activity of the PLpro, confirmed by the binding analysis at GCI. SAXS data show different plasticity and conformational changes in the mutants with respect to the WT. An exhaustive explanation of this behaviour has not been clarified yet; a deeper investigation should

be conducted as well as testing the effectiveness of the known inhibitors towards the mutants. The Delta-variant of SARS-CoV2, which tormented the global population between 2021 and 2022, fortunately disappeared. On the other hand, there is no additional information about these mutants and the following variants (e.g. Omicron) do not show such mutations. However, from the results collected in this thesis, it is possible to take home an important message: working on the PLpro itself is a beneficial way to discover new inhibitors, but it could be restrictive. It could not lead to successful drug development, thus we should always keep in mind that the real context of the PLpro in the nsp3 is much more complicated.

7. REFERENCES

1. Masters, P. S. & Perlman, S. in *Fields Virology* Vol. 2 (eds Knipe, D. M. & Howley, P. M.) 825–858 (Lippincott Williams & Wilkins, 2013).
2. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol. 2019 Mar;17(3):181-192. doi: 10.1038/s41579-018-0118-9. PMID: 30531947; PMCID: PMC7097006.
3. Lu R, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395:565–574.
4. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol. 2021 Mar;19(3):141-154. doi: 10.1038/s41579-020-00459-7. Epub 2020 Oct 6. Erratum in: Nat Rev Microbiol. 2022 May;20(5):315. PMID: 33024307; PMCID: PMC7537588.
5. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. J Med Virol. 2020 Apr;92(4):418-423. doi: 10.1002/jmv.25681. Epub 2020 Feb 7. Erratum in: J Med Virol. 2020 Oct;92(10):2249. PMID: 31967327; PMCID: PMC7167049.
6. Wrapp, Daniel et al. “Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.” Science (New York, N.Y.) vol. 367,6483 (2020): 1260-1263. doi:10.1126/science.abb2507
7. Chen, James et al. “Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex.” Cell vol. 182,6 (2020): 1560-1573.e13. doi:10.1016/j.cell.2020.07.033
8. Yang, Haitao, and Zihe Rao. “Structural biology of SARS-CoV-2 and implications for therapeutic development.” Nature reviews. Microbiology vol. 19,11 (2021): 685-700. doi:10.1038/s41579-021-00630-8
9. V'kovski, Philip et al. “Coronavirus biology and replication: implications for SARS-CoV-2.” Nature reviews. Microbiology vol. 19,3 (2021): 155-170. doi:10.1038/s41579-020-00468-6
10. Zhang JJ, et al., Risk and Protective Factors for COVID-19 Morbidity, Severity, and Mortality. Clin Rev Allergy Immunol. 2023 Feb;64(1):90-107. doi: 10.1007/s12016-022-08921-5. Epub 2022 Jan 19. PMID: 35044620; PMCID: PMC8767775.
11. Chaudhary, Namit et al. “mRNA vaccines for infectious diseases: principles, delivery and clinical translation.” Nature reviews. Drug discovery vol. 20,11 (2021): 817-838. doi:10.1038/s41573-021-00283-5

12. Li G, et al., Therapeutic strategies for COVID-19: progress and lessons learned. Nat Rev Drug Discov. 2023 Jun;22(6):449-475. doi: 10.1038/s41573-023-00672-y. Epub 2023 Apr 19. PMID: 37076602; PMCID: PMC10113999.
13. Hagemeijer MC, Monastyrska I, Griffith J, et al. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. Virology.2014;458–459:125–135
14. WolffG,ZhengS,KosterAJ,et al. A molecular pore spans the double membrane of the coronavirus replication organelle. Science.2020;369:1395–1398
15. Von Soosten L.C. et al., The Swiss army knife of SARS-CoV-2: the structures and functions of NSP3. CRYSTALLOGRAPHY REVIEWS2022, VOL. 28, NO. 1, 39–61https://doi.org/10.1080/0889311X.2022.2098281
16. Carlson CR, Asfaha JB, Ghent CM, et al. Phosphoregulation of phase separation by theSARS-CoV-2 N protein suggests a biophysical basis for its dual functions. Mol Cell.2020;80:1092–1103.e4
17. Lei J, Kusov Y, Hilgenfeld R. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. Antiviral Res.2018;149:58–74
18. Rack JGM, Zorzini V, Zhu Z, et al. Viral macrodomains: a structural and evolutionary assessment of the pharmacological potential. Open Biol.2020;10:200237
19. Kusov Y, Tan J, Alvarez E, et al. A G-quadruplex-binding macrodomain within the 'SARS-unique domain' is essential for the activity of the SARS-coronavirus replication–transcription complex. Virology.2015;484:313–322
20. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. Science. 2020;368(6489):409-412
21. Kneller DW, Phillips G, O'Neill HM, Jedrzejczak R, Stols L, Langan P, et al. Structural plasticity of SARS-CoV-2 3CL Mpro active site cavity revealed by room temperature X-ray crystallography. *Nat Commun*. 2020;11(1):3202
22. Owen DR, Allerton CMN, Anderson AS, Aschenbrenner L, Avery M, Berritt S, et al. An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19. Science. 2021;374(6575):1586-1593
23. Owen, Dafydd R et al. "An oral SARS-CoV-2 Mpro inhibitor clinical candidate for the treatment of COVID-19." Science (New York, N.Y.) vol. 374,6575 (2021): 1586-1593. doi:10.1126/science.abl4784
24. Macip G, Garcia-Segura P, Mestres-Truyol J, Saldivar-Espinoza B, Pujadas G, Garcia-Vallvé S. A review of the current landscape of SARS-CoV-2 main protease inhibitors: have we hit the bullseye yet? Int J Mol Sci. 2021;23(1):259

25. The Role of Structural Biology Task Force: Validation of the Binding Mode of Repurposed Drugs Against SARS-CoV-2 Protein Targets
26. Morasso S., Costanzi E., Demitri N., Giabbai B., Storici P. Exscalate4CoV: High-Performance Computing for COVID Drug Discovery, pp. 51-59 (2023), book chapter 8
27. Costanzi, Elisa et al. “Structural and Biochemical Analysis of the Dual Inhibition of MG-132 against SARS-CoV-2 Main Protease (Mpro/3CLpro) and Human Cathepsin-L.” International journal of molecular sciences vol. 22,21 11779. 29 Oct. 2021, doi:10.3390/ijms222111779
28. Harcourt, Brian H et al. “Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity.” Journal of virology vol. 78,24 (2004): 13600-12. doi:10.1128/JVI.78.24.13600-13612.2004
29. Freitas BT, et al. Characterization and noncovalent inhibition of the deu-biquitinase and deISGylase activity of SARS-CoV-2 papain-like protease. ACS Infect Dis.2020;6:2099–2109
30. Li L, Ma L, et al. Natural biflavones are potent inhibitors against SARS-CoV-2 papain-like protease. Phytochemistry. 2022 Jan;193:112984. doi: 10.1016/j.phytochem.2021.112984. Epub 2021 Oct 12. PMID: 34757253; PMCID: PMC8506144.
31. Osipiuk J, et al. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. Nat Commun. 2021 Feb 2;12(1):743. doi: 10.1038/s41467-021-21060-3. PMID: 33531496; PMCID: PMC7854729.
32. Klemm T, et al. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. EMBO J. 2020 Sep 15;39(18):e106275. doi: 10.15252/embj.2020106275. Epub 2020 Aug 26. PMID: 32845033; PMCID: PMC7461020.
33. Hughes JP, et al. Principles of early drug discovery. Br J Pharmacol. 2011 Mar;162(6):1239-49. doi: 10.1111/j.1476-5381.2010.01127.x. PMID: 21091654; PMCID: PMC3058157.
34. Blay, Vincent et al. “High-Throughput Screening: today's biochemical and cell-based approaches.” Drug discovery today vol. 25,10 (2020): 1807-1821. doi:10.1016/j.drudis.2020.07.024
35. Inglese, J. et al. (2007) High-throughput screening assays for the identification of chemical probes. Nat. Chem. Biol. 3, 466-479
36. Bergsdorf, C. and Ottl, J. (2010) Affinity-based screening techniques: their impact and benefit to increase the number of high quality leads. Expert Opin. Drug Discov. 5, 1095-1107

37. Congreve, M., Carr, R., Murray, C., and Jhoti, H. (2003). A 'Rule of Three' for fragment-based lead discovery? Drug Discovery Today 8, 876–877. doi: 10.1016/S1359-6446(03)02831-9
38. M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, J. Am. Chem. Soc., 123 (2001) 6108-6117
39. Norton, R. S, Leung, E. W, Chandrashekaran, I. R, and MacRaild, C. A (2016). Applications of 19F-NMR in fragment-based drug discovery. Molecules 21:860. doi: 10.3390/molecules21070860
40. Li, Qingxin. "Application of Fragment-Based Drug Discovery to Versatile Targets." Frontiers in molecular biosciences vol. 7 180. 5 Aug. 2020, doi:10.3389/fmolb.2020.00180
41. Maia, Eduardo Habib Bechelane et al. "Structure-Based Virtual Screening: From Classical to Artificial Intelligence." Frontiers in chemistry vol. 8 343. 28 Apr. 2020, doi:10.3389/fchem.2020.00343
42. Liu, S., et al. (2018). Practical model selection for prospective virtual screening. J. Chem. Inf. Model. 59, 282–293. doi: 10.1101/337956
43. Krishnamurthy N, et al. Drug repurposing: a systematic review on root causes, barriers and facilitators. BMC Health Serv Res. 2022 Jul 29;22(1):970. doi: 10.1186/s12913-022-08272-z. PMID: 35906687; PMCID: PMC9336118.
44. Petushkova AI, Zamyatnin AA Jr. Papain-Like Proteases as Coronaviral Drug Targets: Current Inhibitors, Opportunities, and Limitations. Pharmaceuticals (Basel). 2020 Sep 28;13(10):277. doi: 10.3390/ph13100277. PMID: 32998368; PMCID: PMC7601131
45. Calleja DJ, et al. Insights Into Drug Repurposing, as Well as Specificity and Compound Properties of Piperidine-Based SARS-CoV-2 PLpro Inhibitors. Front Chem. 2022 Apr 12;10:861209. doi: 10.3389/fchem.2022.861209. PMID: 35494659; PMCID: PMC9039177.
46. Rut W, et al. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design. Sci Adv. 2020 Oct 16;6(42):eabd4596. doi: 10.1126/sciadv.abd4596. PMID: 33067239; PMCID: PMC7567588.
47. Napolitano V, et al. Acriflavine, a clinically approved drug, inhibits SARS-CoV-2 and other betacoronaviruses. Cell Chem Biol. 2022 May 19;29(5):774-784.e8. doi: 10.1016/j.chembiol.2021.11.006. Epub 2022 Jan 11. PMID: 35021060; PMCID: PMC8751734
48. Báez-Santos YM, St John SE, Mesecar AD. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds.

Antiviral Res. 2015 Mar;115:21-38. doi: 10.1016/j.antiviral.2014.12.015. Epub 2014 Dec 29. PMID: 25554382; PMCID: PMC5896749

49. Swaim CD, et al. 6-Thioguanine blocks SARS-CoV-2 replication by inhibition of PLpro. iScience. 2021 Oct 22;24(10):103213. doi: 10.1016/j.isci.2021.103213. Epub 2021 Oct 2. PMID: 34632326; PMCID: PMC8487320.
50. Arya R, Prashar V, Kumar M. Evaluating Stability and Activity of SARS-CoV-2 PLpro for High-throughput Screening of Inhibitors. Mol Biotechnol. 2022 Jan;64(1):1-8. doi: 10.1007/s12033-021-00383-y. Epub 2021 Aug 22. PMID: 34420183; PMCID: PMC8380414
51. Zhao Y, et al. High-throughput screening identifies established drugs as SARS-CoV-2 PLpro inhibitors. Protein Cell. 2021 Nov;12(11):877-888. doi: 10.1007/s13238-021-00836-9. Epub 2021 Apr 17. PMID: 33864621; PMCID: PMC8052528.
52. Santos LH, et al. Structure-based identification of naphthoquinones and derivatives as novel inhibitors of main protease Mpro and papain-like protease PLpro of SARS-CoV-2. bioRxiv [Preprint]. 2022 Jan 5:2022.01.05.475095. doi: 10.1101/2022.01.05.475095. Update in: J Chem Inf Model. 2022 Aug 12;: PMID: 35018373; PMCID: PMC8750648
53. Ratia, Kiira et al. “A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication.” Proceedings of the National Academy of Sciences of the United States of America vol. 105,42 (2008): 16119-24. doi:10.1073/pnas.0805240105
54. Fu Z, et al. The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. Nat Commun. 2021 Jan 20;12(1):488. doi: 10.1038/s41467-020-20718-8. PMID: 33473130; PMCID: PMC7817691.
55. Shiraishi Yutaro and Ichio Shimada. “NMR Characterization of the Papain-like Protease from SARS-CoV-2 Identifies the Conformational Heterogeneity in Its Inhibitor-Binding Site.” Journal of the American Chemical Society vol. 145,30 (2023): 16669-16677. doi:10.1021/jacs.3c04115
56. Ma C, et al. Discovery of SARS-CoV-2 Papain-like Protease Inhibitors through a Combination of High-Throughput Screening and a FlipGFP-Based Reporter Assay. ACS Cent Sci. 2021 Jul 28;7(7):1245-1260. doi: 10.1021/acscentsci.1c00519. Epub 2021 Jun 18. PMID: 34341772; PMCID: PMC8265724.
57. Komander, David, and Michael Rape. “The ubiquitin code.” Annual review of biochemistry vol. 81 (2012): 203-29. doi:10.1146/annurev-biochem-060310-170328

58. Finley, Daniel. “Recognition and processing of ubiquitin-protein conjugates by the proteasome.” Annual review of biochemistry vol. 78 (2009): 477-513. doi:10.1146/annurev.biochem.78.081507.101607
59. Gold IM et al. “Coronaviral PLpro proteases and the immunomodulatory roles of conjugated versus free Interferon Stimulated Gene product-15 (ISG15).” Seminars in cell & developmental biology vol. 132 (2022): 16-26. doi:10.1016/j.semcdb.2022.06.005
60. Han H G et al. “ISG15 in cancer: Beyond ubiquitin-like protein.” Cancer letters vol. 438 (2018): 52-62. doi:10.1016/j.canlet.2018.09.007
61. Perng Yi-Chieh, and Deborah J Lenschow. “ISG15 in antiviral immunity and beyond.” Nature reviews. Microbiology vol. 16,7 (2018): 423-439. doi:10.1038/s41579-018-0020-5
62. Basters A, et al. Structural basis of the specificity of USP18 toward ISG15. Nat Struct Mol Biol. 2017 Mar;24(3):270-278. doi: 10.1038/nsmb.3371. Epub 2017 Feb 6. PMID: 28165509; PMCID: PMC5405867.
63. Swaim C D et al. “Extracellular ISG15 Signals Cytokine Secretion through the LFA-1 Integrin Receptor.” Molecular cell vol. 68,3 (2017): 581-590.e5. doi:10.1016/j.molcel.2017.10.003
64. Narasimhan J et al. “Crystal structure of the interferon-induced ubiquitin-like protein ISG15.” The Journal of biological chemistry vol. 280,29 (2005): 27356-65. doi:10.1074/jbc.M502814200
65. Dzimianski JV, et al. ISG15: It's Complicated. J Mol Biol. 2019 Oct 4;431(21):4203-4216. doi: 10.1016/j.jmb.2019.03.013. Epub 2019 Mar 16. PMID: 30890331; PMCID: PMC6746611.
66. Wydorski PM, Osipiuk J, et al. Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and K48-linked di-ubiquitin. Nat Commun. 2023 Apr 25;14(1):2366. doi: 10.1038/s41467-023-38031-5. PMID: 37185902; PMCID: PMC10126577.
67. Ma C, Wang J. Validation and Invalidation of SARS-CoV-2 Papain-like Protease Inhibitors. ACS Pharmacol Transl Sci. 2022 Jan 24;5(2):102-109. doi: 10.1021/acsptsci.1c00240. PMID: 35178512; PMCID: PMC8806001
68. Petrone D et al. “Reduction of the risk of severe COVID-19 due to Omicron compared to Delta variant in Italy (November 2021 - February 2022).” International journal of infectious diseases: IJID: official publication of the International Society for Infectious Diseases vol. 129 (2023): 135-141. doi:10.1016/j.ijid.2023.01.027

69. Kumar P, et al. Integrated genomic view of SARS-CoV-2 in India [version 1; peer review: 3 approved]. Wellcome Open Res 2020, 5:184 (https://doi.org/10.12688/wellcomeopenres.16119.1)
70. Corsello, S M et al. "The Drug Repurposing Hub: a next-generation drug library and information resource." Nature medicine vol. 23,4 (2017): 405-408. doi:10.1038/nm.4306
71. Kuzikov, M et al. "Identification of Inhibitors of SARS-CoV-2 3CL-Pro Enzymatic Activity Using a Small Molecule in Vitro Repurposing Screen." ACS pharmacology & translational science vol. 4,3 1096-1110. 11 Mar. 2021, doi:10.1021/acsptsci.0c00216
72. Kathy Huynh and Carrie L Partch , Analysis of protein stability and ligand interactions by thermal shift assay, Curr Protoc Prot Sci, 79:28.9.1-28.9.14, 2015
73. Boivin, S et al. "Optimization of protein purification and characterization using Thermofluor screens." Protein expression and purification vol. 91,2 (2013): 192-206. doi:10.1016/j.pep.2013.08.002
74. Niesen, F H et al. "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nature protocols vol. 2,9 (2007): 2212-21. doi:10.1038/nprot.2007.321
75. Malinovska, L et al. "Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: an advanced protocol for high-throughput applications." Nature protocols vol. 18,3 (2023): 659-682. doi:10.1038/s41596-022-00771-x
76. Viegas, A et al. "Ligand-based nuclear magnetic resonance screening techniques." Methods in molecular biology (Clifton, N.J.) vol. 572 (2009): 81-100. doi:10.1007/978-1-60761-244-5_6
77. Meyer, Bernd, and Thomas Peters. "NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors." Angewandte Chemie (International ed. in English) vol. 42,8 (2003): 864-90. doi:10.1002/anie.200390233
78. Walpole, S et al. "STD NMR as a Technique for Ligand Screening and Structural Studies." Methods in enzymology vol. 615 (2019): 423-451. doi:10.1016/bs.mie.2018.08.018
79. Ludwig, Christian, and Ulrich L Guenther. "Ligand based NMR methods for drug discovery." Frontiers in bioscience (Landmark edition) vol. 14,12 4565-74. 1 Jan. 2009, doi:10.2741/3549

80. M. Mayer, B. Meyer, Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, J. Am. Chem. Soc., 123 (2001) 6108-6117
81. Clore GM and Gronenborn AM, Theory and applications of the transferred nuclear overhauser effect to the study of the conformations of small ligands bound to proteins, J. Magn. Res., 48 (1982) 402-417
82. Louche, A et al. "Protein-Protein Interactions: Pull-Down Assays." Methods in molecular biology (Clifton, N.J.) vol. 1615 (2017): 247-255. doi:10.1007/978-1-4939-7033-9_20
83. Konarev PV et al. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 2003, 36, 1277–1282
84. Manalastas-Cantos K et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. J. Appl. Cryst. 2021, 54, 343—355
85. Guinier, A. La diffraction des rayons X aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques. Ann. Phys. 1939, 12, 161–237
86. Svergun, D.I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. 1992, 25, 495–503.
87. Porod, G. Small Angle X-ray Scattering"; Academic Press, London, 1982
88. Franke, D and Svergun, DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Cryst., 42, 342-346
89. Kozin, MB and Svergun, DI Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. 2001, 34, 33–41
90. Volkov VV and Svergun DI Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. 2003, 36, 860–864
91. Tuukkanen AT, Kleywegt GJ and Svergun DI (2016) Resolution of ab initio shapes determined from small-angle scattering IUCrJ 3, 440-447
92. Petoukhov MV et al. New developments in the ATSAS program package for small-angle scattering data analysis. J. Appl. Cryst. 2012, 45, 342-350
93. Waterhouse, A et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296-W303 (2018)
94. Tria G et al. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. IUCr J. 2015, 2, 207-217
95. Svergun, DI et al. CRYSOL – A Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Crystallogr. 1995, 28, 768–773

96. Shin D, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. Nature. 2020 Nov;587(7835):657-662. doi: 10.1038/s41586-020-2601-5. Epub 2020 Jul 29. PMID: 32726803; PMCID: PMC7116779
97. Gotoh, Y et al. Novel antibacterial compounds specifically targeting the essential WalR response regulator. J Antibiot 63, 127–134 (2010). doi.org/10.1038/ja.2010.4
98. Ma, Liang et al. "A novel 8-hydroxyquinoline derivative induces breast cancer cell death through paraptosis and apoptosis." Apoptosis: an international journal on programmed cell death vol. 27,7-8 (2022): 577-589. doi:10.1007/s10495-022-01737-w
99. Han, Yusheng et al. "NAD(P)H: quinone oxidoreductase-1-dependent and -independent cytotoxicity of potent quinone Cdc25 phosphatase inhibitors." The Journal of pharmacology and experimental therapeutics vol. 309,1 (2004): 64-70. doi:10.1124/jpet.103.059477
100. Blum, Gil et al. "Small-molecule inhibitors of SETD8 with cellular activity." ACS chemical biology vol. 9,11 (2014): 2471-8. doi:10.1021/cb500515r
101. Liljebris, C et al. "Oxidation of protein tyrosine phosphatases as a pharmaceutical mechanism of action: a study using 4-hydroxy-3,3-dimethyl-2H-benzo[g]indole-2,5(3H)-dione." The Journal of pharmacology and experimental therapeutics vol. 309,2 (2004): 711-9. doi:10.1124/jpet.103.062745
102. Li, C J et al. "beta-Lapachone, a novel DNA topoisomerase I inhibitor with a mode of action different from camptothecin." The Journal of biological chemistry vol. 268,30 (1993): 22463-8
103. Li, Yitang et al. "Pretreatment with phosphatase and tensin homolog deleted on chromosome 10 (PTEN) inhibitor SF1670 augments the efficacy of granulocyte transfusion in a clinically relevant mouse model." Blood vol. 117,24 (2011): 6702-13. doi:10.1182/blood-2010-09-309864
104. Wehner, S et al. "Inhibition of p38 mitogen-activated protein kinase pathway as prophylaxis of postoperative ileus in mice." Gastroenterology vol. 136,2 (2009): 619-29. doi:10.1053/j.gastro.2008.10.017
105. Kim, Ji Yeun et al. "Safe, high-throughput screening of natural compounds of MERS-CoV entry inhibitors using a pseudovirus expressing MERS-CoV spike protein." International journal of antimicrobial agents vol. 52,5 (2018): 730-732. doi:10.1016/j.ijantimicag.2018.05.003
106. Zhao, Wenwen et al. "Dihydrotanshinone I Attenuates Atherosclerosis in ApoE-Deficient Mice: Role of NOX4/NF-κB Mediated Lectin-Like Oxidized LDL

Receptor-1 (LOX-1) of the Endothelium." Frontiers in pharmacology vol. 7 418. 8 Nov. 2016, doi:10.3389/fphar.2016.00418

107. Lee H, et al. Reducing agents affect inhibitory activities of compounds: results from multiple drug targets. Anal Biochem. 2012 Apr 1;423(1):46-53. doi: 10.1016/j.ab.2012.01.006. Epub 2012 Jan 18. PMID: 22310499; PMCID: PMC3299889

108. Cotto-Rios XM, et al. Deubiquitinases as a signaling target of oxidative stress. Cell Rep. 2012 Dec 27;2(6):1475-84. doi: 10.1016/j.celrep.2012.11.011. Epub 2012 Dec 6. PMID: 23219552; PMCID: PMC3534866

109. Zhou YB, et al. LGH00031, a novel ortho-quinonoid inhibitor of cell division cycle 25B, inhibits human cancer cells via ROS generation. Acta Pharmacol Sin. 2009 Sep;30(9):1359-68. doi: 10.1038/aps.2009.131. PMID: 19730430; PMCID: PMC4085684

110. Gopinath, P et al. (2016). Chemical and semisynthetic approaches to study and target deubiquitinases. Chemical Society Reviews, 45(15), 4171-4198. doi:10.1039/c6cs00083e

111. Bradley, W D et al. "EZH2 inhibitor efficacy in non-Hodgkin's lymphoma does not require suppression of H3K27 monomethylation." Chemistry & biology vol. 21,11 (2014): 1463-75. doi:10.1016/j.chembiol.2014.09.017

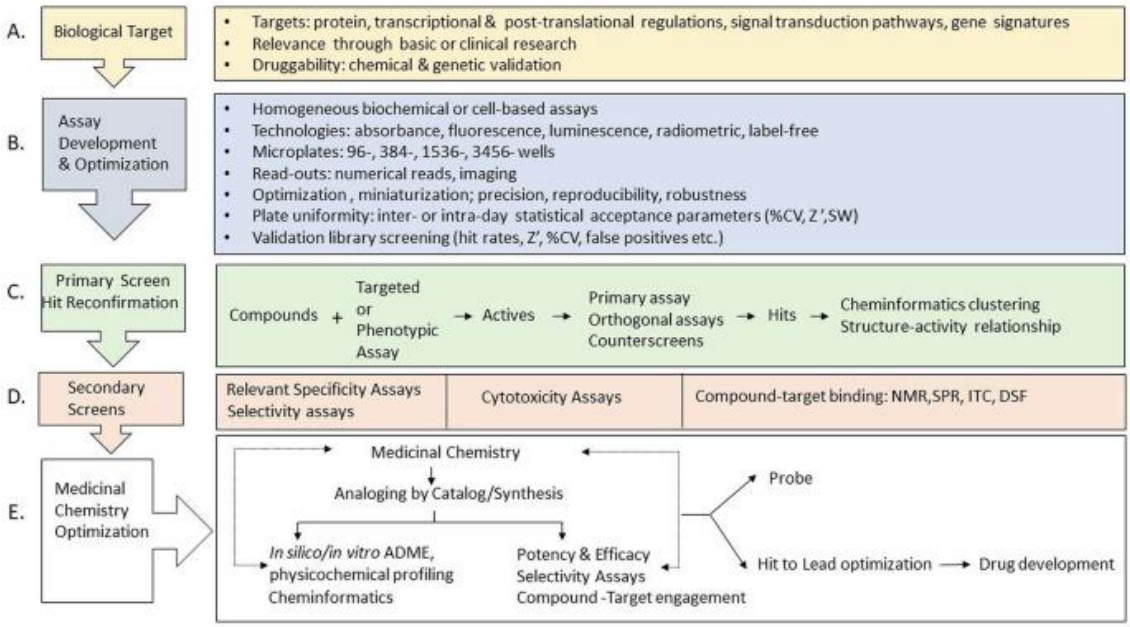
112. Wallace A. C., et al (1995). LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. Protein Engineering, 8(2), 127-134

113. Stefanelli I et al. "Broad-spectrum coronavirus 3C-like protease peptidomimetic inhibitors effectively block SARS-CoV-2 replication in cells: Design, synthesis, biological evaluation, and X-ray structure determination." European journal of medicinal chemistry vol. 253 (2023): 115311. doi:10.1016/j.ejmech.2023.115311

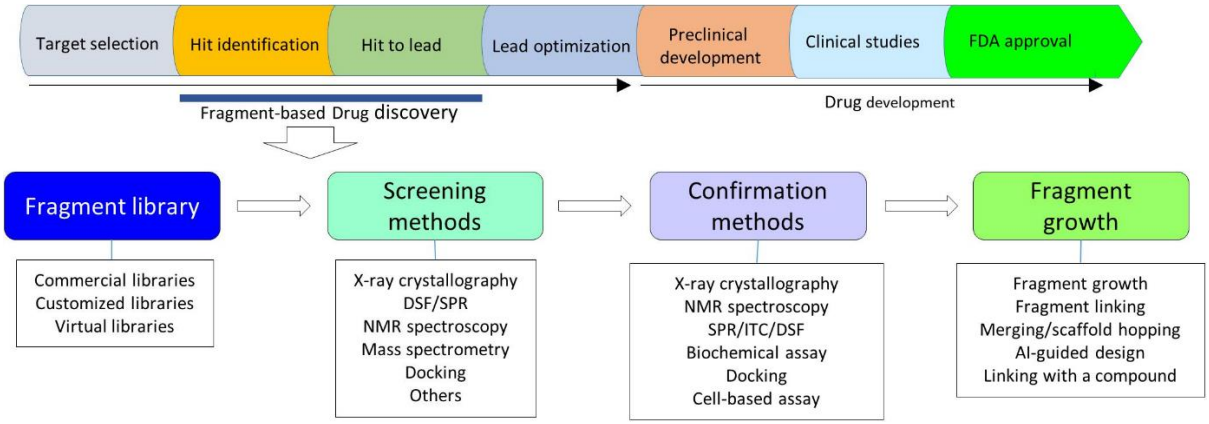
114. Roy, A. (2021). High-Throughput Screening (HTS) Technology. In: Offermanns, S., Rosenthal, W. (eds) Encyclopedia of Molecular Pharmacology. Springer, Cham. https://doi.org/10.1007/978-3-030-57401-7_73

115. Baell JB and Nissink JWM (2018). "Seven Yaer Itch: Pan-Assay Interference Compounds in 2017-Utility and Limitations"; ACS Chem. Biol. 13, 36-44

SUPPLEMENTARY MATERIAL

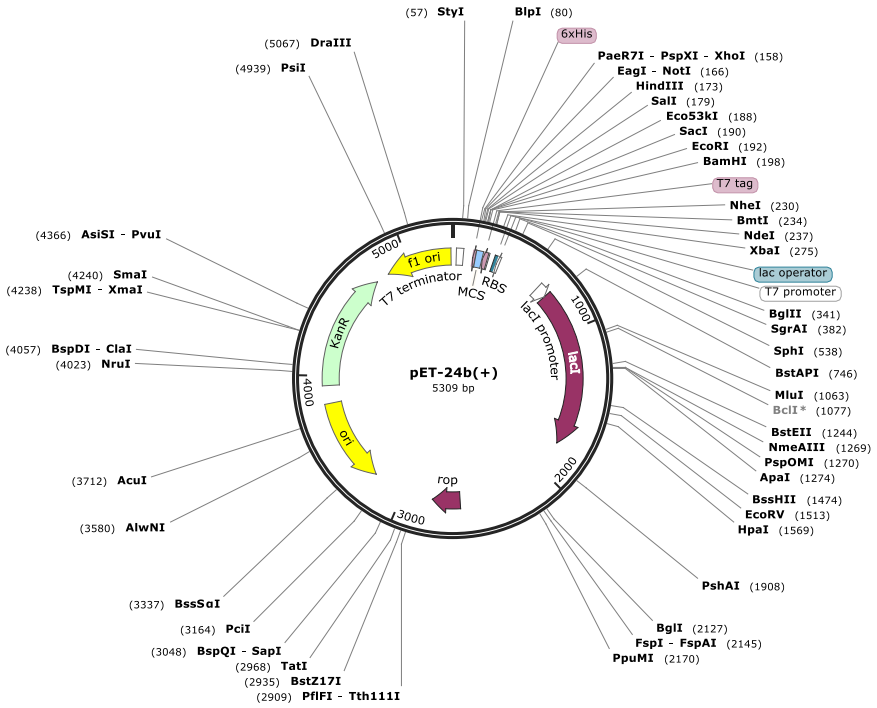


Supp. Figure 1: Scheme of the HT screening workflow (from Anuradha Roy, 2021) [114].

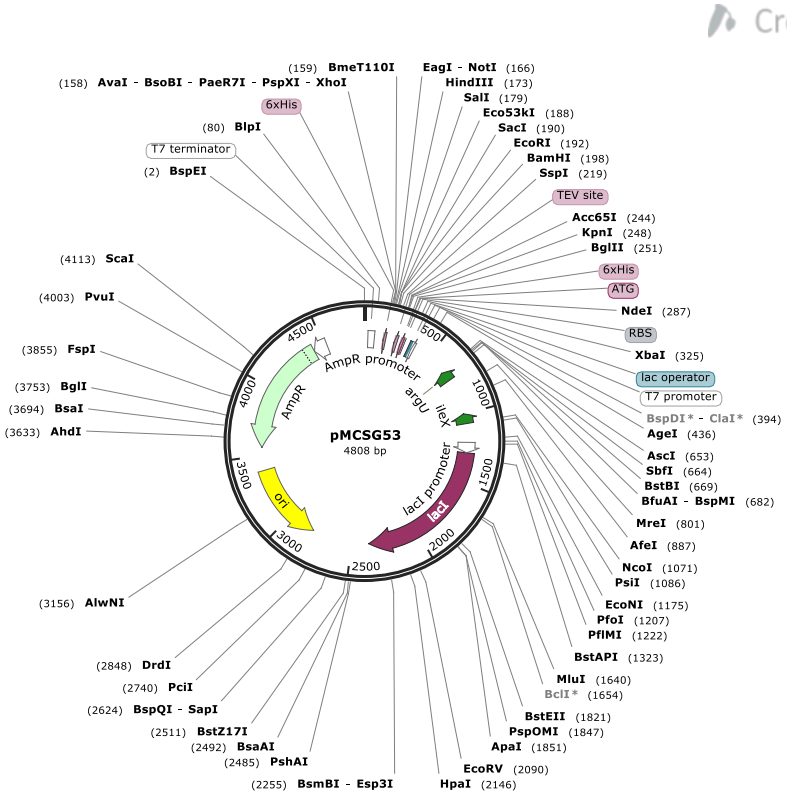


Supp. Figure 2: Scheme of fragment-based screening and fragment-based drug design process (from Li, Qingxin 2020) [40].

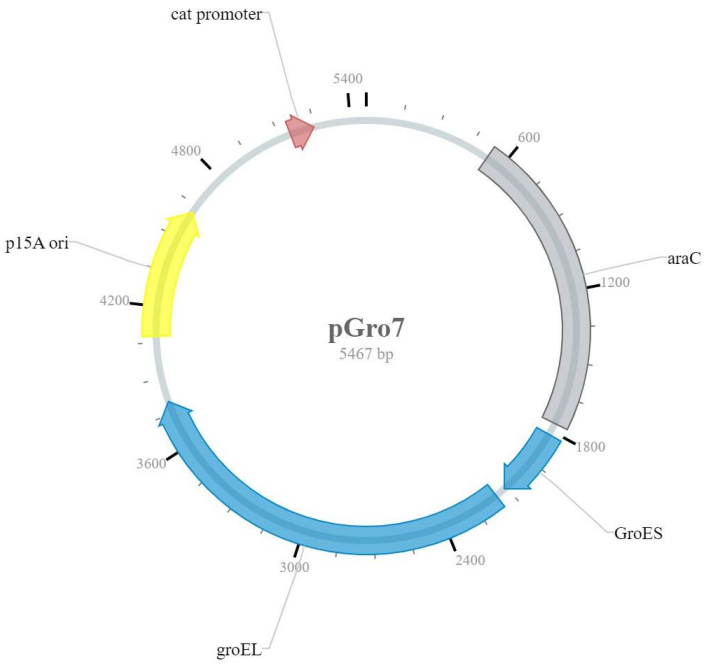
Expression vector maps



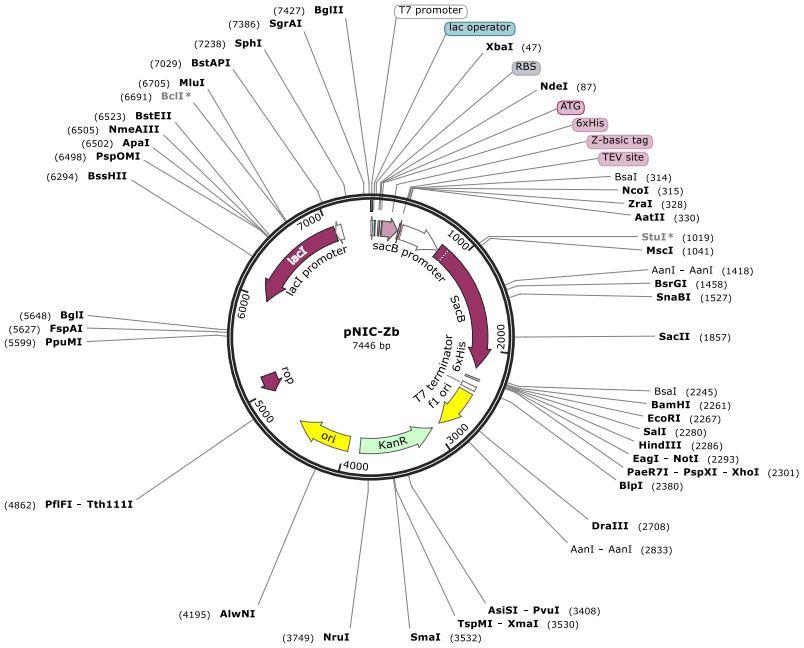
Supp. Figure 3: pET24b+ vector map



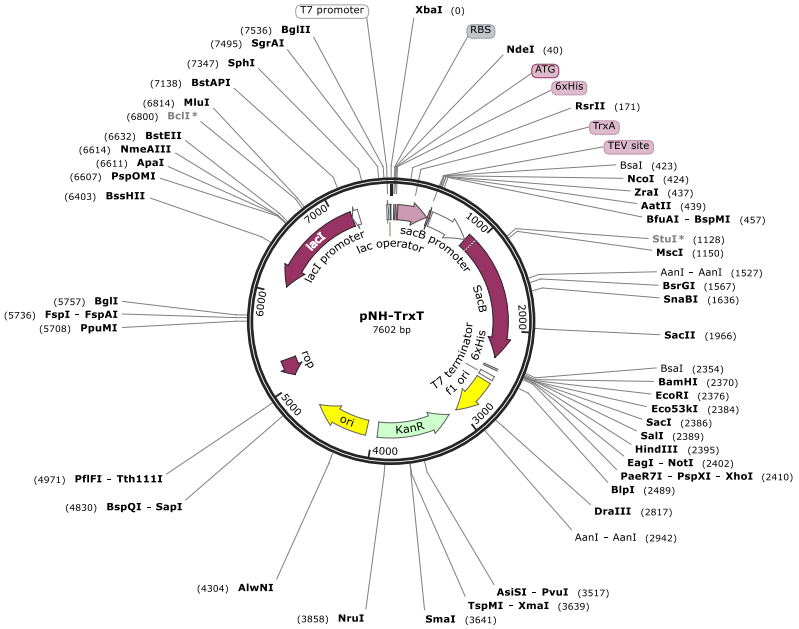
Supp. Figure 4: pMCSG53 vector map.



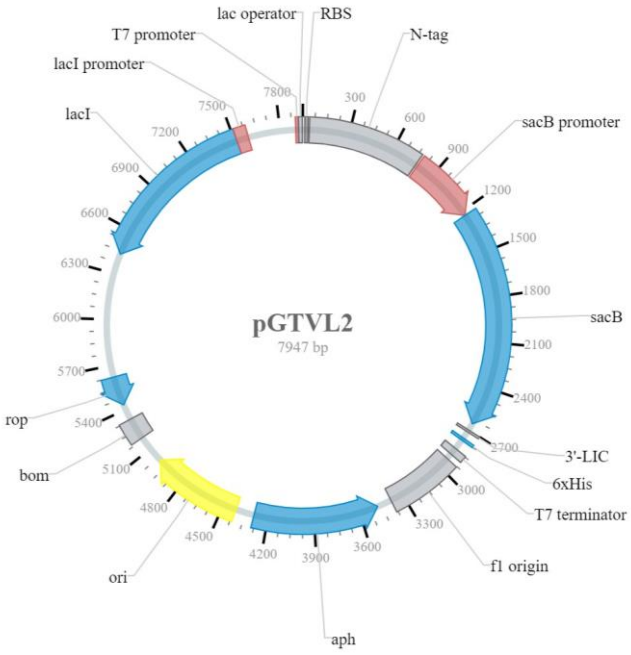
Supp. Figure 5: pGro7 vector map.



Supp. Figure 6: pNIC-ZB vector map.



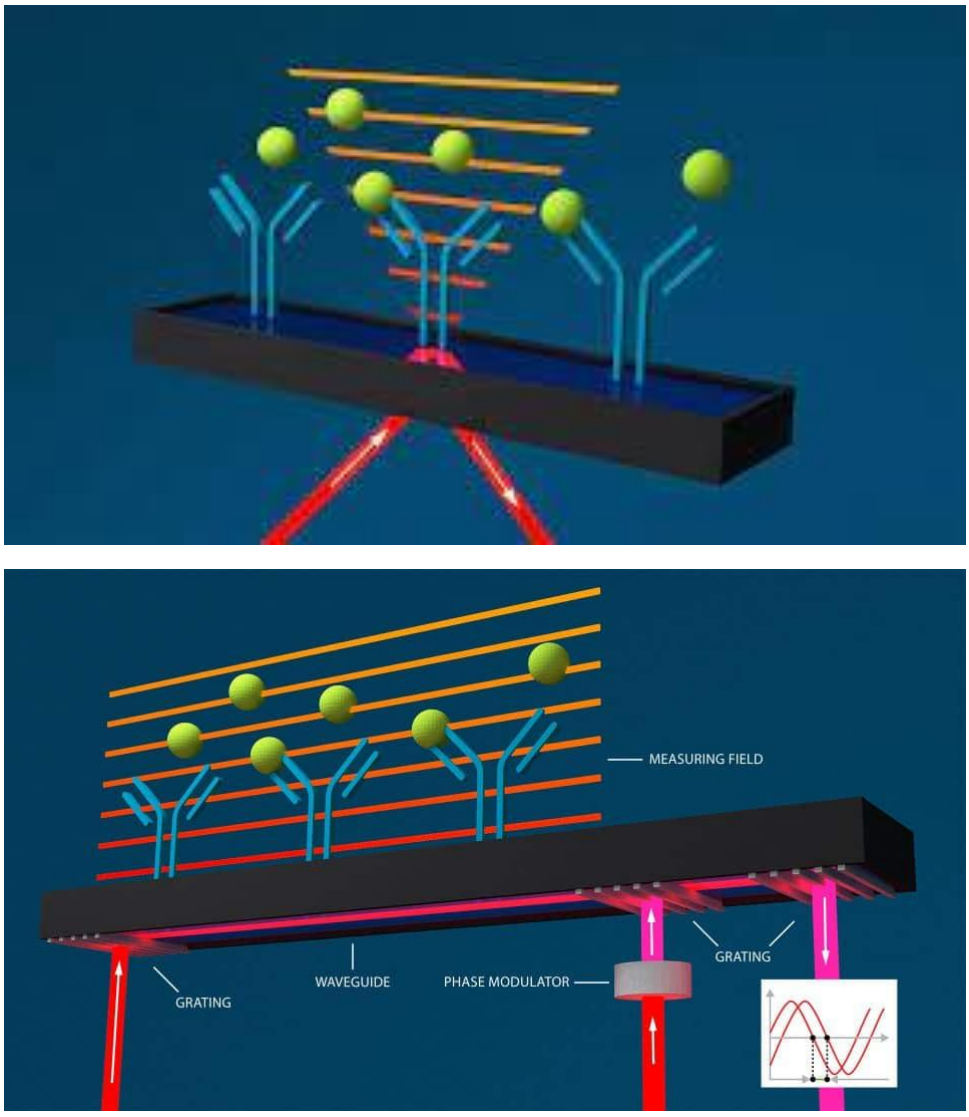
Supp. Figure 7: pNH-TrxT vector map.



Supp. Figure 8: pGTvL2 vector map.

Supp. Table 1: primers used for LIC of ISG15 and proISG15. In green: LIC sequence; in red: stop codons; in yellow: sequence for insert annealing.

Primer	Sequence
LIC ISG15 forward	TACTTCCAATCCATGGGTTGGGACCTGAC
LIC ISG15 reverse	TATCCACCTTTACTGTCATTAACCGCCGCCGCGCA
LIC proISG15 reverse	TATCCACCTTTACTGTCATTAGCTACGGCCAC



Supp. Figure 9: representation of the experimental setup of the SPR (above); experimental setup of GCI, explaining the increased readout and the consequent higher signal and sensitivity than the SPR (below). From Malvern Paranalytical web site.

Supp. Table 2: Parameters of the data collection performed on PLpro crystals at the XRD2 beamline in Elettra Sincrotrone Trieste. The autoproc processing generated from best diffracting crystals of each condition is reported. Data were not refined. 1: 0.1 M Acetate buffer, 0.8 M NaH2PO4 / 1.2 M K2HPO4, pH 4.5, reported by Osipiuk et al, 2021 [31]; 2: Sodium-malonate pH 7 1.5M; 3. 0.1M Tris pH 8.0, 1.6/1.8 M Ammonium sulfate, pre-incubating proteins with 8% of No. 26 of Silver Bullets™ (Hampton Research) additive, reported by Zhao et al, 2021 [51].

	Space group	Cell parameters	Resolution Å	<I/σ(I)>	$CC^{1/2}$	Completeness %
1	$P3_2$ 2 1	a=82.32 b=82.32 c=134.58	1.73 - 71.29	32.2 (2.3)	1.0 (0.8)	100
2	$P3_1$ 2 1	a=81.51 b=81.51 c=135.17	1.68 - 70.59	27.3 (2.1)	1.0 (0.8)	100
3	P6 2 2	a=115.99 b=115.99 c=254.18	2.1 – 49.27	25.8 (1.9)	1.0 (0.76)	99

ACKNOWLEDGEMENTS

I would like to thank our precious collaborators who participated in the PLpro project. In particular, Maria Kuzikov, Dr Andrea Zaliani and Dr Philip Gribbon from Fraunhofer Institute in Hamburg for the development of the screening and the intense networking. Vittoria Monaco and Prof Maria Monti from CEINGE institute in Naples, for the MS characterization and the essential role in demonstrating the oxidative reaction of false-positive compounds. A special thanks go to our collaborators from the National Institute of Chemistry in Ljubljana, in particular Prof Simona Golič Grdadolnik, Dr Primož Šket and Prof Janez Plavec, who welcomed me in their wonderful NMR facility and helped me in the NMR experiments which marked a turning point in my PhD project.

I would like to express my gratitude to Dr Sonia Covaceuszach, for the great help in the PLpro-ISG15 characterization and for SAXS data processing and analysis, but above all for being such a nice colleague and person.

I also thank the Central European Research Infrastructure Consortium, CERIC-ERIC, to have funded my PhD scholarship and given to me the opportunity to explore new techniques at the NMR facility of Ljubljana.

All my gratitude goes to my research group who have made me feel an integral part of the team. I would like to thank my supervisor Dr Paola Storici to have bet on me and to have given this intense but thrilling opportunity to me. I want to thank also our Master thesis student, Livia Palumbo for the great help in the ISG15 branch, to my past and present colleagues, in particular Barbara Giabbai and Dr Andrea dalle Vedove and all people who made me grow during these three years. A special thanks goes to the angel of our lab, our technician Federica Ruzzier.

My heartfelt thanks to my favourite post-docs, Dr Marta Stefania Semrau and Dr Theo Battista. They supported me in the bright and odd days, always making me feel part of a family.