Università Politecnica delle Marche
Biomolecular Sciences XXXIII cycle
Doctoral thesis

Recombinant protein for diagnostic purposes: expression optimization, purification and structural characterization

Lucrezia Savini

Supervisor: Prof. Paolo Mariani

2020

Abstract

In recent years, the production of recombinant proteins has become a basic tool for structural, biochemical and biophysical studies, but has been also considered fundamental for the develpment of therapeutic and biotechnological applications. For this reason, innovative studies and approaches are necessary for the optimization of the expression, purification and characterization of these proteins.

This thesis focuses on the production of two proteins, the human hemoglobin receptor (aka CD163) and the SARS-CoV-2 Main protease (aka Mpro) obtained by recombinant technology. In both cases, the production procedure has been optimazied to obtain material as pure as possible and in appropriate quantities.

The *human hemoglobin receptor* CD163 is a multifunctional endocytic receptor and is characterized by a repetition of the cysteine-rich scavenger receptor (SRCR). It is known to play a critical role in various physiological and pathological pathways. However, the structural and functional details of this protein have not yet been elucidated, even less so if considering the complex mechanism of interaction with its ligand (Haptoglobin-hemoglobin (Hp-Hb) complex). Indeed, once produced in the necessary quantities, the objective of the thesis was to clarify the structural properties of CD163, both before and after the complex formation, by using X-ray crystallography and/or cryoEM techniques (in collaboration with the Roma Sapienza University). CD163 was then recombinantly produced in two different eukaryotic expression systems: *mammalian* and *insect* cells. For both expression systems, the produced CD163 resulted homogeneous after purification both by metal affinity chromatography and by size exclusion. The Baculovirus system, however, seems to allow a

higher protein expression yield, resulting more suitable for large-scale production. Subsequent studies showed that the physiological link of recombinant CD163 with the Hp-Hb complex is not stable and the complex formation was not quantitatively relevant. In particular, the ligand-receptor complex seems to form after mixing, but the investigations carried out at equilibrium by dynamic light scattering show two distinct populations of molecules within the solution. However, calorimetric analysis conducted on the individual components of the ternary complex did not reveal any structural or folding problems. Such a result, confirmed by some first cryoEM tests performed by Roma Sapienza group on the complex, led us to further study the binding mechanism between recombinant CD163 and the Hp-Hb complex, essentially trying to reconsider the stability and purity of the produced protein and the modalities of formation of the complex. However, the COVID-19 pandemic made the work progress very complicated and it also seemed very necessary to try to provide a possible contribution to research against the spread of the SARS-CoV-2 virus by applying the weapons we know.

Therefore, in the last year the research activity has focused on the production and characterization of a protein, the *main protease* (Mpro) that can be considered the heart of the functioning of the coronavirus. The maturation of SARS-CoV-2, in fact, requires a protease (Mpro) to cleave the polyproteins encoded by the virus. Despite a large amount of experimental information already available, there is wide disagreement on the active form of Mpro: a monomer-dimer equilibrium characterizes the protease in solution, and several evidences suggest that the active functional unit is the dimeric. Therefore, detailed knowledge of the thermodynamics of the monomer-dimer equilibrium and of the possible effects of small molecules on the dimer formation could be a key information for possible therapeutic interventions: if the binding to Mpro and the interference with the dimerization process and then to the protease activity could be proved, small molecules can be used as potential, broad

spectrum antiviral drugs. Inside a national collaboration, the production of recombinant Mpro in *E. coli* Expression System has been started, testing different temperatures of expression in order to find the best one. Once the protein was produced both at small and medium scale, Mpro was purified and activity assays and inhibition tests were performed by considering a series of small inhibitors selected by virtual screening by a group from University of Palermo. We find that the selected inhibitors affect dimerization and enzymatic activity to a different extent and sometimes in an opposite way, likely due to the different molecular mechanism of the two processes. The amount of Mpro produced was also sufficient to perform a series of biophysical experiments to investigate the structural characteristics of the monomer-dimer equilibrium. In particular, small angle X-ray scattering (SAXS) experiments were performed at Diamond Synchrotron. The study was performed both in the absence and in the presence of the selected small inhibitors, in order to simultaneously test their therapeutic potential with respect to dimerization inhibition. By measuring the large-scale structural features of Mpro as a function of temperature, protein concentration and in the presence of different amounts of inhibitors an accurate thermodynamic picture of the Mpro inhibitor-dependent dimerization process has been provided. Moreover, the analysis suggested a few Mpro amino acid residues that emerge as key to optimize both dissociation and enzymatic activity inhibition and that can be the basis for the development of future drug molecules.

Table of contents

General introduction

In recent years, the production of recombinant proteins has become a basic tool for structural, biochemical and biophysical studies, but has been also considered fundamental for the development of therapeutic and biotechnological applications. For this reason, innovative studies and approaches are necessary for the optimization of the expression, purification and characterization of these proteins.

The purpose of this doctoral thesis is to evaluate strategies to obtain an appreciable amount of recombinant proteins both from the quantitative and qualitative aspects for, structural and biotechnological characterization. Moreover, after the application of several biochemical methods, the resulting protein should be pure and biologically active.

The aim was to achieve this by comparing and using different experimental techniques and conditions.

The research project was initially focused on the study of the expression and structural characterization of a human protein: the receptor for hemoglobin (aka CD163). However, the COVID-19 pandemic has made the progress of the work very complicated and it was necessary to assess a world-wide contribution to research against the spread of the SARS-Cov-2 virus. For this reason, at a later stage, the work focused on the study of the expression and biological activity of the main protease of SARS-Cov2 (aka Mpro).

For those reasons, the research conducted for the development of this doctorate work is illustrated in two different sections, both focusing on molecular and biochemical strategies for the production of high-scale recombinant proteins.

The first part illustrates the study conducted on CD163 and the second part shows the study conducted on the SARS-CoV-2 main protease (Mpro). The latter was completed by the publication of the article attached in the appendix.

In the first chapter, The human hemoglobin receptor CD163 is described. It is a multifunctional endocytic receptor and is characterized by a repetition of the cysteine-rich scavenger receptor (SRCR). It is known to play a critical role in various physiological and pathological pathways.

However, the structural and functional details of this protein have not yet been elucidated, even less so if considering the complex mechanism of interaction with its ligand (Haptoglobin-hemoglobin (Hp-Hb) complex). Indeed, once produced in the necessary quantities, the objective of the thesis was to clarify the structural properties of CD163, both before and after the complex formation, by using X-ray crystallography and/or cryoEM techniques (in collaboration with the Roma Sapienza University). CD163 was then recombinantly produced in two different eukaryotic expression systems: mammalian and insect cells. For both expression systems, the produced CD163 resulted homogeneous after purification both by metal affinity chromatography and by size exclusion.

The Baculovirus system, however, seems to allow a higher protein expression yield, resulting more suitable for large-scale production. Subsequent studies showed that the physiological link of recombinant CD163 with the Hp-Hb complex is not stable and the complex formation was not quantitatively relevant. In particular, the ligand-receptor complex seems to form after mixing, but the investigations carried out at equilibrium by dynamic light scattering show two distinct populations of molecules within the solution. However, calorimetric analysis conducted on the individual components of the ternary complex did not reveal any structural or folding problems. Such a result, confirmed by some first cryoEM tests performed by Roma Sapienza group on the complex, led us to further study the binding mechanism between recombinant CD163 and the Hp-Hb complex, essentially trying to reconsider the stability and purity of the produced protein and the modalities of formation of the complex.

As previously reported due to complications and restrictions due to the spread of the COVID-19 pandemic, the final experiments on the CD163 study were postponed. Although seemingly limiting, for the work carried out, the pandemic has nevertheless made it possible to contribute to the progress of scientific discoveries concerning SARS-Cov-2.

In fact, in the last year the research activity has focused on the production and characterization of a protein, the main protease (Mpro) that can be considered the heart of the functioning of the coronavirus. The maturation of SARS-CoV-2, in fact, requires a protease (Mpro) to cleave the polyproteins encoded by the virus. Despite a large amount of

experimental information already available, there is wide disagreement on the active form of Mpro: a monomer-dimer equilibrium characterizes the protease in solution, and several evidences suggest that the active functional unit is the dimeric. Therefore, detailed knowledge of the thermodynamics of the monomer-dimer equilibrium and of the possible effects of small molecules acting as inhibitors of the dimer formation could be a key information for possible therapeutic interventions: if the binding to Mpro and the interference with the dimerization process and then to the protease activity could be proved, small molecules can be used as potential, broad spectrum antiviral drugs. Inside a national collaboration, the production of recombinant Mpro in E.coli Expression System has been started, testing different temperatures of expression in order to find the best one. Once the protein was produced both at small and medium scale, Mpro was purified and activity assays and inhibition tests were performed by considering a series of small inhibitors selected by virtual screening by a group from University of Palermo. We find that the selected inhibitors affect dimerization and enzymatic activity to a different extent and sometimes in an opposite way, likely due to the different molecular mechanism of the two processes.

The amount of Mpro produced was also sufficient to perform a series of biophysical experiments to investigate the structural characteristics of the monomer-dimer equilibrium. In particular, small angle X-ray scattering (SAXS) experiments were performed at Diamond Synchrotron. The study was performed both in the absence and in the presence of the selected small inhibitors, in order to simultaneously test their therapeutic potential with respect to dimerization inhibition. By measuring the large-scale structural features of Mpro as a function of temperature, protein concentration and in the presence of different amounts of inhibitors an accurate thermodynamic picture of the Mpro inhibitor-dependent dimerization process has been provided. Moreover, the analysis suggested a few Mpro amino acid residues that emerge as key to optimize both dissociation and enzymatic activity inhibition and that can be the basis for the development of future drug molecules.

Cd163

1. Introduction

The superfamily SRCR-SF, includes protein receptors soluble or bound to the membrane, rich in cysteine and characterized by the presence of a highly conserved domain: the SRCR domain.

To date, the SRCR-SF has more than 30 members, but this number is expected to increase.

The aminoacid residues of the SRCR domain are highly conserved and have not shown any significant mutations in the course of evolution; on the contrary, the aminoacid residues flanking SRCR domain, have undergone, over time, a considerable evolution generating a pool of similar structurally proteins but profoundly different in function.

1.1 The history of super family SRCR-SF

The first information regarding the superfamily of Scavenger receptors rich in cysteines, go back to the early nineties thanks to the discovery of a series of scavenger receptors containing one or more extracellular domains homologous to the C-terminal domain cysteine-rich scavenger receptor expressed in macrophages (SR-AI) (Freeman et al., 1990).

Initially, some characteristics of these proteins seemed common to both the immunoglobulin superfamily (Huang et al., 1987; Resnick et al., 1994) and a bacterial chaperonin Papd, which has the role of guiding the assembly of Pili in *E. coli* (Holmgren and Branden, 1989). This fact suggested that these proteins could be related even if, this theory was not supported by a detailed alignment analysis.

Then it became clear that these proteins form a separate superfamily, now called SRCR-SF.

Members of the SRCR family are closely related from a structural point of view but share few common functions. The fact that some SRCR-SF members behave functionally as Scavenger Receptors has created much confusion about the identification of the members of these two families of receptors. Scavenger (SR) receptors were initially defined for their ability to absorb modified lipoproteins (Brown and Goldstein, 1983). To date it is known the ability of these proteins to bind different molecules, both endogenous and pathogenic. The factor that unites these ligands is their polyanionic nature, in fact with the term SR to date defines an extracellular glycoprotein (soluble or bound to the membrane) involved in the recognition and / or endocytosis of negatively charged molecules (Sarrias et al., 2004a). However, no common domain has yet been identified which is capable of conferring scavenger activity on these proteins (Gough and Gordon, 2000). Therefore, SR is a functional family of structurally unrelated receptors; while SRCR-SF is a family of structurally related but not functionally related proteins.

Actually, more than 30 proteins form SRCR-SF, but this number is constantly and progressively expanding (Sarrias et al., 2004a).

Most members of the SRCR-SF have been described in mammals as: man, mouse, rat, pig, cow, rabbit, sheep; but some of these receptors have also been detected in other species of vertebrates such as birds, amphibians and fish; some also in invertebrates such as sponges, echinoderms, insects and even in algae (Wheeler et al., 2008).

These facts indicate that this superfamily is really very old and preserved on the whole phylogenetic scale.

1. 2 Structure of super family SRCR-SF

SRCR domains consist of approximately 90 to 110 amino acids and are characterized by the fact that their cysteine content is quite high and well defined (Resnick et al., 1994; Sarrias et al., 2004a).

Based on the characteristics of individual SRCR domains, there are two different class for SRCR-SF members. Class A members differ because they are encoded by at least two exons and have six cysteine residues. Class B members are encoded by a single esone and have eight cysteine residues (Martinez et al., 2011). *(Fig. 1 and Fig. 2)*

There is no evidence to date of an SRCR-SF protein that simultaneously contains domains that are part of both cases, A and B.

By the analysis of aminoacid sequences, it is known that: despite all, SRCR domains sharing different levels of homology, the relative position of cysteines is well preserved within each domain, as well as the pattern of disulfide bonds (Resnick et al., 1996).

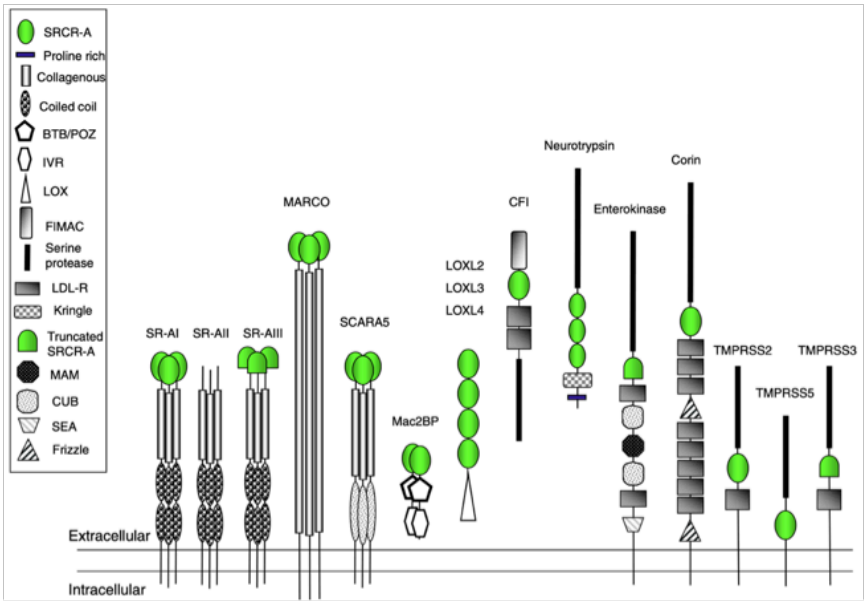


Figure 1 :Structure of Human group A SRCR members (Martinez G.V.et al.,2011)

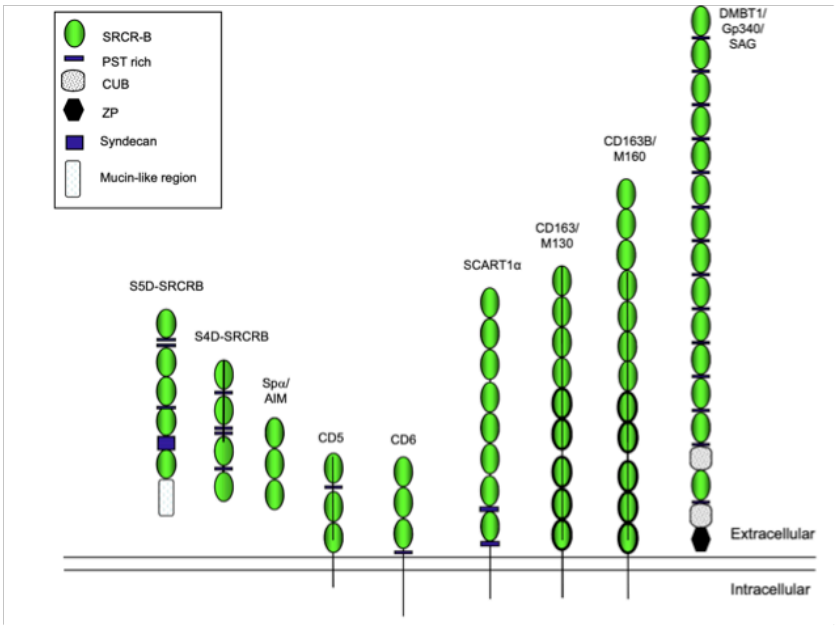


Figure 2 : Structure of Human group B SRCR members (Martinez G.V.et al.,2011)

It is interesting to note that, from the literature to date, the basic three-dimensional structures of the SRCR domains of both classes seem to be perfectly overlapping and therefore indistinguishable. This raises the question of whether the division into two different classes, from a structural point of view, is really useful (Martinez et al., 2011).

1. 3 Functions of super family SRCR-SF

Within the SRCR superfamily are located secreted and/or membrane-bound proteins; a diverse group of proteins involved in different functions.

A single common biological function attributable to the sole SRCR domain has not yet been determined.

To date, it is known that many SRCR domains are needed to mediate protein-protein interactions, such as those described for CD6 and ALCAM (Bowen et al., 1996) or CD163 (Kristiansen et al., 2001). Some studies report that many proteins belonging to this superfamily act as "pathogenic sensors" and represent a first line of defence against inflammation (Aruffo et al., 1997). Some of these proteins appear to be able to modulate the immune response, epithelial homeostasis, stem cell biology, and tumour cell development. Other proteins of the superfamily SRCR appear to be involved in a number of disabling diseases such as atherosclerosis, Alzheimer's disease and autoimmune diseases. (Graversen et al., 2015).

So all these proteins seem to contribute significantly to the regulation of the internal environment of organisms, maintaining the homeostasis of tissues.

Although further research on the functional aspects of the SRCR-SF is still ongoing, their involvement in various physiological and pathological situations must be considered an attractive targets for diagnostic and therapeutic studies (Martinez et al., 2011).

1. 4 The Human Hemoglobin receptor: CD163

The human hemoglobin receptor, identified by the name of Cd163, but also as M130, has been described as a type I transmembrane glycoprotein, of 130KDa as molecular weight.

Glucocorticoids regulated and expressed mostly in monocytes and macrophages (Pulford, et al., 1992 ; Zwadlo et al., 1987).

From the structural point of view, CD163 is composed of a set of 9 SRCR domains repeated in tandem connected, through a small transmembrane domain to an intracellular cytoplasmic tail to the C-terminal (Law et al., 1993).

A very similar structure had been seen during the study of the protein CD163-L1 (Gronlund et al., 2000). The only structural difference between CD163 and CD163-L1 is that it has 12 SRCR domains. It is thought that CD163 may result from the loss of the first three domains by CD163-L1 (Gronlund et al., 2000; Stover et al., 2000).

CD163 is class B of the superfamily SRCR-SF, in fact, the 9 SRCR domains are encoded by a single exon and contain 8 cysteine residues.

Four of these extracellular domains, in particular the domains 2,3,7 and 9, have bonding sites preserved for calcium consent, which are considered very important for the interactions of this protein with different ligands. On the other domains, however, were found mutations not conservative of essential residues, in the sites where you should have found the link sites for the calcium consent.

Several isoforms of CD163 have been described due to the fact that this protein is subjected to alternative splicing.

In particular, regarding the extra cellular region, two isoforms deriving from alternative splicing have been reported: the first is a truncated and secreted form, generated by a stop codon after the first SRCR domains; The second is a protein that has 33 additional amino acids, which form a linker between domain 5 and 6 SRCR (Law et al., 1993).

With regard to the region of the CD163 cytoplasmic tail is concerned, three isoforms have been described, of different lengths, deriving from alternative splicing. The most abundant form has about 49 amino acids, while the other two are longer, respectively 84 and 89 amino acids (Law et al., 1993; Ritter et al., 1999).

However, all three isoforms contain the first 42 amino acids that have target sequences for phosphorylation of creatine kinase and protein kinase C (Van Gorp et al., 2010). Finally, there is a further variant : in addition to the common CD163, there is a soluble cytoplasmic version of this protein, identified as sCD163 (Droste et al., 1999). This isoform is not attributable to alternative splicing phenomena, in fact it is generated thanks to a proteolytic cleavage of CD163 linked to the membrane, by the protease ADAM17 (Etzerodt et al., 2010). Despite the

cleavage made by ADAM17, sCD163 has a molecular weight such as to suggest that it contains almost all the extracellular portion of CD163 bound to the membrane (Moller et al., 2002).

The physiological role of this variant is not yet clear, but it seems to have a minor affinity for the typical ligands of CD163 linked to the membrane (Moller et al., 2010). However, it appears to be particularly implicated in anti-inflammatory processes caused by autoimmune infections and disorders (Hogger and Sorg, 2001; Van Gorp et al., 2010).

1. 5 Expression of CD163

The expression of CD163, until now, is documented only in macrophage or monocitarian cells.

In particular, CD163 is more expressed in liver macrophages (in particular in Kupffer cells) (Pulford et al., 1998), in the spleen, lung and bone marrow. On the contrary, it has been found weakly expressed in the islands of Langerhans (Kodelja et al., 1994) and in dendritic cells (Maniecki et al., 2006) *(Fig. 3)*.

The variability of CD163's expression raised questions about the effectiveness of the different antibodies used to detect it. Indeed, it was observed that the use of EDTA as a coagulating agent interfered with the binding of some antibodies to CD163; other antibodies, however, were unable to bind the scavenger receptor when it was bound to the membrane, since antibody recognition sites were inaccessible (Maniecki et al., 2011).

The use of monocyte cell cultures for *in vitro* studies has allowed to evaluate the significant increase in CD163 expression (Zwadlo et al., 1987) and to study its regulation by specific factors. It was therefore studied how, the expression of CD163 in cell cultures of mature monocytes and macrophages, both Up-regulated by the administration of hemoglobin or free eme group, glucocorticoids, interleucine-6 and interleucine-10 (IL-6 and IL-10); on the contrary the administration of lipopolysaccharides (LPS), Tumor Necrosis factor (TNF), Interleucin-4 (IL-4) and the stimulating factor the colonies of granulocytes-macrophages causes a down-regulation (Etzerodt et al., 2013).

In vivo, the expression of CD163 is often influenced by inflammatory phenomena (Moller, H.J. 2012).

High CD163 expression has been found in macrophages within the microglial region of men with Alzheimer's bites, but also in macrophages of the brainstem of people with Parkinson's disease. It is unclear whether the macrophages examined were residents or infiltrations by systemic macrophages within brain regions (Pey et al., 2014).

Also in the case of people with HIV, an increase in CD163 expression was detected at inflammation sites (Tavazzi et al., 2014).

The expression of CD163, as well as in cells of the monocitarian and macrophage line has been detected, thanks to the presence of antigens, also on cancer cells (Shabo et al., 2011).

The hypothesis, now demonstrated by *in vitro* studies, was that cancer cells were able to merge with macrophages, thus leading, thanks to their ability to reproduce much faster than normal cells, an increase in the expression also of CD163 (Larizza et al., 1984; Munzarova et al., 1992; Maniecki et al., 2012; Chakraborty et al., 2001; Huysentruyt et al., 2008).

It has also been evaluated how these cancer cells "fusion" showed a higher metastatic potential than normal cancer cells, influencing the survival rate for the cancer categories to which they have been associated (Pancione et al., 2014); these include: breast cancer (48%) (Shabo et al., 2008), bladder cancer (39%) (Maniecki et al., 2012), rectal cancer (23%) (Shabo et al., 2009) colon cancer (20%) (Shabo et al., 2014).

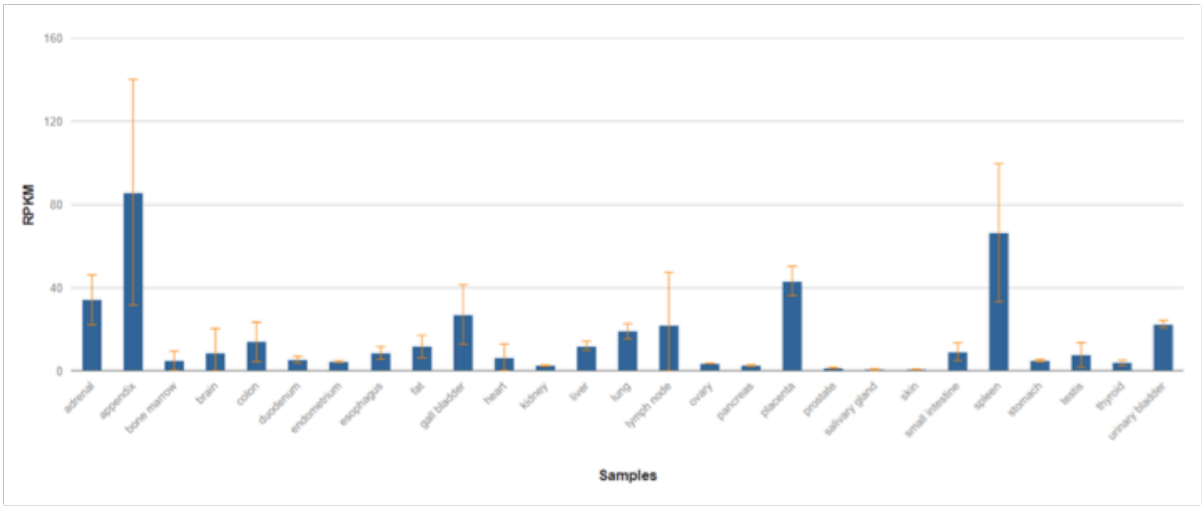


Figure 3 : Graph report of Cd163 tissue expression

1. 6 Functions of CD163

CD163 is known as the scavenger receptor of the haptoglobin-hemoglobin (Hp-Hb) complex (Kristiansen et al., 2001).

The main role of CD163 is the detoxification of the organism from the oxidative potential of the heme group contained within hemoglobin.

Hemoglobin is released, following the intravascular hemolysis of red blood cells, into the circulatory stream. This process is physiological in nature, about 10 % of the red blood cells are degraded daily by the human organism; about one gram of hemoglobin is detoxified by CD163 every day (Garby et al., 1959 ; Garby et al., 1959; Thomsen et al.,2013) contributing to maintaining the balance of cellular homeostasis.

In case of autoimmune infections or diseases, the hemolysis process becomes pathological, accelerating the degradation of red blood cells and hemoglobin by CD163. This receptor therefore plays a very important role in the prevention of the potential toxic effects of the heme group present within hemoglobin during a hemolytic disease.

To date, CD163 shows a no relevant binding affinity for free hemoglobin in the blood. Infact, thanks to the SRCR 2 and 3 domains and the presence of calcium ions, CD163 is able to bind a specific region of haptoglobin, a protein produced by the liver and released within the bloodstream (Madsen et al., 2004). The importance of haptoglobin is that it forms a complex with a molecule of hemoglobin (Kristiansen et al., 2001). It is known that the Hp-Hb complex is one of the most stable complexes in human plasma, thanks to a strong protein-protein interaction.

Once CD163 binds to the Hp-Hb complex, the neoformed ternary complex, it is endocytosed within macrophages (Schaer et al., 2006). At this point the single proteins forming the ternary complex are released as a result of a lowering of the pH and the concentration of calcium within the macrophage (Nielsen et al., 2013).

Hemoglobin is degraded by specific lysosomal enzymes and the heme group released into the cytoplasm. The EME group will be transformed into biliverdin, CO and iron by eme oxygenase-1 (HO-1) (Nielsen et al., 2010) *(Fig. 4)*.

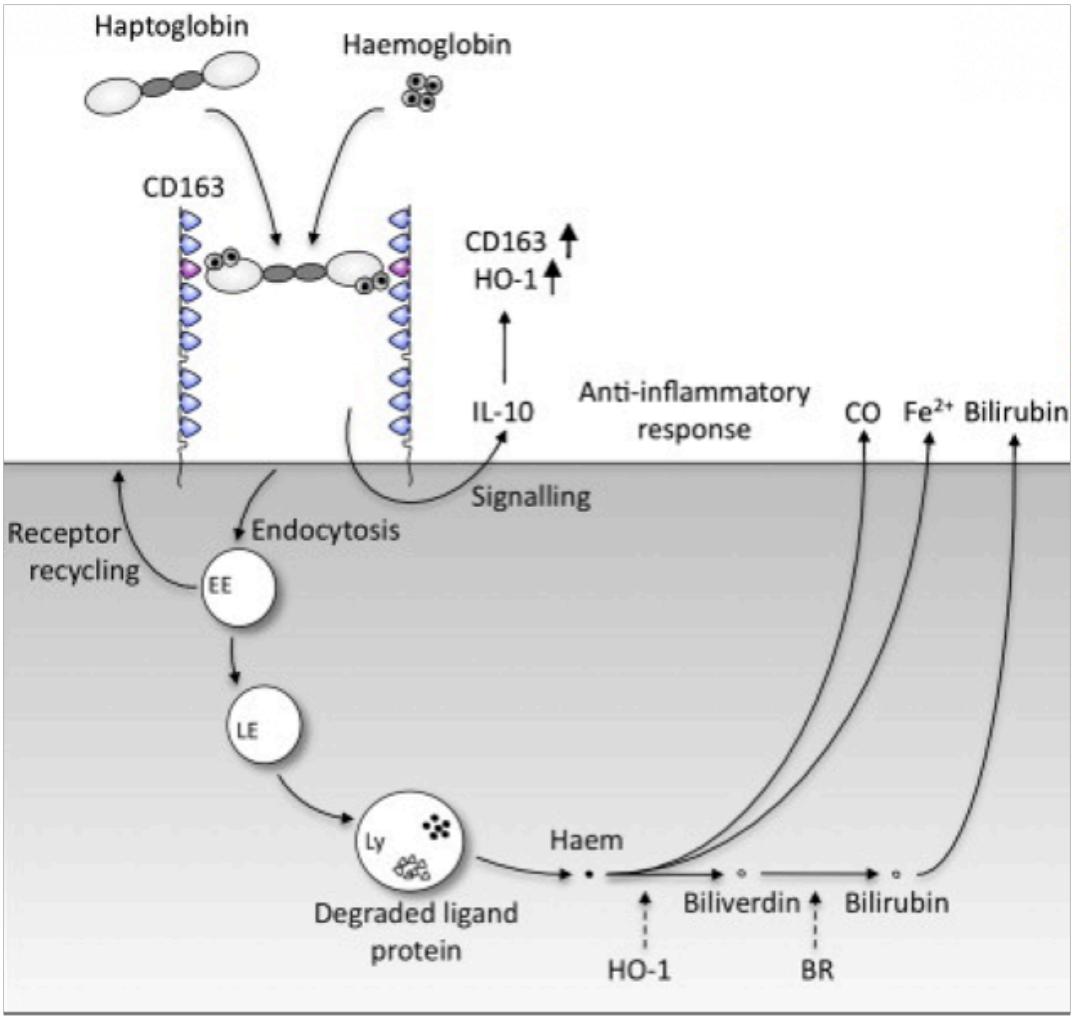


Figure 4: Representing mechanism function of CD163 in macrophage (Gorp V.H. et al.; 2010)

These elements, deriving from the degradation of heme group of hemoglobin contribute, notoriously, to increase the anti-inflammatory response (Grochot-Przeczek et al., 2012) mediated by the immunoregulatory cytokines. In particular, Interleukin-10 (IL-10) seems to be closely related to both CD163 and HO-1 regulating its expression and activity during pathological inflammatory phenomena, favouring the capture of excess hemoglobin and detoxification of the heme group (Graversen et al., 2002;Philippidis et al., 2004).

Parallel to this, it has been hypothesized that the elimination of excess heme group from the bloodstream can help the anti-inflammatory response, since it would reduce the availability of iron to pathogens (Madsen et al., 2004 and Weaver et al., 2006).

During particular infectious phenomena, it has been seen that in case of a reduced presence in the circle of the haptoglobin, the CD163 shows a weak affinity also for the single free hemoglobin (Schaer et al., 2006).

In addition to its "detox" function towards hemoglobin, it has been studied that, as happens for other members of the superfamily SRCR-SF, CD163 is able to bind both Gram-positive and Gram-negative bacteria (Fabriek et al., 2009). This bond triggers the production of pro inflammatory cytokines, but does not appear to be at the basis of phagocytosis of the bacterium by the macrophage. For this reason, this receptor has been attributed to the role of "pathogen sensor" (Fabriek et al., 2009).

Except for the variant sCD163: it seems that it is able to bind some bacteria, by means of the fibronectin present on the bacterial surface. In this way sCD163 would effectively strengthen the phagocytosis of bacteria by macrophages by going to protect the organism from infection (Kneidl et al., 2012).

Since then, evidence has been recorded that CD163 is able to also bind some viruses. In particular, has been researched that African swine fever virus (ASFV) uses CD163 as a target molecule for the admission of viral particles into the host. Thanks to the use of a specific monoclonal antibody, this interaction can now be inhibited (Sanchez-Torres et al., 2003). Swine Reproductive Syndrome (PRRSV) virus infection also appears susceptible to CD163 linkage (Calvert et al., 2007). It seems that this link not only favours the viral infection by PRRSV, but also influences its replication efficiency (Patton et al., 2009), this hypothesis was confirmed by a study in which it is evidenced that CD163 recombined, without its cytoplasmic tail, promotes viral replication of PRRSV to a greater extent than complete CD163 (Lee and Lee, 2010).

Contrary to what has been described for bacteria, in which CD163 seems to play a role of marker and/or defence; it is clear that with regard to viruses, it seems to promote the cellular infection (Lee and Lee, 2010).

It is interesting to note that unlike the link with the Hp-Hb complex in which the SRCR 2 and 3 domain were involved, the binding sites for bacteria and viruses were located respectively on the second and fifth dominoes SRCR (Bikker et al., 2004; Van Gorp et al., 2010), suggesting a specific bond for each domain.

2. Materials and methods

The gene encoding the ectodomain of human CD163 was extracted by PCR from the plasmid pAcGP67A-TfR (Addgene, Cambridge, MA) and BamHI / EcoRI inserted using the Gibson cloning method and fused at the 3' of the Kozak sequence of the pα–H mammalian expression vector modified by the addition of the hydrophobic leader peptide from the baculovirus protein gp67 (pα–H BiP). An octa-histidine tag was also placed at the C-terminus of the CD163 gene.

This work was carried out in collaboration with colleagues from the biochemistry institute of the Sapienza University of Rome.

The recombinant plasmid containing the gene of interest was propagated in E. coli DH5 alpha cells. Subsequently, a transient transfection in HEK293s mammalian cells was tested first, followed by a transfection in insect cells, to evaluate the difference in yield in the expression of the protein of interest.

2.1 Mammalian cells expression system

2.1.1 The HEK293s Cells

Human embryonic kidney cell line 293, also recognized as HEK 293, HEK-293, 293, or less precisely as HEK cells, is a specific cell line originally derived from human embryonic kidney cells taken from an aborted female fetus.

The studies that led to the isolation of this cell line date back to the early 1970: Alex van der Eb and Frank Graham studied and cultivated these cells for the first time in their laboratory in Belgium.

The cells are called "HEK" because they originated in human embryonic kidney cultures, while the number 293 comes from Graham's habit of numbering his experiments: the original HEK 293 cell clone was from his 293rd experiment.

HEK 293 cells were adapted to grow both in suspension culture and on plastic plates in 1985 (Stillman BW, Gluzman Y 1985).

These cells are widely used in research as they prove to be particularly easy to cultivate and fairly efficient in transfections with specific constructs for the production of recombinant proteins. They are used above all, for the expression of large proteins, or with more functional domains, for which the use of bacterial cell cultures would not be desirable, as they would not be able to assemble the entire protein and perform the post-translational changes necessary for its functioning.

HEK 293 cells are also used by the biotechnology industry to produce therapeutic proteins and viruses for gene therapy.

The Human Embrionic Kidney 293 Suspension (HEK 293s) cells used in this study were cultured in FreeStyle™ 293 medium supplemented with 0.5% Fetal Bovine Serum (FBS; Gibco Life Technologies). All cell cultures were maintained at a constant temperature of 37 ° C, with a rotation of 100 RPM in the presence of 8% CO_2.

2.1.2 Expression in HEK293

The colonies resulting from the transformation of the recombinant plasmid with the gene of interest were used for the transient transfection of HEK293s cells, through the use of polyethyleneimide (PEI). PEI is a cationic polymer capable of forming complexes with DNA. These complexes are incorporated within endosomes to facilitate the entry of recombinant DNA into the cell. Once inside the cell, thanks to their positive charge, the endosomes recall chlorine ions and water which cause the endosome wall to break. In this way the DNA is released into the cytoplasm.

The DNA-PEI complex was prepared by combining a first solution containing FreeStyle medium and recombinant plasmid, with a second solution containing FreeStyle medium and a

1mg/ml PEI solution. The ratio of DNA to PEI used is 1: 4. The relationship between FreeStyle medium and DNA is dependent on the volume of cell culture to be analyzed.

Each solution was left to incubate individually at room temperature for 15 minutes before being combined with the other. The mixture obtained at the end was left to incubate for another 30 minutes at room temperature. The solution was then carefully added to the flasks containing HEK293s cell cultures at a concentration of 1mln of cells / ml.

All these operations were performed in a sterile environment, under a laminar flow extractor hood.

At the end of the procedure described above, the cell cultures were incubated for 72 hours at a constant temperature of 37 ° C, with a rotation of 100 RPM and in the presence of 8% of CO_2.

2.1.3 Expression with expifectamine kit

The Expi293TM Expression System uses suspension adapted human embryonic kidney (HEK) cells similar to common expression methods.

Transient transfection performed with the Expi293TM kit allows for much more efficient expression yields than the methods normally used. This is thanks to the use of the reagents contained in this kit. The Expi293TM reagents are the result of a patented chemical formulation, free of serum or proteins of animal origin, based on cationic lipids and with specialized transfection stimulators, designed for supported high intensity transfections. The components of the Kit work together to generate protein yields that are 2 to 10 times higher than the expression systems normally in use. Expression yields greater than 1 gram per liter of transfected culture have been demonstrated for some proteins.

The Expi293TM Kit uses the same transient expression protocols typically used with HEK293 suspension culture systems. In fact, a solution containing plasmid DNA and OptiMeM Medium is mixed with a solution containing Expifectamine293 reagent and Optimem Medium. The final solution is then added to the cell culture, the concentration of which, however, is doubled compared to the normal protocols for transient transfection described above.

The cell culture is then incubated for 24 hours at a constant temperature of 37 ° C with 8% CO_2 in an orbital shaker incubator, with a rotation of 100 RPM.

The day after this procedure, the culture is taken from the orbital shaker and expifectamine enhancer1 and expifectamine enhancer2 are inoculated. These reagents chemically stimulate the expression of the protein of interest by the cells in culture.
The cell culture is placed back in the incubator for another 72 hours.

2.1.4 Purification of Recombinant Cd163

Within pα – H BiP-Cd163 there is a sequence for secretion; this means that the expressed protein is secreted and released by the cells into the culture medium.
For this reason, the cell culture, at the end of the transfection, was subjected to a centrifuge of 800xg for 20 minutes. The cell pellet, given by the fragmented cells, is discarded; instead the supernatant containing the protein to be purified was processed.
First, it was mixed with a buffer, to restore the correct physiological pH typical of Cd163, i.e. pH 8.
Different samples of recombinant CD163 were purified with three different buffers, such as: 1) Sodium phosphate 2) Tris-HCl 3) Hepes.
Once the pH of the sample was stabilized, the purification protocol was the same for each buffer used.
The protein was purified by affinity chromatography on Ni-NTA agarose resin (Quiagen). The binding of the histidine tail to the resin was carried out by keeping the two components in solution (resin and sample) in incubation overnight on a rotating wheel at 4 ° C. The next day the column was washed with 30 ml of washing buffer (1) 50 mM NaPb, 200 mM NaCl; 2) 50 mM Tris-HCl, 200 mM NaCl; 3) 50 mM Hepes, 200 mM NaCl) and eluted in various fractions (1 ml each) of elution buffer (1) 50 mM NaPb, 200 mM NaCl; 500 mM imidazole 2) 50 mM Tris-HCl, 200 mM NaCl; 500 mM imidazole 3) 50 mM Hepes, 200 mM NaCl, 500 mM imidazole).
The eluate was dialyzed overnight at 4 ° C against the dialysis buffer (1) 50 mM NaPb, 150 mM NaCl; 2) 50 mM Tris-HCl, 150 mM NaCl; 3) 50 mM Hepes, 150 mM NaCl) to eliminate imidazole, an annoying molecule for subsequent absorbance spectra.
The concentration of CD163 was determined by detecting the absorbance of the sample at 280 nm and its purity was evaluated by SDS-PAGE (Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis), followed by the Coomassie-blue stain.

2.1.5 Salting out method With Ammonium sulfate

The salting out technique allows proteins to be precipitated in solution without denaturing them through the use of precise salt concentrations: The water will solvate the added salt and at the same time decrease the solvation of the protein. The protein molecules, therefore, will interact with each other, forming aggregates that will precipitate. This technique was used to reduce the sample volume to be purified.

On the basis of the molecular weight of the expressed protein and the volume of sample obtained, 60% by weight of powdered ammonium sulphate was added to the supernatant already buffering. The mixture was first dissolved with the help of a magnetic plate, then it was incubated at room temperature for about 2 hours to favor the precipitation of protein aggregates. The sample was then centrifuged at 6500xg for 30 minutes and the pellet was resuspended with 5ml of 50mM Hepes.

The obtained solution was dialyzed over night to eliminate the ammonium sulphate residues with 50mM Hepes, 200mM NaCl.

The next day, the sample was purified as described above.

2.2 Baculovirus-insect cell Expression system

2.2.1 Sf9 / Hi5 cells

The Sf9 cell line originates in turn from the IPLBSF-21 cell line. Both of these cell lines were initially isolated from pupal ovarian tissue of the worm Spodoptera frugiperda. They are widely used for the expression of recombinant proteins through the use of the baculovirus system. They are characterized by the fact that they are able to grow even in the absence of serum and can be grown attached or in suspension. Furthermore, the dimensions of these cells are quite small and regular, which makes them easier to manipulate than those commonly used. For all these reasons they are particularly suitable for biomolecular studies on recombinant proteins.

The Hi5 cell line, on the other hand, originates from cabbage looper ovary cells, Trichoplusia-ni. This cell line is characterized by a very rapid cell reproduction (less than 24 hours). Like sf9 it can grow in the absence of serum and can be grown in both adherent and

suspension cultures. Unlike Sf9, however, it has a greater expression of secreted proteins (up to 5/10 times greater). For this reason Hi5 cells are excellent for expressing recombinant proteins.

2.2.2 Expression test in Baculovirus System

In order to express the protein of interest in insect cells, the Cd163 gene was inserted into a Bacmid using the LIC Ligation protocol.

The design of special LIC primers for CD163 made it possible to avoid the use of restriction enzymes, since the primers allow the amplification of only the region of interest within the plasmid.

The PCR reaction was performed with the addition of different percentages (2%, 5%, 10%) of DMSO to favor the KOD polymerase reaction.

The obtained amplification product was purified both from solution and from gel, with the same kit (Omega bio-tek).

At this point, the LIC, Annealing and transformation protocol by *Bruni R. and Kloss B.* (2013) was performed, using DH10Bac cells, that is cells suitable for the growth of a Bacmid, which is larger than that of a common bacterial vector.

The grown colonies were purified using the E.Z.N.A Plasmid DNA Mini Kit-1 (Omega Bio-Tek) kit.

The purified DNA was first quantized by using a Nanodrop and subsequently sequenced at the GENE WIZ.

After confirming the presence, within the recombinant bacmid, of the Cd163 gene, the efficiency of infection was tested with a mini-scale transfection.

Approximately 3 mL of cell culture of Sf9, at a concentration of 0.8 $x10^6$ cells / mL was sampled for this assay.

Cellfectin® II was used as the transfection reagent. This reagent is a patented cationic lipid formulation that offers the highest transfection efficiencies and protein expression levels on the widest variety of insect cell lines.

Similarly to the transfection protocol for HEK293 cells, two solutions were mixed: one solution containing the culture medium (100 µl) with 5µg of bacmid DNA, while another containing the culture medium (100 µl) and 10 µl of Cellfectin® II. The solutions were

incubated for about 15 minutes at room temperature; then they were mixed together and incubated again for another 15 minutes at room temperature.

The solution obtained from the mixture of the first two was then gently administered to the cell culture used for transfection.

The cells thus treated were incubated at 27 ° C for 96 hours.

The transfection efficiency was checked daily by using a fluorescence microscope, thanks to the presence of a gene for GFP present within the Bacmid.

The small-scale test was used to evaluate the efficiency of Baculovirus and is defined as "V_0".

At the end of the 96 hours of incubation, the 3 mL of V_0 were centrifuged at 800xg for 10 minutes. The supernatant was used to infect 25 mL of Sf9 cells at a concentration of $1x10^6$ cells / mL. Thus began the viral propagation phase, also known as V_1.

The cell culture of V_1 was incubated as previously: for 96 hours at 27 ° C. At the end of the incubation, the sample was centrifuged at 800xg for 10 minutes and the supernatant was stored in dark falcons (since the virus is prone to deterioration if exposed to light) at 4 ° C.

Part of the V_1 obtained was used to continue viral propagation, namely V_2, in order to always have a new vector available.

Another part was used to test the expression of Cd163 in Sf9 and Hi5.

Different concentrations of V_1 were tested in parallel: as regards Sf9, 0.5% and 1% of V_1 were used to infect them; as regards Hi5 cells, 3% and 6% were used. In order to evaluate differences in efficiency in protein production.

The expression was carried out for a duration of 96 hours at 27 ° C.

2.2.3 Western Blot analysis

The cell cultures of Sf9 and Hi5 previously used to test the expression of recombinant CD163 were collected and centrifuged at 5000xg for 5 minutes. The supernatant was discarded, while the pellet was washed three times with 1 mL of PBS solution. Finally, the samples were centrifuged again at 5000g x 5 minutes to remove the excess of PBS solution and frozen at -80 ° C. The following day, the frozen pellet was resuspended with 150 µl of Resuspension Buffer containing: PBS solution, 2% DDM, Dnase1 and AeBSF.

At this point the samples were centrifuged at 20,000g x 15 minutes; the supernatant was mixed with Sample Buffer 2x, boiled for 10 min at 98 ° C and loaded on 12% polyacrylamide gel as standard protocol (Laemmli, 1970).

After electrophoresis, the samples were transferred to a 0.2 μm nitrocellulose membrane (GE Healthcare) using a semi-dry blotting apparatus (Trans-blot SD Semi-Dry Transfer Cell, BioRad). Protein transfer was performed at 15 V for 20 minutes in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol). After blocking non-specific binding with 5% skim milk in TBST Buffer (20 mM Tris, 150 mM NaCl, 0.1% Tween 20) for at least 1 hour at 37 ° C, the blots were probed with con anti-histidine (Thermo Scientific) in a 1: 6000 dilution in 1% TBST-BSA and left to incubate overnight at 4 ° C. The following day the membranes were washed three times with TBST. An IgG-HRP anti-goat mouse antibody (Santa Cruz Biotechnology) was used to wash away the excess anti-histidine. The solution was incubated with the membrane for 1 hour at 37 ° C. At this point, the membranes were washed with TBST four times and subsequently developed with the Western Blotting Enhanced Chemiluminescence Substrate LiteAblot PLUS (Euroclone), according to the manufacturer's instructions.
The images were viewed with the BioRad ChemiDoc Imaging System.

2. 3 Size exclusion chromatography

Size exclusion chromatography (SEC), also called gel filtration, is a technique that allows, thanks to the use of columns with different types of resins, to separate the proteins present in a solution based on their molecular weight. The instrument used (ÀKTA purifier GÈ Healthcare) for this type of technique essentially consists of two pumps, a spectrophotometer (the instrument supplied has a lamp that emits at 280nm), a valve to regulate the flow path inside of the device, a capillary tube loop for the insertion of precise quantities of solution and a chromatographic column to be used.

The proteins were loaded on a Superdex GL 200 Increase 10/300 column (GE Life Science); this column is capable of separating proteins with dimensions between 10 kDa and 600 kDa.
The chromatographic buffer was subjected to filtration and contained 20mM of Tris-HCl pH 8, 150mM NaCl. The column was equilibrated with 3 volumes of this buffer.
The solutions containing the proteins of interest were centrifuged at 10,000 g for 10 minutes, to exclude the presence of any precipitate that could interfere with the analysis. The elution

fractions were collected and those containing proteins, selected based on their absorbance at 280 nm.
Subsequently the absorbance data were confirmed by an SDS-PAGE electrophoresis.

2. 4 Dynamic Light Scattering

The Dynamic Light Scattering was used to observe the stability of the recombinant CD163 complex + the Hp-Hb complex.

The analyzed samples are: Hp-Hb complex at a concentration of 1mg/mL maintained at a temperature of 4 ° C and CD163-Hp-Hb complex, at two different conditions: 1) 0.4 mg/mL and 4 ° C; 2) 0.05 mg/mL and Frozen. The conditions of the analyzed samples are different as they are considered irrelevant since the survey was conducted for the sole purpose of verifying the monodispersity of the ternary complex.
The samples were centrifuged at 10000xg for 15 minutes at 4 ° C and transferred to new eppendorf.
The analysis and study of the data obtained was carried out by the technician responsible for the instrument.

2. 5 Tycho

TychoTM NT.6 is the tool used to perform calorimetry analyzes to determine if the analyzed protein samples are structurally intact and properly folded.

The instrument requires the use of only 10 µl of sample.
Given the information to be sought, it was not necessary to normalize the concentration of the sample analyzed.
The sample was taken, by capillarity from the eppendorf, using special stylets. These were placed in a specific machine housing.
Once the instrument is started, the sample profiles are automatically defined in a few minutes. In particular, the inflection temperature (Ti) is disclosed, given by the ratio between the

fluorescence emission of the residues inside the tryptophan and tyrosine proteins detected at both 350 nm and 330 nm during a temperature scale ranging from 35 ° C to 95 ° C.
The instrument automatically generates easy to interpret diagrams: the higher the inflection temperature, the more stable the analyzed protein is and consequently it is correctly folded.

3. Results & discussions

3.1 Production and purification of recombinant CD163

The recombinant plasmid containing the gene encoding for cd163 was cloned by the research group of the biochemistry department of the *La Sapienza University*, with which this project was carried out in collaboration.

Once received and propagated in competent *E.coli* gc5-α cells, the recombinant pα–HBiP-Cd163 plasmid, a small-scale recombinant protein expression assay was performed using HEK293 cells, to verify the correct production of the protein of interest.
It is known in the literature that this protein does not have a good expression yield as a recombinant protein, for this reason it is usually preferable to extract it from the tissues, however this involves a series of complications, easily avoidable using a recombinant system. The search was therefore made for the best method to produce this protein in a recombinant system by increasing the production yield as much as possible.
From the small-scale expression test (30 ml), it was possible to observe that the protein expresses itself, therefore the cloning of the cd163 gene within the pα–HBiP-Cd163 plasmid was successful.
The cell pellet does not have the band corresponding to the protein of interest, so it can be excluded that it forms inclusion bodies. Flow through and buffers used to wash the chromatographic column were used as negative controls, as the protein does not appear in their fractions. Finally, its post purification product appears perfectly clean and free from contaminants *(Fig. 1)*.

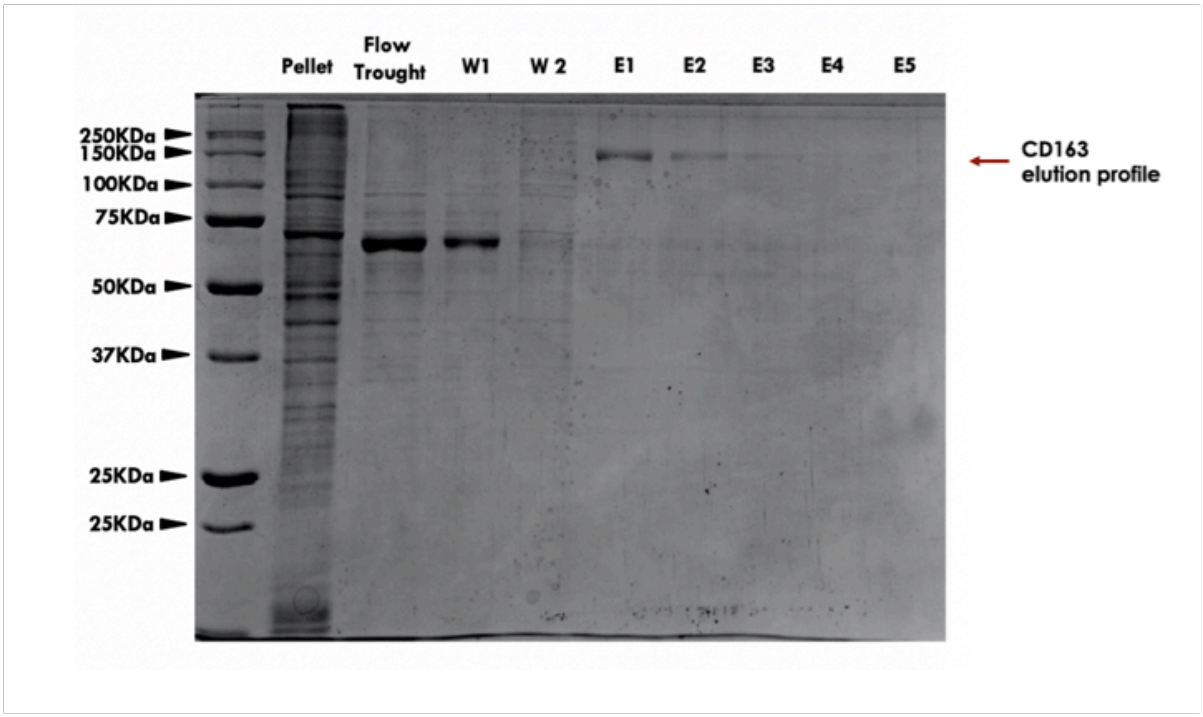


Figure 1 :Purification products of Recombinant CD163 in SDS-PAGE. Pellet; Flow trough: elements not binding to resin; W1: washing 1; W2: washing2; E1,E2,E3,E4,E5: elution fraction.

This means that the choice of a eukaryotic expression system and a plasmid with signal for the secretion of the expressed protein proved to be an optimal choice for the production of recombinant Cd163.

Proceeding on a large scale, the already known problem became evident: the expression yield of this protein is very low. In fact, from about 500 ml of HEK293 cell culture, we obtain a yield equal to 0.3 ug of protein.

The protein is expressed even if the amount obtained is not too high as the eukaryotic expression system does not allow a conspicuous production of recombinant proteins, on the other hand the result described above does not allow the desired analyzes to be carried out.

For this reason, by analyzing the data obtained, a method was sought to optimize the expression of the recombinant protein of interest.

First, a study was conducted on the ratio of transfection agents used, mainly polyethylenimine (PEI). Tests were conducted that involved small-scale expression of Cd163 with different concentrations of PEI.

Since none of these tests showed obvious improvements in the expression yield of the protein of interest, different purification protocols were examined.

In particular, the purification efficiency was compared on the basis of the use of different buffers.

Cell cultures of HEK293 of the same volume were grown, transfected and collected with the same protocol. Subsequently the samples were purified in parallel, with: Hepes, Sodium phosphate, and Tris-HCl. All the buffers listed have been used at pH 8 to ensure the stability of the protein in solution.

At the end of the purification, the corresponding eluates were loaded into an SDS-PAGE, but no buffer used for the purification proved particularly efficient in optimizing the yield of this protein. The first three buffers have a fairly equivalent yield of the protein, even if the Hepes shows a more "compact" elution profile *(Fig. 2)*.

Finally, an attempt was made to precipitate the protein product of the entire cell culture using ammonium sulfate.

Ammonium sulfate is a highly corrosive and dangerous salt for tissues and surfaces, but it is often used for the precipitation of proteins in solution, as it is able to concentrate them without denaturing them (Protein precipitation using Ammonium Sulfate Curr prot protein 2001 May).

The flow through of the cell culture was then collected and 60% by weight of ammonium sulphate was added. The sample was then incubated for 5 hours at 4 ° C, monitoring the pH which was always kept on a neutral value. The sample was then centrifuged in order to recover the precipitated proteins. The resulting pellet was resuspended and dialyzed over night to allow the elimination of the precipitating agent. At this point the purification was continued as per protocol.

Despite the precipitation with ammonium sulphate, and therefore the concentration of all the proteins produced by the eukaryotic cell culture examined, the SDS PAGE gel does not show different results compared to the previously supported tests *(Fig. 2)*.

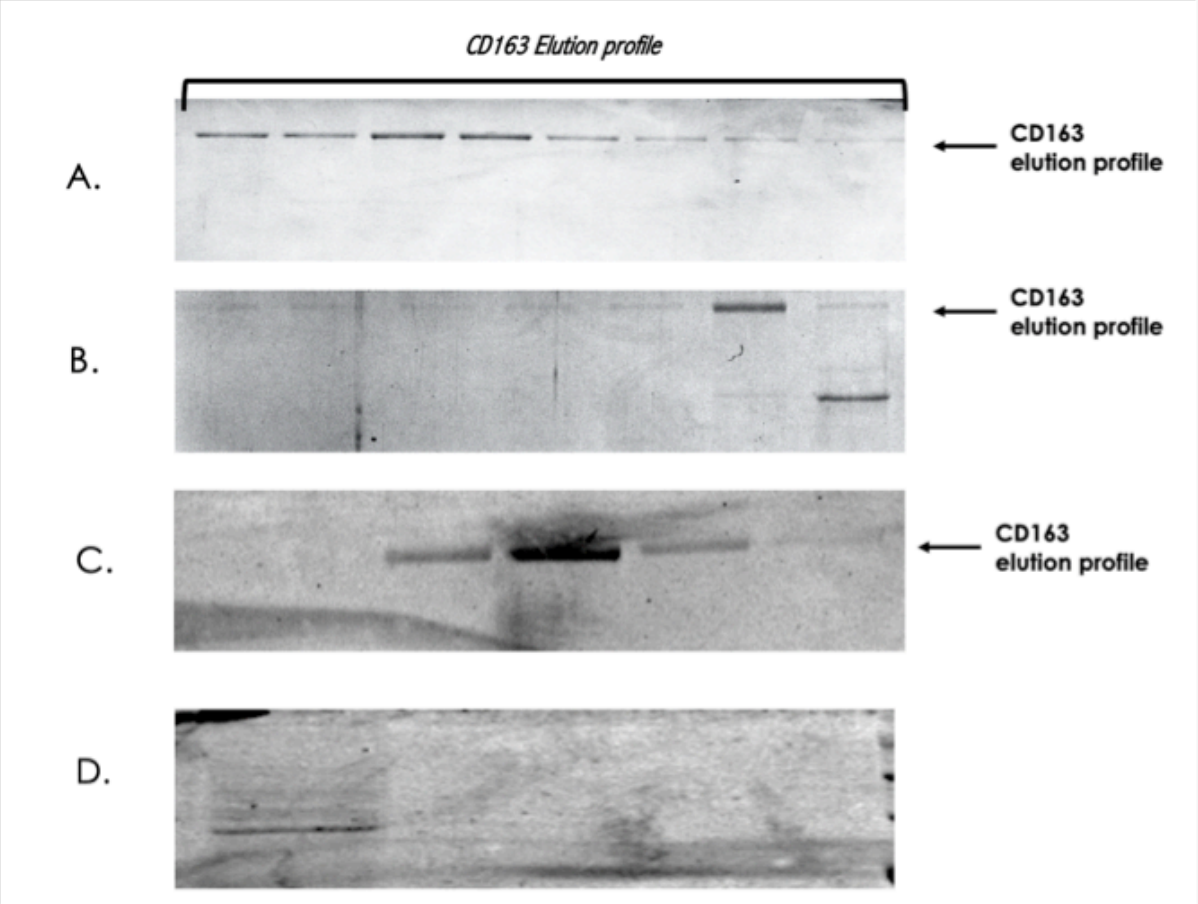


Figure 2 : Elution profile of Recombinante CD163 with different buffer for purification. A.= Sodium phosphate buffer; B.= TRIS-HCl buffer; C.= Hepes Buffer; D.=Precipitation protein with Ammonium posphate.

At this point it was decided to change strategy and re-examine the transfection reaction. For this reason, the reagents used so far, namely PEI, have been replaced with a kit specifically designed and sold by Gibco to increase the yield of transient expression in eukaryotic systems.

At this point it was decided to change strategy and re-examine the transfection reaction. For this reason, the reagents used so far, namely PEI, have been replaced with a kit specifically designed and sold by Gibco to increase the yield of transient expression in eukaryotic systems.

The expifectamine kit (or Expi293TM kit), is a very expensive kit, has been designed and created specifically for the optimization of the expression of recombined proteins in specific

eukaryotic systems such as human kidney cells in suspension (HEK293). This kit is characterized by the presence of the ExpiFectamine TM 293 reagent, based on cationic lipids with specialized transfection enhancers. The components of this reagent work together to generate 2 to 10 times higher protein yields than previous generation expression systems.
The result obtained with the use of this kit was truly amazing. From the same volume of cell culture previously used, from which a rather poor yield of recombinant protein was obtained, more than double, about 0.8 ug of protein was obtained with the Expi293TM kit *(Fig. 3)*.

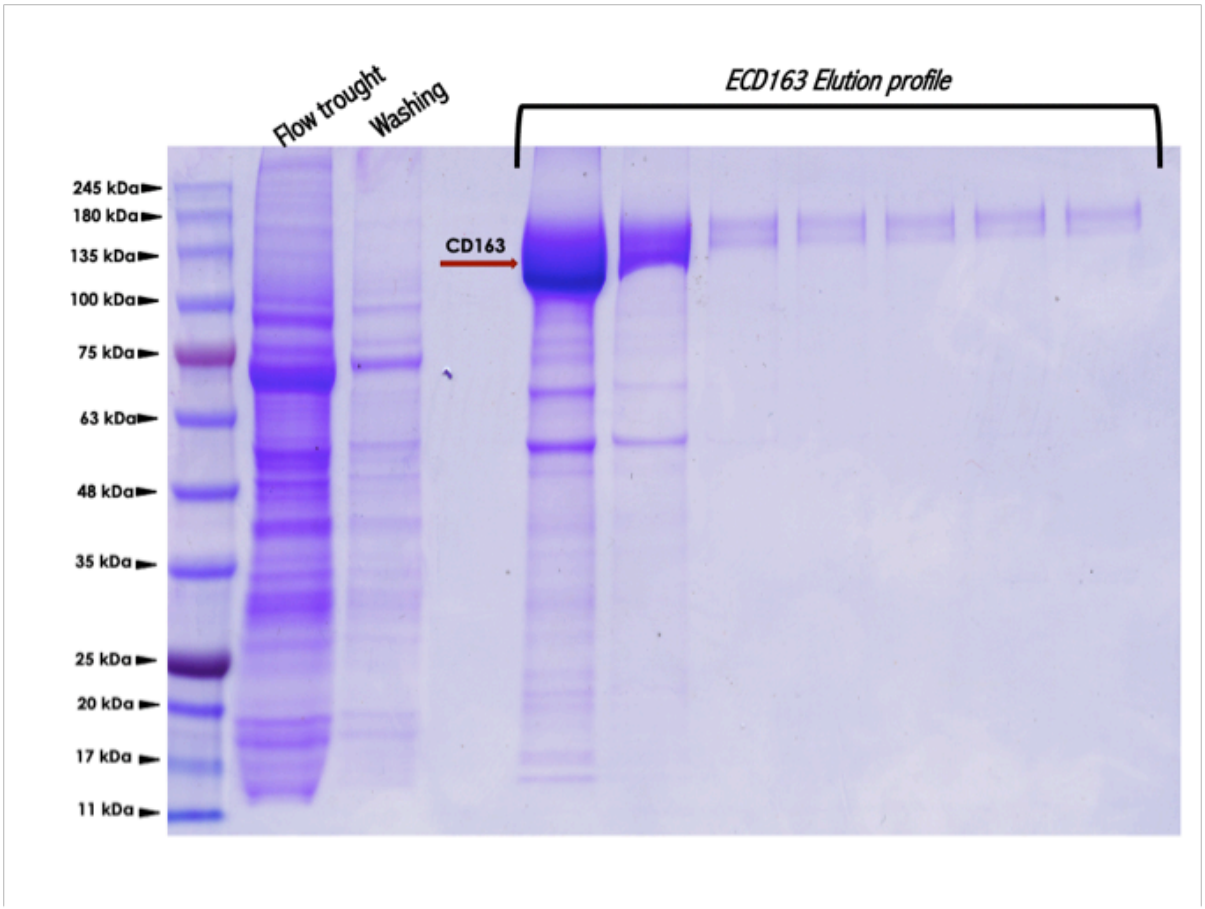


Figure 3: Purification products of Recombinant CD163, expressed with expifectamine kit, in SDS-PAGE.

As an alternative to a commercial kit, which is also quite expensive, it was decided to test another expression system, always eukaryotic, that of insect cells. The protocols used for this viral infection system are, in fact, very well known and used precisely to increase the production yield of recombinant proteins.

To express Cd163 in insect cells, a dedicated expression vector was required, PfastBac, an engineered baculovirus capable of infecting insect cells.

First, primers were designed, starting from the nucleotide sequence of the protein of interest in order to be able to insert it within the PfastBac.

The primers were designed with computer programs such as Serial cloner and Primer Primer.

Once the primers were designed they were purchased by GENE WIZ.

The primers were used to amplify the Cd163 insert.

The amplification was carried out with a highly efficient enzyme (KOD DNA Polymerase High fidelity). To be sure of the best possible yield of this amplification, different concentrations of DMSO (Dimethyl sulfoxide) were tested in parallel, knowing that this additive can favor the denaturation of the DNA double helix and therefore optimize the amplification yield.

As shown by the gel, the insert bands are well amplified and despite the different concentrations of DMSO, there is no real difference between the samples *(Fig. 4)*

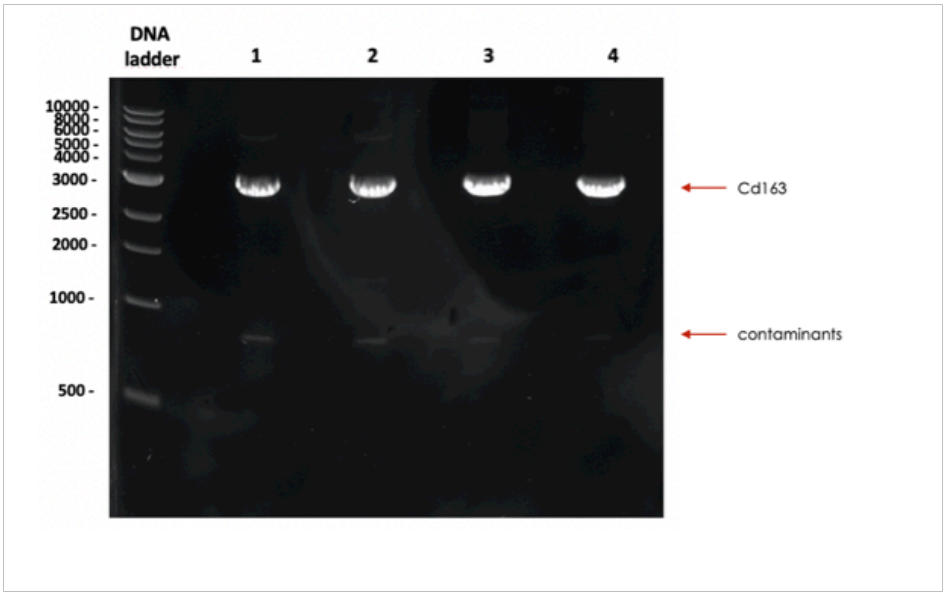


Figure 4 : Agarose gel electrophoresis of the amplification product obtained for CD163. 1=No DMSO in PCR mix; 2= 2% DMSO in PCR mix; 3= 5% DMSO in PCR mix; 4= 10% DMSO in PCR mix.

The PCR product was subsequently purified with the Omega bio-teck Kit first from the solution and subsequently, given the presence of bands indicating contaminants, from gel to obtain a product, as clean as possible and to avoid that some contaminants could hinder the subsequent ligation of the insert with the plasmid.

Once the LIC, annealing and transformation protocol was performed, the clones present on the plates were propagated on a small scale (3ml) and purified with the appropriate kit.
To check that the gene had been correctly cloned within PfastBAC, a control amplification, also known as pic colony, was performed.
The gel shows a precise fragment and in the absence of contaminants, exactly corresponding to the Cd163 insert, in fact it has a size of 3000 bp *(Fig. 5)*.

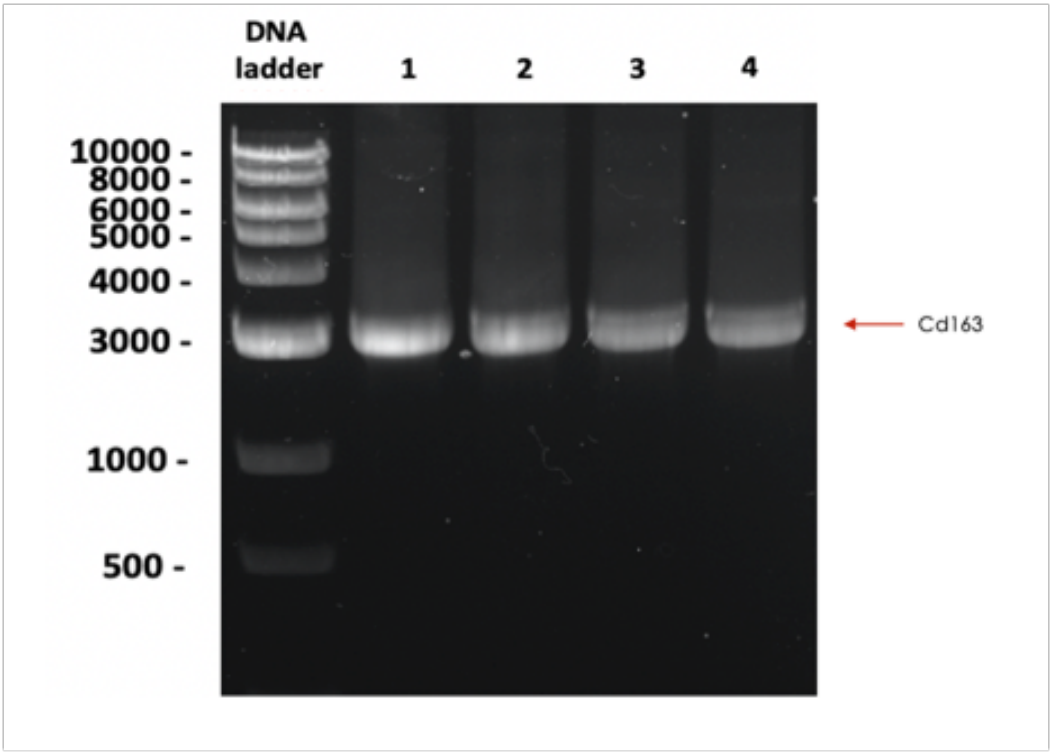


Figure 5: Agarose gel to check for the presence of CD163 in PfastBAC.

Once the concentration of DNA obtained was estimated, the entire nucleotide sequence of the cloned gene was also sequenced to ensure that there were no unwanted mutations inside.
Subsequently, the DNA was transformed into EmBACy cells, competent cells specifically designed for bacmid transformation. Bacmide positive colonies were purified with a specific kit, since bacmides were larger than normal expression vectors.
At this point the recombinant Baculovirus for Cd163 was used to infect a very small volume of SF9 cells (about 3 ml), in order to verify its infection efficiency.

It emerged that the baculovirus engineered with the cd163 gene is particularly efficient, in fact in just three days, cell growth was inhibited with a minimum concentration administered, equal to 1 ug of bacmid DNA *(Fig. 6)*.

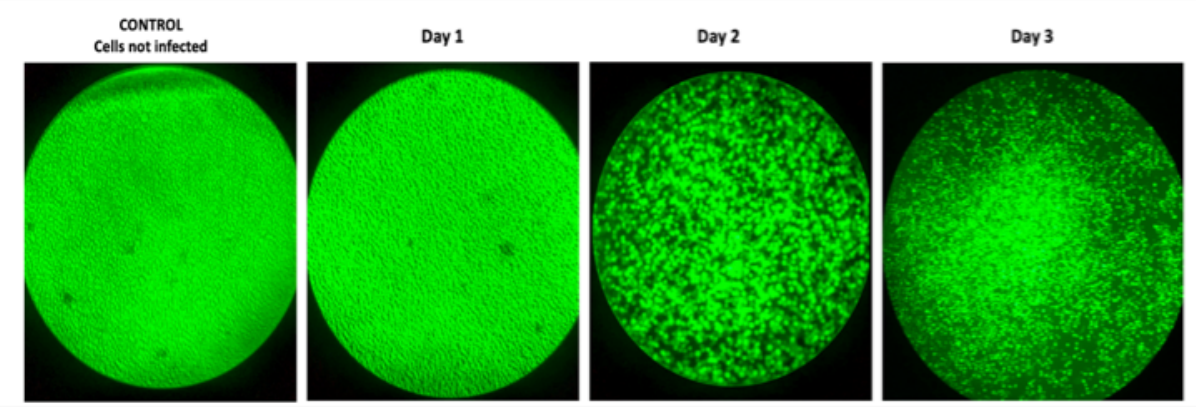


Figure 6 : Verify infection efficiency of Cd163-Bac through a fluorescence microscope.

Subsequently, a further infection was made, in a larger cell volume (250ml), in order to obtain a greater amount of virus that could be reused to infect ever greater volumes of cells.
The first real test of recombinant Cd163 expression was performed on a small scale, since we wanted to compare the expression efficiency of two different insect cell lines. SF9 cells, previously used for viral propagation, and HI5 cells, which are generally used for the expression of recombinant proteins in this system, as they are believed to be more efficient in the expression of the actual protein.
Indeed, the western bloth revelation shows that HI5s allow for significantly greater expression of the protein of interest.
Furthermore, the investigation conducted daily to verify the expression time of Cd163 in this expression system, shows us that it is produced in acceptable quantities only after the 3rd day, as regards HI5. If you want to use the SF9s, however, the incubation should be extended by one day.
Furthermore, a further difference can be noted in the amount of Baculovirus used to infect cell cultures. In fact, as a percentage, it is noted that SF9 cells need a smaller percentage of Baculovirus to express the protein of interest, while Hi5 cells, even if they take a shorter incubation time to express Cd163, however, need a higher percentage of viruses *(Fig. 7)*.

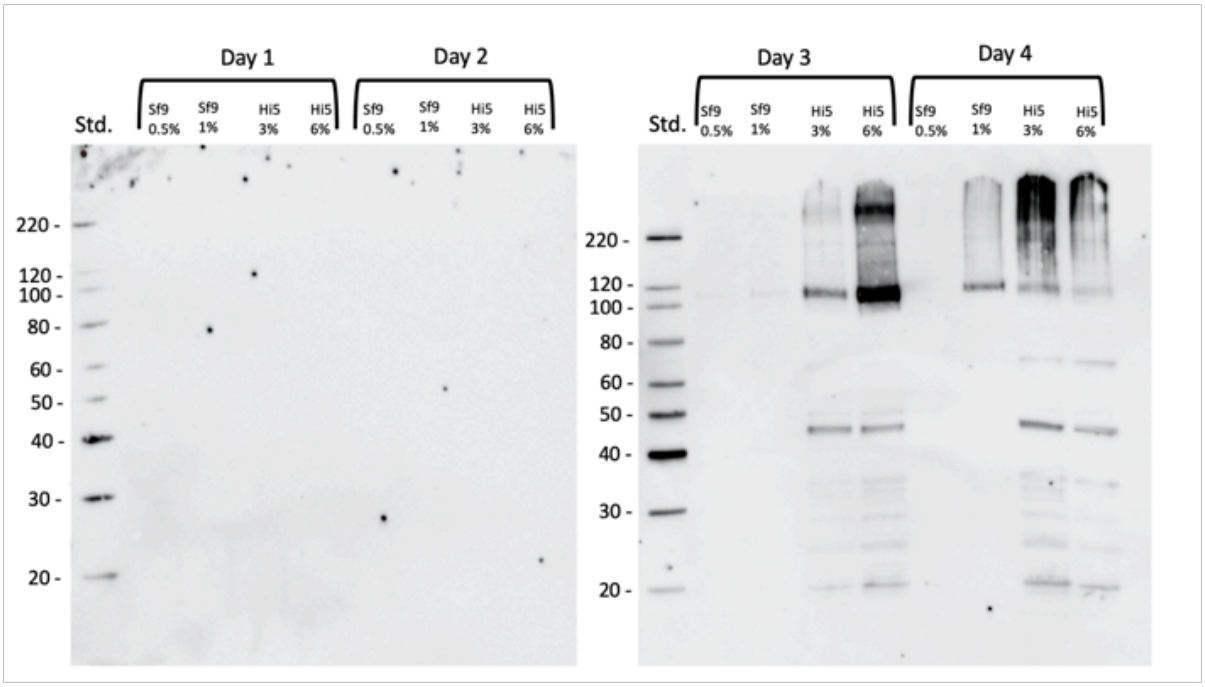


Figure 7 : Western bloth detection of Cd163 expression in Sf9 and Hi5. Expression time (day) and percentage of infecting virus.

Unfortunately, the investigations could not go on due to the outbreak of the coronavirus pandemic. For this reason it is not possible to show results that can correctly compare the different methods used for the expression of recombinant Cd163.

In any case, comparing the band obtained from the Western blots of a sample of about 5 ml of SF9 / Hi5 culture, with those obtained from the previous gels, corresponding to about 400 ml of HEK293s cell culture, we can say that: as an expression system, that of insects ensures a certainly more efficient expression yield, as regards the protein under examination.

3.2 Protein analysis

Despite the problems in optimizing the expression yield of the protein of interest, the amount obtained was sufficient to carry out various analyzes.

First of all a Size Exclusion Cromatography was performed to verify the dimensions and evaluate the purity of the obtained recombinant protein.

The analysis was performed using a Superdex 200 Increase 5/150 GL column, with previously filtered and degassed sodium phosphate / NaCl, as a running buffer.

The peak obtained shows a decidedly clean recombinant protein corresponding to the expected molecular weight (130KDa) *(Fig. 8)*.

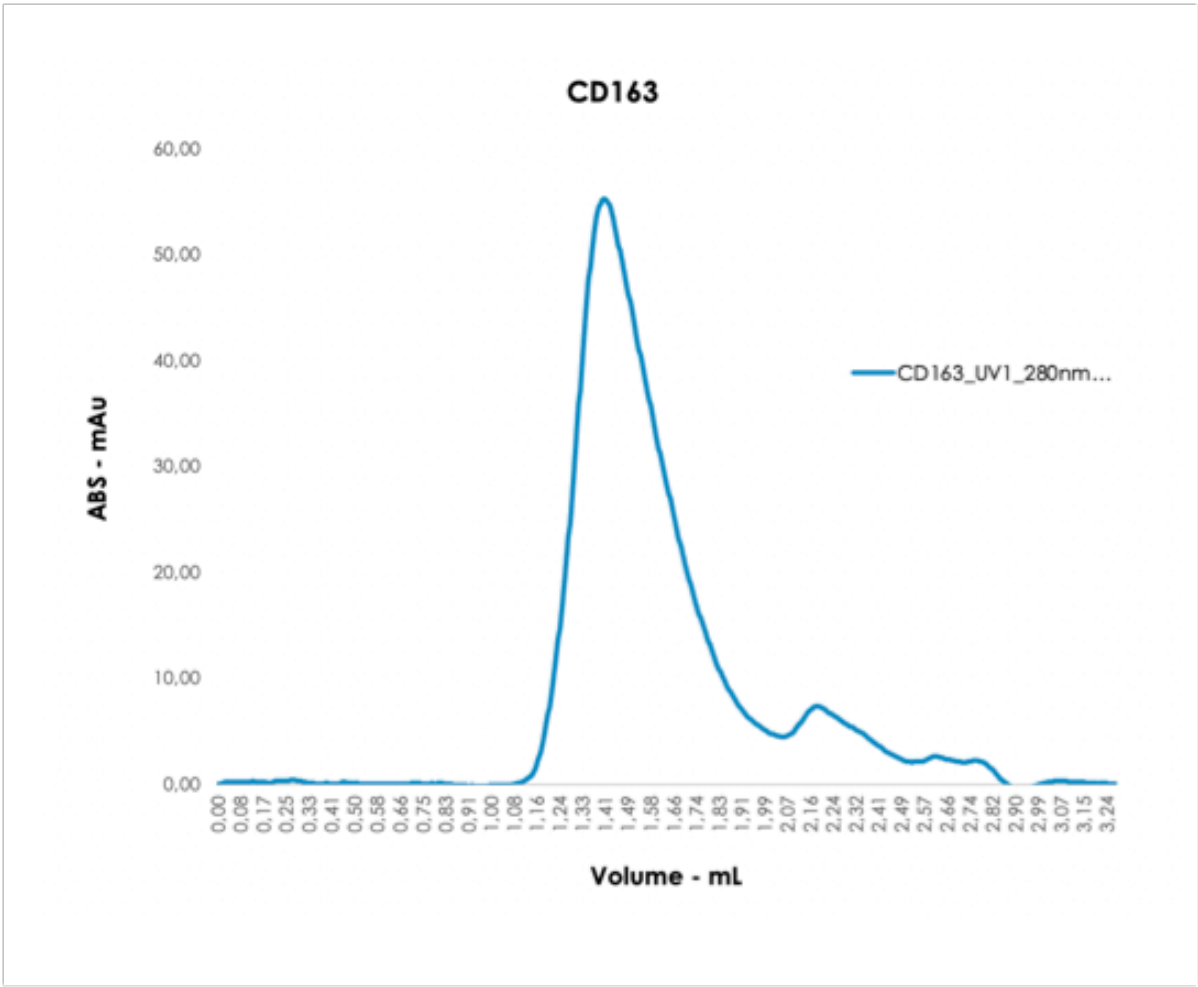


Figure 8 :Size-exclusion chromatography profile of Cd163 expressed in HEK293 cells.

Given that physiologically, Cd163 plays a fundamental role within human macrophages, that is, it binds the haptoglomin-hemoglobin complex, in order to eliminate the excess of heme groups that are potentially toxic to the body. The ability of recombinant Cd163 to perform this activity was evaluated.

At this point the Hp-Hb complex was incubated for about 4 hours, at 4 ° C, with the recombinant Cd163, in a 3: 1 ratio.

Since the recombinant Cd163 has a 6Histag tail, it was possible to eliminate the excess of Hp-Hb through affinity chromatography.The formation of the ternary complex Cd163 / Hp-Hb was verified by means of a new Size exclusion chromatography and an SDS page of the loaded sample (*Fig. 9 and Fig. 10)*

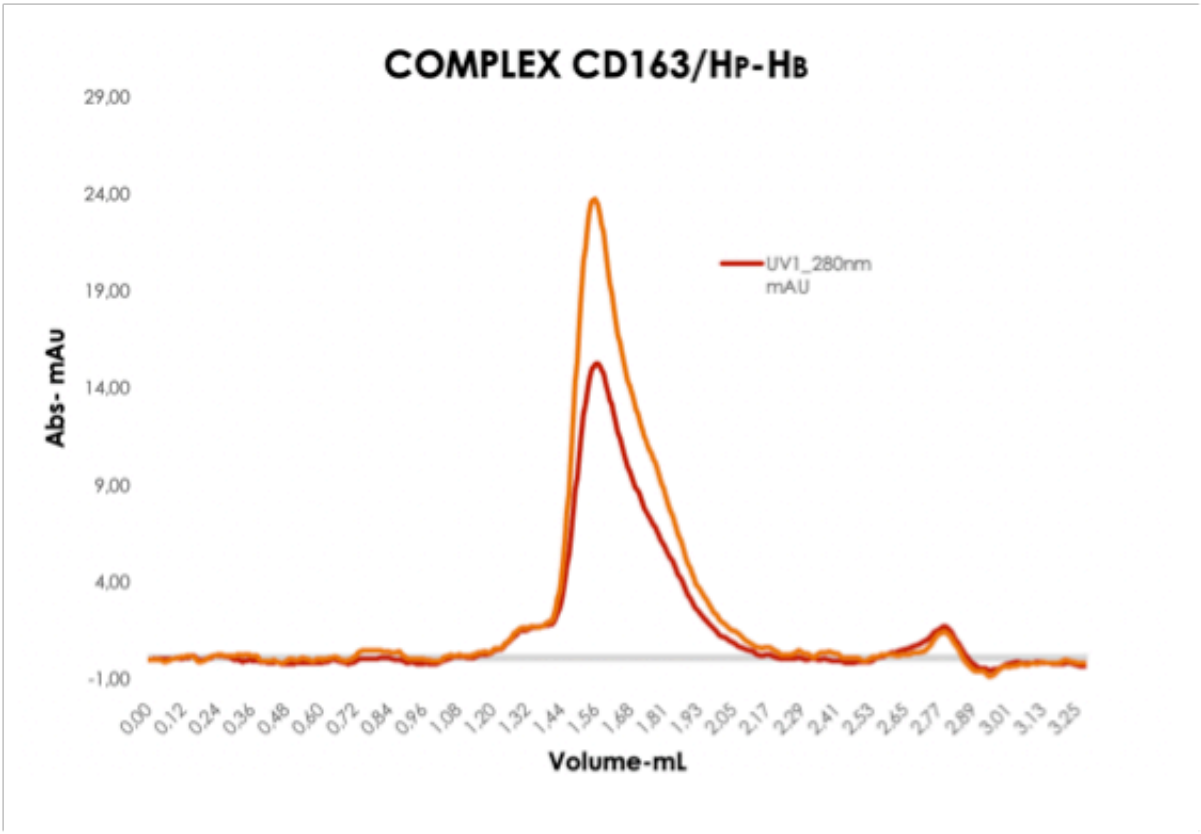


Figure 9 : Size-exclusion chromatography profile of ternary complex CD163/Hp-Hb

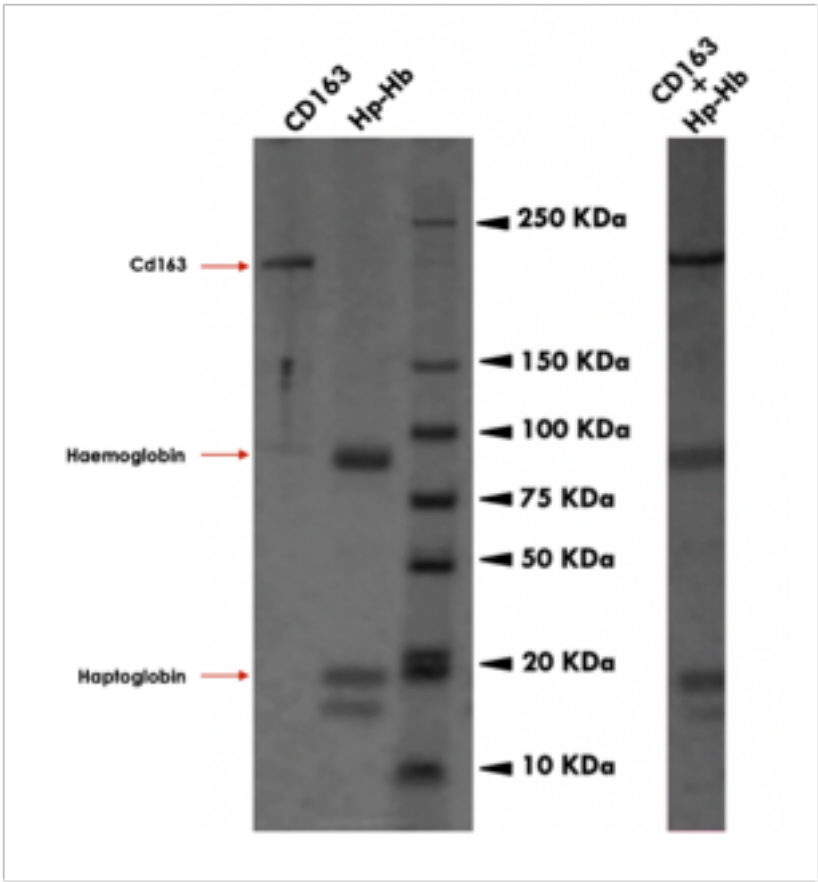


Figure 10 :SDS-PAGE to confirm the presence in the ternary complex of single part: Cd163 and Hp-Hb.

Although the results obtained seemed encouraging; from the superimposition of the single curves obtained by Size Exclusion chromatography it is clear that the protein populations that should form the ternary complex are not actually able to bind.

The peak corresponding to the Cd163 / Hp-Hb complex, therefore given by the sum of the molecular weights of the molecules that compose it (130KDa + 109KDa), should be decidedly off to the left. The peak that should describe the ternary complex is much smaller than one would expect *(Fig. 11)*.

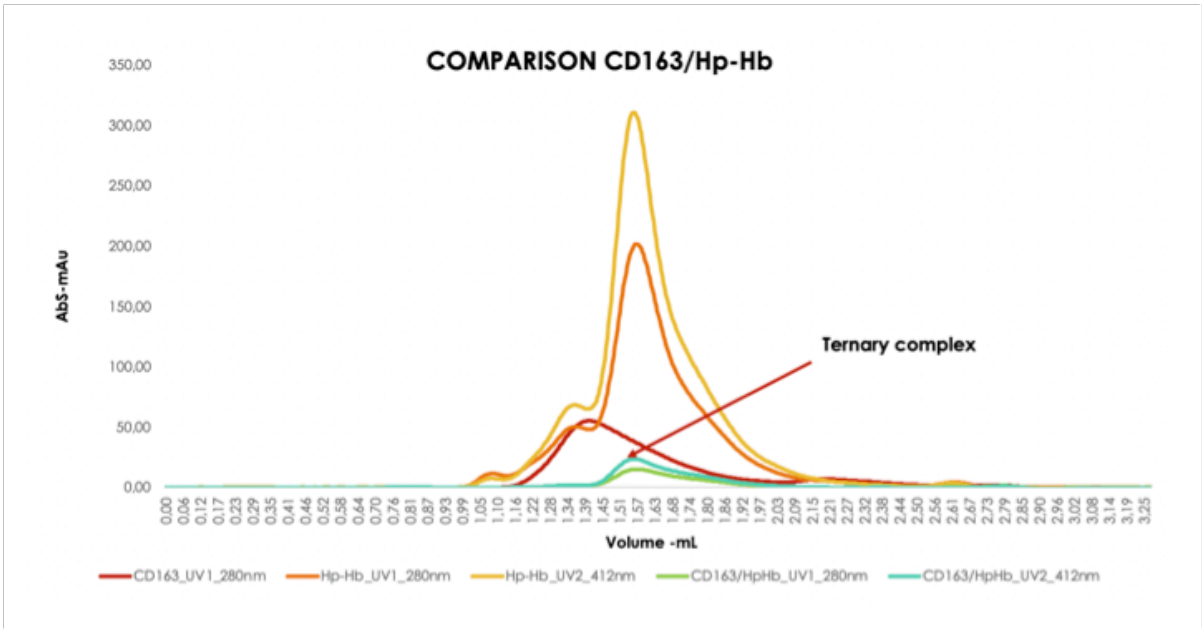


Figure 11 : Size-exclusion chromatography profile comparison. NB: absorbance measurements were also conducted at the wavelength of 412 nm, as hemoglobin has heme groups. RED= CD163 (280nm); ORANGE= Hp-Hb complex(280nm); YELLOW= Hp-Hb (412nm); GREEN= CD163/Hp-Hb complex (280nm); BLUE= CD163/Hp-Hb complex(412nm).

This discordant data has led to several questions on the stability of the complex, which if it did not form could not be characterized, bringing the study back to the starting point.

By using a Dynamic Light Scattering, two samples of Cd163 / Hp-Hb at different concentrations (0.4 mg / ml; 0.05 mg / ml) and kept in different conditions (4 ° C) were then compared. ; Frozen), with a sample of the Hp-Hb complex with a concentration of 1 mg / ml, kept at 4 ° C.

The result of the analysis showed how the cd163 / Hp-Hb ternary complex is not stable and tends to disintegrate, in both samples, therefore at any concentration and condition of maintenance, producing two different populations of molecules *(Fig. 12)*.

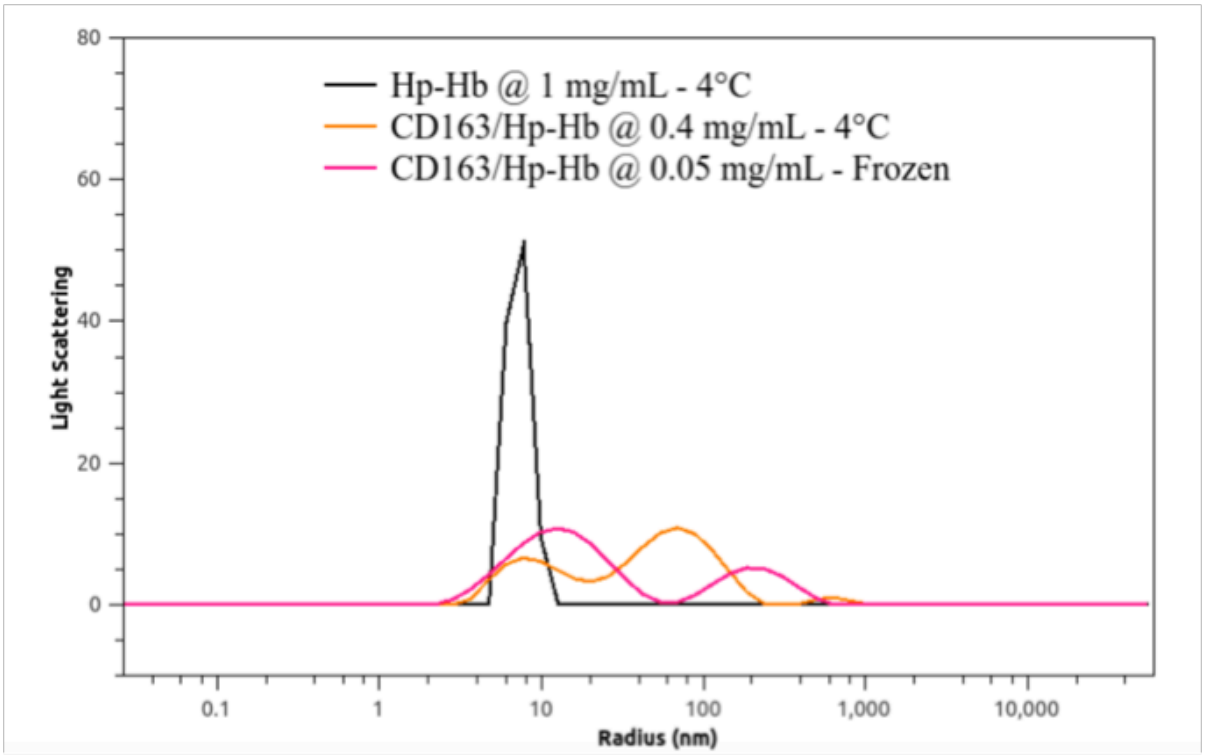


Figure 12 : Light scattering data performed by Sapienza Group. Hp-Hb complex VS two different sample of ternary complex Cd163/Hp-Hb.

On the contrary, the Hp-Hb complex, stored at 4 ° C, shows perfect stability, presenting itself as a single population.

Superimposed on the peak, represent the Hp-Hb complex, we find one of the two populations produced by the samples containing the tertiary complex, as confirmation of the hypothesis that the complex, although initially formed, tends to break down into its main components, namely Hp-Hb and Cd163 , resulting unstable.

The lack of formation of the ternary complex could be explained by a folding problem of the recombinant protein.

For this reason it was decided to deepen the study with the integration of calorimetric surveys on all components of the complex in order to understand if one of these was unfolded and consequently interfered with the bond of the others.

The analysis was conducted at the Ny-SBC laboratories. The tycho NT6 tool was used. This instrument allows to measure the structural integrity of the protein of interest by detecting the fluorescence of the intrinsic residues of tryptophan and tyrosine detected at both 350 nm and 330 nm depending on a temperature that reaches up to 95 ° C.

In addition, it allows you to analyze several samples simultaneously.

In fact, the analysis was conducted simultaneously, first on a sample containing the ternary complex and later, on the individual components of the same *(Fig. 13)*.

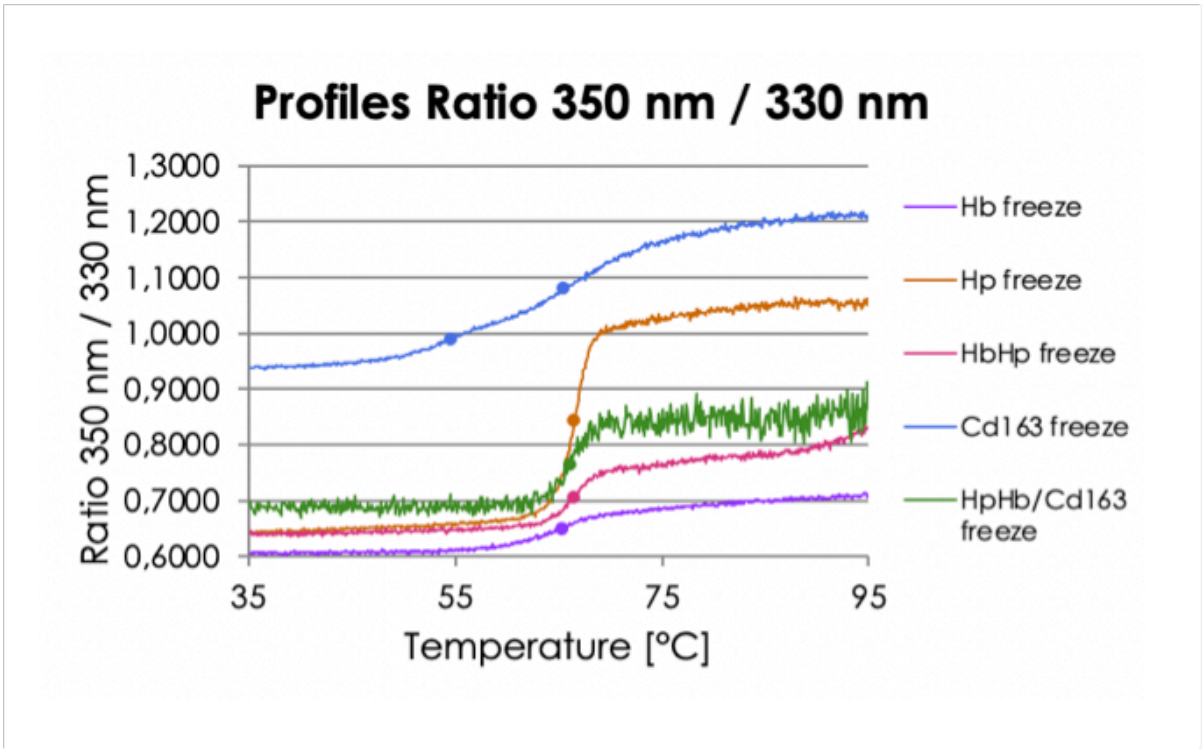


Figure 13 : Denaturation profile, detecting by Ratio between the fluorescence issued by tryptophan and tyrosine in single part in ternary complex according to temperature

The results shown in Fig.13 show that none of the tested components show signs of degradation or unfolding.

The confirmation of the fact that the production of a Cd163 with recombinant DNA technology does not affect its structure, however, does not resolve the inefficiency of the same, in the formation of the complex with Hp-Hb.

4. Conclusion

From the data shown in this thesis chapter we can assume that within eukaryotic expression systems even small protocol changes can make a difference.

In our case, a recombinant human protein shows satisfactory expression in human embryonic cells, but only with the use of excessively expensive reagents.

On the contrary, the Baculovirus system and the expression in insect cells proved particularly efficient for the production of recombinant Cd163.

Technically speaking, this can be explained by the fact that: Insect cell cultures are notoriously easier to use, as they are less delicate than human embryonic cells. This allows them to be subjected to greater stress (such as a high concentration of cells in small volumes, shorter or longer incubation times, the use of buffers with a more acidic pH, etc.) without altering the product of expression (32-33). Moreover, trivially, since insect cells allow the use of high concentration cultures compared to human embryonic cells, it is obvious that a greater quantity of expressed proteins will be obtained from this system (*Fig. 1*).

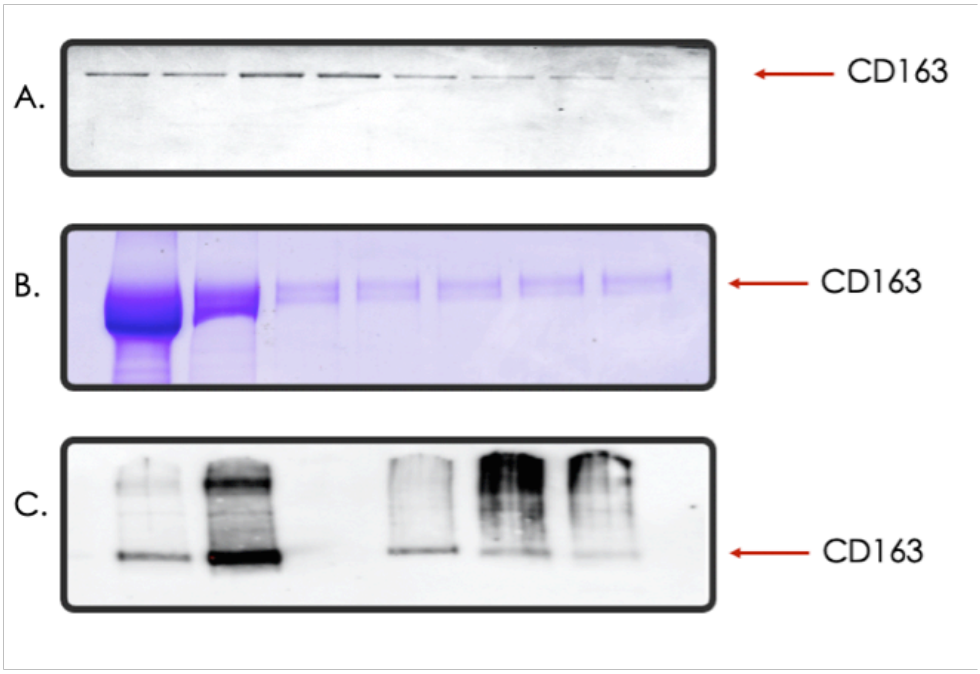


Figure 1: comparative elution profile of recombinant CD163. A. = 400 mL of HEK293s cell culturestransfected with PEI; B. = 400 ml of HEL293s cell cultures transfected with expifectamine kit; C. = 1 ml of cell cultures of insects transfected with cellfectin II

The use of the Baculovirus expression system has proved to be quite important for the production of recombinant proteins, above all because it preserves the characteristics of a eukaryotic system, in fact it does not show post-translational modifications or folding problems in the expressed proteins, but allows to increase remarkably the quantitative yield of expressed protein. Furthermore, many protein structures have been resolved thanks to expression in this system (34-43).

All this cannot be demonstrated, but only hypothesized, since the experiments have been temporarily slowed down to give more attention, by the whole scientific community, to investigations on the new Coronavirus. For this reason, more accurate and repeatable comparative data are not reported within this thesis, i.e. a yield of recombinant Cd163 obtained in parallel from the same cell culture volume in both mammalian embryonic and insect cells.

Of course, this comparison analysis is only temporarily postponed; full analyzes will be performed as soon as the situation makes it possible.

The analyzes carried out with calorimetric assays, to verify the purity and structural stability of the recombinant Cd163 have all given a positive result: the recombinant protein appears correctly folded and there are no contaminants present that could interfere with its biological activity. Despite this, to date the inefficiency of the protein produced in binding the Hp-Hb complex is not clear, as it should physiologically do.

It is known that the binding between CD163 and the Hp-Hb complex within the human organism does not occur suddenly and that a small mutation, even a specific site of individual amino acids, can prevent recognition between the molecules analyzed (Marianne JN et al., 2013).

It is evident that further analyzes are needed to complete and deepen the problems encountered in this work. A sequencing of the entire protein construct would be appropriate to confirm that the protein of interest has not undergone mutations over time that prevent binding with its substrate. Furthermore, cryoelectronic microscopy studies are among the objectives indicated for this study, but at the moment, due to the restrictions imposed by the new Coronavirus, it has not been possible to carry out them. Hopefully, they can be done soon in order to provide more complete data for this job.

Mpro

1. Introduction

In November 2019, a new Coronavirus triggered a global health emergency, known today as COVID-19. The onset of this epidemic represented a great challenge for all scientists in the world.

The first coronaviruses (CoVs) have been reported since 1947, when the first prototype mouse strain JHM was discovered (Cheever F.S. et al.; 1949; Bailey O.T. et al.; 1949).

CoVs are enveloped viruses, made up of positive single-stranded RNA and infect various vertebrates (bats, pets , livestock, poultry and humans) (Goyal B. & Goyal D.;2020).

Among humans, CoV are responsible for respiratory, gastrointestinal and neurological problems (Zumla A. et al.; 2016; de Wit E. et al.; 2016).

In recent years, Coronaviruses have also been recognized as responsible for other epidemics: the first SARS epidemic, in 2003, and MERS (Middle East Respiratory Syndrome), in 2012. However, these two epidemics have been so efficiently controlled that vaccines products to limit its spread have never been used (Lin J. et al.; 2007; Yong C. et al.; 2019).

The new Coronavirus, is currently called "SARS-CoV-2", (Gorbalenya A. E. et al.; 2020) from Severe Acute Respiratory Syndrome Coronavirus 2 since the RNA genome is 82% similar to that of the coronavirus (SARS-CoV) responsible for the SARS epidemic of 2003 (Zhou P. et al.; 2020 Wu F. et al.; 2020).

1.1 The SARS-CoV 2

Coronaviruses owe their name to the particular morphology that distinguishes them: precisely the proteins that protrude from their capsid make them appear to be surrounded by a crown. For this reason their family is called Coronaviridae.

In particular, Covid-19 belongs to the Orthocoronavirinae subfamily. Which is further divided into four genera (CoV): Alphacoronavirus, Betacoronavirus, Deltacoronavirus and Gammacoronavirus.

In detail, SarS-CoV-2 is a Betacoronavirus, more specifically, belonging to the subgenus of Sarbecoviruses.

Beta coronaviruses have a positive sense single-stranded RNA genome of approximately 29.9 kilobases encapsulated by a membrane envelope. It is the largest genome among RNA viruses(Kim D. et al.; 2020).

The BetaCoronaviruses are quite well known to date, as they have caused outbreaks of severe acute respiratory syndromes (SARS) in the past, including SARS-CoV (2002/2003 in Guangdong, China) and Middle Eastern respiratory syndrome virus MERS-CoV (2012 in Saudi Arabia) (Masters PS 2006).

To date, the source of β Coronavirus SARS-CoV-2 is believed to be bats, which carry this virus with no obvious disease manifestations (Sola I. et al.; 2015).

A single viral particle (the virion) of SARSCoV-2 is composed of a nucleocapsid composed of genomic RNA and covered with phosphorylated proteins that interact with the viral membrane during virion assembly, playing a critical role in enhancing virus replication. These proteins are found embedded in a bilayer of phospholipids and play crucial roles for both infection and replication: protein S, membrane protein (M), hemagglutinin esterase (HE) and lining protein (IS).

The viral particle is 60–100 nm (10 m) in diameter and appears round or oval (McBride R. et al.; 2014). (*Fig. 1*)

In particular, the transmembrane spike glycoproteins (S proteins) form homotrimers that protrude from the viral surface recreating the renowned corona around the viral particle.

SARS-CoV and SARS-CoV-2 proteins S, exhibit structural homology and conserved ectodomains, so previous strategies employed to study, prevent or treat SARS-CoV infection may be relevant (Berger A. & Schechter I. 1970; Anand K. et al.; 2002).

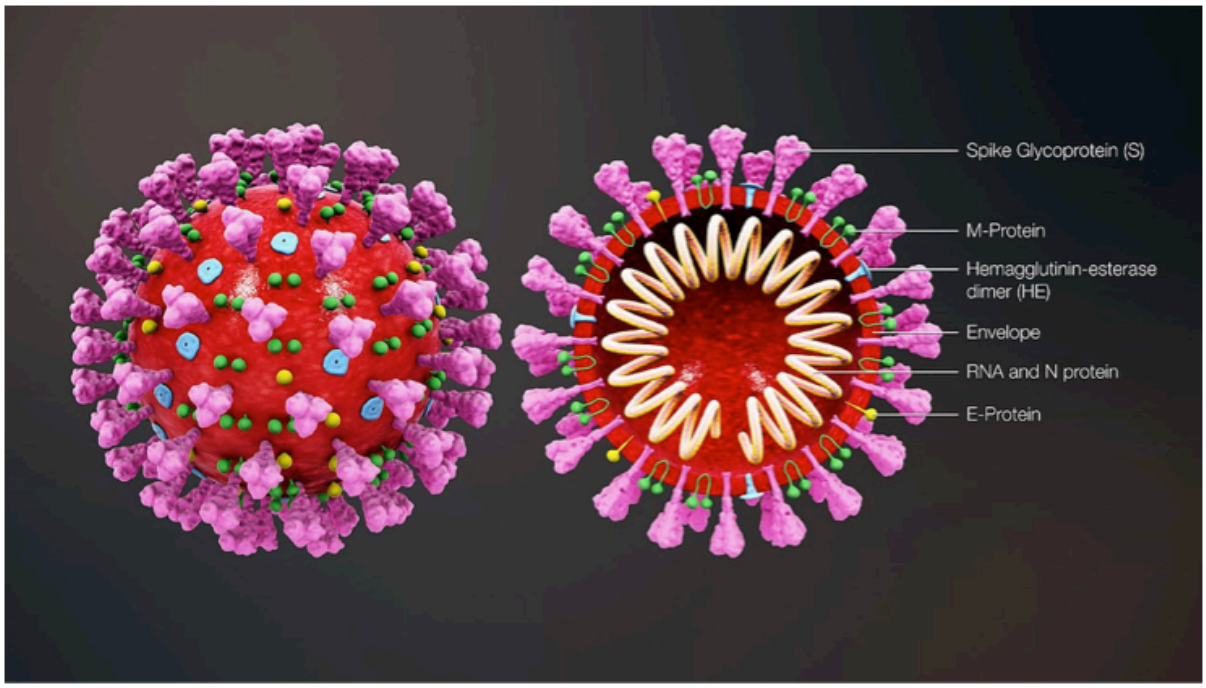


Figure 1 : SARS-CoV-2 viral particle composition (https://www.scientificanimations.com/wiki-images/)

Although SARS-CoV and SARS-CoV-2 have only 79% genomic sequence similarity, they share a highly conserved receptor binding domain for their S proteins (Colavita F. et al.; 2020). There are also other highly conserved proteins associated with SARS-CoV and SARS-CoV-2, including RdRp and 3Clpro (also called Mpro), which share more than 95% similarity between the two viruses, despite only 79% sharing genomic sequences (Hanife P. et al.; 2021).

1. 2 Viral replication of SARS-CoV-2

The replication and transcription mechanism of the SARS-CoV-2 genome appears to be common to that of other coronaviruses. In particular, 80% is identical to that of his brother SARS-CoV, head of SARS. These viruses share many biological characteristics, including the mechanism and proteins they use to enter the host cell. SARS-CoV-2 begins its replication

cycle, with the virion binding to the target cell membrane. The Spike protein, present on the surface of the virion, specifically recognizes and binds the ACE-2 receptor located on the surface of the target cell. Once the link between the virion and the receptor has been established, the former is absorbed into the host cell by endocytosis. At this point, the Spike protein makes changes to its conformation to facilitate the fusion of the endosomal membrane with the membrane of the target cell. Once inside the cytoplasm of the host cell, thanks also to the presence of digestive enzymes of the latter, the structure of the nucleocapsid is degraded, resulting in the release of the viral genome inside the cell.

Once inside the cell, the Coronavirus RNA genome is immediately translated by the ribosomes and host cell specific proteins.

It is possible to divide the Coronavirus genome into two parts: in the first we find the replicase gene, in turn divided into two ORFs (ORF1a and ORF1b), which encodes a giant polyprotein, renamed pp1ab. The pp1ab will then be “cut” into 16 parts by the main protease of the virus, resulting in smaller non-structural proteins (nsp, non structural proteins) (Snijder E.J. et al .; 2016; Sola I. et al .; 2015). All 16 nsp are necessary for the replication of genomic RNA (gRNA) and the production of various messenger RNAs (mRNAs) and viral sub-genomic RNAs (sgRNAs, the fragments of the genome from which viral proteins are translated).

Downstream of the replicase gene, however, we find the second part of the genome, where the information for the production of structural proteins is contained: spike protein; envelope (virion coat protein); M (membrane protein); N (nucleocapsid protein, which complexes viral RNA).

These proteins "pack" the viral genome and are necessary for the production of new viral particles. In addition, six other accessory proteins (3a, 6, 7a, 7b, 8) are produced from the segment downstream of the replicase gene, the role of which is not yet fully understood (Masters P. et al.; 2006).

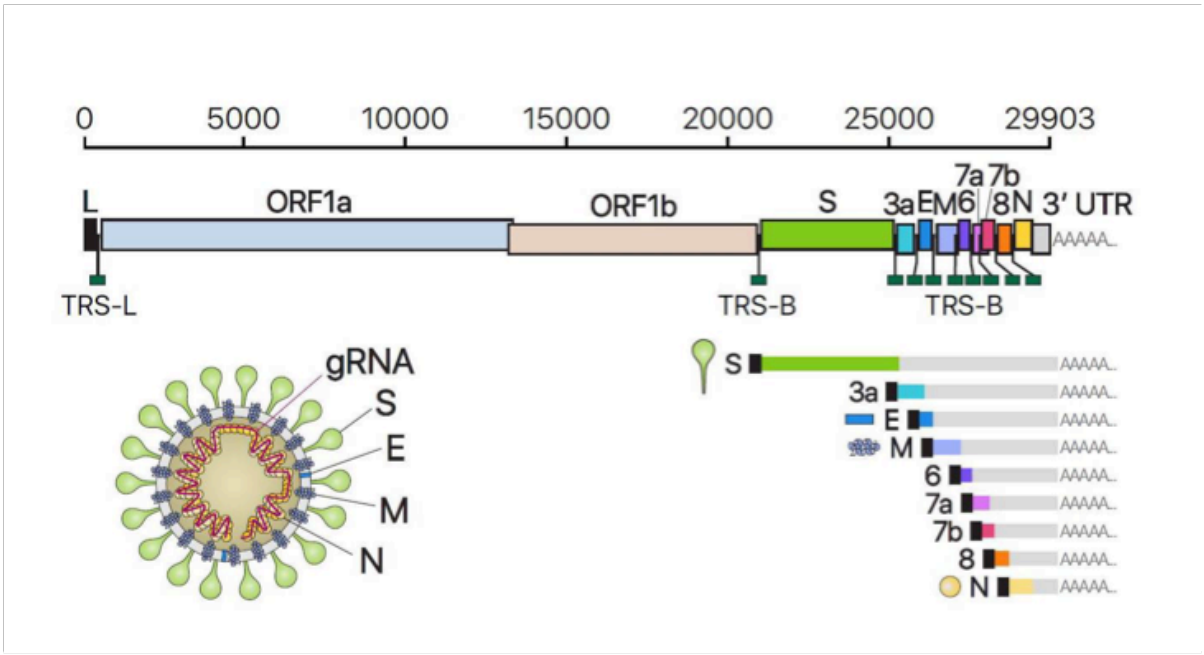


Figure 2 : gRNA structure and viral proteins of SARSCoV-2. Abbreviations: L, Leader; TRS-L, Transcription Regulatory Sequence At The Leader; TRS-B, Transcription Regulatory Sequence At The Body (regulatory sequence of transcription within the body of the genome); ORF, Open Reading Frame; S, Spike; E, Envelope (viral coat protein); M, Membrane (membrane protein); N, Nucleocapsid (nucleocapsid protein that complexes the gRNA); the abbreviations 3a, 6, 7a, 7b, 8, 10 indicate the accessory genes; nsp, non structural protein; PLP, Papain-Like Protease (papain-like protease); 3CL, Chymotrypsin-Like Protease (chymotrypsin-like protease); RdRp, RNA-dependent RNA polymerase (RNA-dependent RNA polymerase); HEL, HELicase (helicase); ExoN, ExoNuclease (exonuclease); EndoU, EndonUclease (endonuclease); Mtase, Methyltransferase (methyltransferase); UTR, UnTranslated Region (untranslated region); ST, structural protein; A, accessory protein. (Kim et al., Cell, 2020 and Sola et al., Ann Rev Virol, 2015).

The coronavirus genome is, in fact, a very long mRNA. It is a positive single strand, that is, it is read by the structures responsible for translation in the so-called 5 ′ → 3 ′ direction.

RNA replication, i.e. the production of multiple copies of the viral genome, is a continuous process. The new gRNA is produced in its entirety thanks to the synthesis of a negative intermediate strand (which runs in the 3 ′ → 5 ′ direction), which serves as a template to produce a new positive gRNA. This process mainly involves the nsp12 protein, which is an RNA-dependent RNA polymerase (RdRp), and which synthesizes the first negative strand by binding to the 3 'end of the gRNA. The bond is also possible thanks to the formation of structures that flex the RNA in order to facilitate replication thanks to the approximation of some portions.

After transcription and translation, the structural and non-structural viral proteins and genomic RNA are then assembled into virions, which are subsequently transported through the exocytotic vesicles and released from the target cell spreading throughout the human organism (Kim D. et al.; 2020).

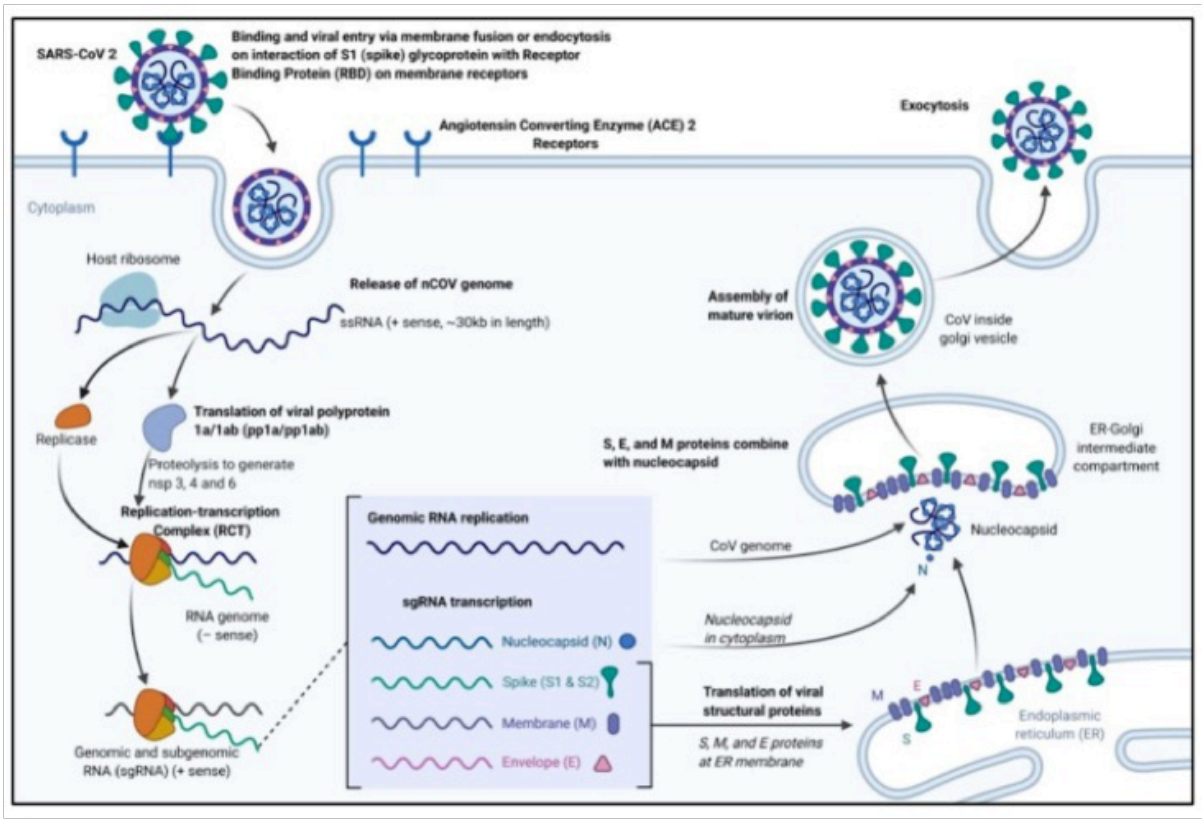


Figure 3 : Binding, viral entry and replication circle of sars cov 2 (Cascella M et al.; 2021)

1. 3 The Main Protease of SARS-CoV-2

Main protease (Mpro) is a protein present in known β coronaviruses, which are responsible for SARS (EC 3.4.22.69).

As for the SARSCoV2 Main protease, it weighs 33.8 kDa, constitutes the main protease of this Coronavirus and is essential for the processing of polyproteins translated by viral RNA (Anand K. et al .; 2002) .

This protease belongs to the C30 family of endopeptidases; it is also often called 3C like protease (3CLpro) as it is homologous to a protease found in picornaviruses.

It is a cysteine protease, as well as a member of the PA protease clan (proteases of mixed nucleophile, superfamily A) in fact it has a folded structure and proteolysis mechanisms similar to chymotrypsin (Anand K. et al .; 2002).

Its active site is characterized by a cysteine-histidine catalytic dyad and is capable of cleaving a peptide bond between a glutamine and a small amino acid (serine, alanine or glycine).

Mpro corresponds to the non-structural protein 5 of SARSCoV-2 (nsp5). Its mature form, of about 300 residues, is released from pp1ab by proteolytic self-cleavage (Anand K. et al .; 2002; Yang H. et al.; 2003; St John SE et al .; 2015; Wango F. et al. .; 2015). Once released, it dimerizes and cuts the coronavirus polyprotein at eleven conserved sites.

The protein structure of M-pro comprises three domains. Domains I and II have a typical six-stranded antiparallel β barrel structure and together resemble the architecture of chymotrypsin and picornavirus 3C proteinases. Domain III, on the other hand, is formed by five α-helices (Jin Z. et al .; 2020).

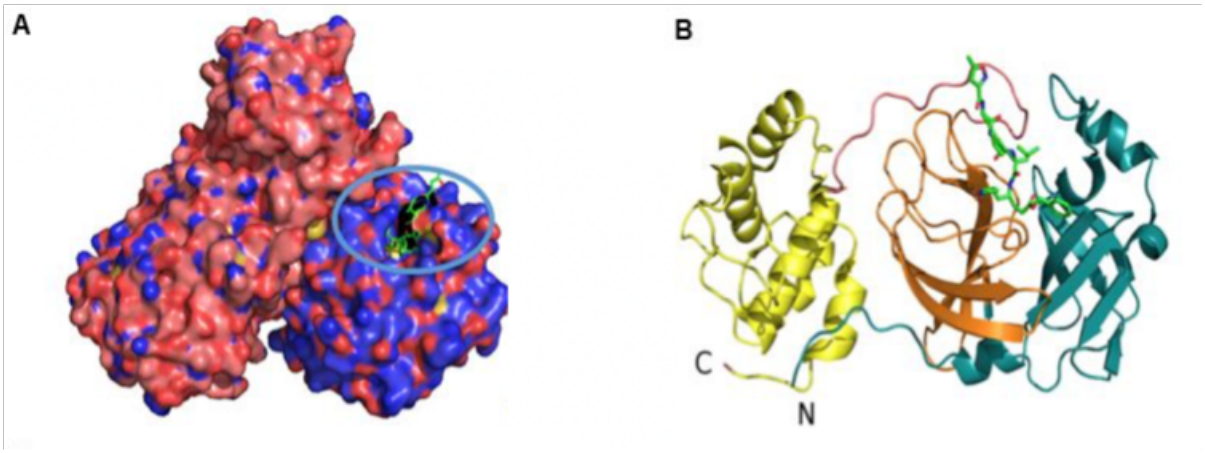


Figure 4: Crystallographic structure of the CoV-2 Mpro protein in complex with the N3 inhibitor. A: Surface representation of the Mpro homodimer. The carbon atoms of the two monomers A and B are colored pink and blue, respectively. The nitrogen atoms are colored blue, the oxygen atoms red and the sulfur atoms yellow. The N3 inhibitor is represented with the rod structure and is colored green. B: "Tape" representation of the Mpro monomer. Doimini I, II and III are colored in light green, orange and yellow, respectively. The N3 molecule is colored green. (PDB Source).

The substrate binding site is located in a deep gap between domain I and domain II. The catalytic site, on the other hand, is located in the center of the slit and contains the catalytic dyad formed by Cys145-His41. In contrast to the other cysteine proteinases, which have a catalytic triad, a third catalytic residue is not present. A long folded ring, of about 15 amino acid residues, connects domain II to the C-terminal domain (domain III). This latter domain, a globular cluster formed by five helices, seems to be involved in the self-cleveage activity of Mpro.

The mechanism of action of Mpro is similar to that of other proteases. All proteases possess two key amino acid residues: an activating residue (usually a histidine, His) that removes protons from a hydroxyl or thiol group of the side chain of a second residue (usually a serine, Ser, or a cysteine, Cys) which acts as a powerful nucleophile, i.e. a powerful electron donor. The reaction mechanism with which Mpro catalyzes the hydrolysis of the viral polyprotein is described in *Fig. 5*.

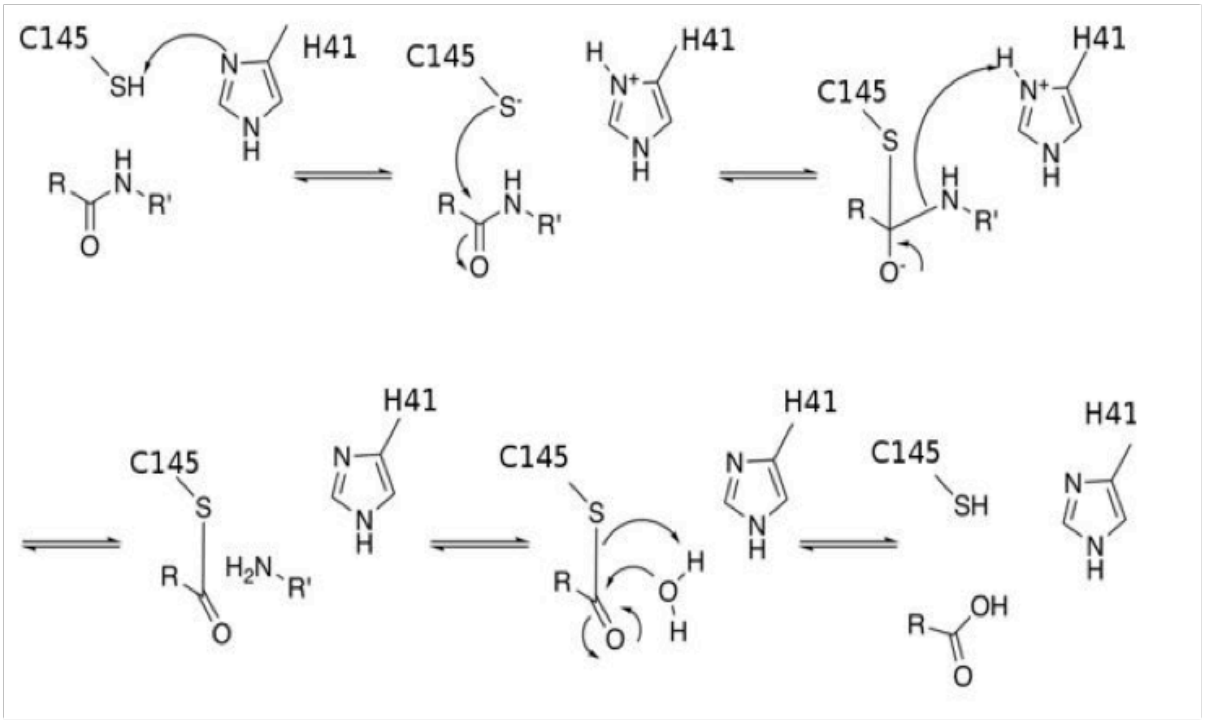


Figure 5 : Catalitic Mechanism of cysteine protease (Mpro).

The functional importance of Mpro in the viral life cycle, together with the absence of closely related homologs in humans, make this protease an interesting target for the design of new antiviral drugs: blocking its functions would in fact be lethal for the virus. but safe for humans. Furthermore, the overlap of Mpro sequences of 12 coronaviruses, including SARS-CoV-2, SARS-CoV and MERS-CoV, shows that the 9-12 residues lining the pocket that binds the substrate are strongly conserved. This suggests that a potent CoV-2 inhibitor could be a leading compound to drive broad-spectrum drug synthesis against all coronavirus infections.

2. Materials and Methods

The main protease of SarS-Cov2 was expressed in competent BL21 DE3 Plyss cells of Escherichia coli.

To be precise, the gene encoding Mpro was inserted into an expression vector, such as pGEX-6P-1 (purchased from GenScript) in order to obtain a construct encoding a protein with six histidine residues in the C-terminal position.

The recombinant plasmid containing the gene of interest was propagated to the DH5 alpha cells of E. coli and subsequently transformed into E. coli BL21 (DE3) PLyss cells, in order to allow the expression of our protein.

2.1 pGEX-6P-1 vector

PGEX vectors are designed for high-level, inducible intracellular expression of glutathione S-transferase (GST) fused genes or gene fragments.

Expression of this type of plasmid in E. coli produces proteins labeled with the GST fraction at the amino end, while the protein of interest at the carboxyl end.

The binding of a GST-labeled protein to the ligand is reversible and the protein can also be eluted under mild, non-denaturing conditions by adding reduced glutathione to elution buffer or by affinity chromatography using immobilized glutathione. The technique then provides a delicate purification process that does not affect the native structure and function of a protein.

In addition, the high solubility of GST also increases the solubility of the target protein, favoring its accumulation within the cytoplasm of the bacterial cell.

GST-labeled proteins are constructed by inserting a gene or gene fragment into the multiple cloning site (MCS). Expression is under the control of the tac promoter, which is induced by the lactose analogue isopropyl β-D-thiogalactoside (IPTG).

All pGEX vectors are also designed with an internal lacIq gene. The lacIq gene product is a repressive protein that binds to the operator region of the tac promoter, preventing expression until induction by IPTG, thereby maintaining tight control over insert expression.

In particular, pGEX-6P-1 encodes the recognition sequence for site-specific cleavage by the PreScission protease between the GST domain and the MCS.

Site-specific cleavage is performed with simultaneous immobilization of the protease on the column. The protease has a high low temperature activity so that all steps can be performed in the cold room to protect the integrity of the target protein. The cleavage enzyme and the GST tag are removed in one step.

2. 2 E.Coli BL21 (DE3) PLysS Cells

The bacterial strain BL21 (DE3) pLysS, used, is part of a rather large family of possible hosts, prokaryotic, for vectors based on the use of the T7 phage promoter.

These cells are particularly suitable for the expression of exogenous proteins induced by Isopropyl-β-D-1-thiogalactopyranoside (IPTG) as they have a high expression efficiency for any gene under the control of the T7 phage promoter.

These cells are able to express the RNA polymerase of another lambda phage, because a sequence (indicated with DE3) that contains the gene encoding this enzyme has been integrated into their genome. Furthermore, they contain the plasmid pLysS, which constitutively expresses low levels of T7 lysozyme which reduces the basal expression of target genes by inhibiting T7 RNA polymerase. This provides tight control of T7 RNA polymerase, which is especially needed when the recombinant protein to be expressed is toxic.

A further interesting feature of the BL21 strain and its derivatives is the absence of two proteases (lon and ompT), a situation that reduces the possibility of degradation of the synthesized proteins.

Finally, BL21 (DE3) pLysS E. coli cells offer greater mRNA stability so that abundant mRNA is available for protein expression. This increased stability is due to a mutation in the RNaseE (rne131) gene, which is involved in the degradation of mRNA.

2. 3 Expression tests in E. coli cells

The colonies resulting from the transformation of the recombinant plasmid pGEX-6P-1 were individually resuspended in 5 mL of LB containing 34 μg/mL of chloramphenicol and 100 μg/mL of ampicillin and grown over night, at 37 °C and 200 RPM. The next day the cultures were transferred to flasks containing 500 mL of LB with 34 μg / mL of chloramphenicol and 100 μg/mL of ampicillin and grown to an optical density (measured with a spectrophotometer at a wavelength of 600 nm) equal to at an OD600 = 0.4-0.8. At this point, the expression of the protein of interest was induced by adding IPTG (isopropyl-β-D-thio-galactopyranoside), a non-hydrolyzable derivative of lactose, in a concentration of 1 μM, and maintained for 3 hours at the temperature of 30 ° C and 37 ° C, while a whole night for the temperature of 16 °; and 200 RPM. In the induction phase the bacteria, which are in logarithmic growth, must stop using their resources to multiply and instead divert them into the production of the protein of interest.

2. 4 Purification of Recombinant Mpro

Affinity chromatography was used to purify the protein.

The column in question is a Ni-NTA column (NTA = nitrilotriacetic acid), where NTA is a high affinity chelator for metals, nickel in particular. This system allows the purification of proteins fused with a histidine tail because it exploits the selectivity of binding between the nickel ions exposed on the silica resin and the tail of 6 histidine residues (6x His tag) linked to the recombinant protein of interest.

The elution of the protein from the column is subsequently carried out with an imidazole gradient (a competitor for the coordination bond to the metal) or pH (lowering the pH protonates the histidines, which are no longer able to coordinate with the metal).

The purification procedure consists of three stages:

1. Cell lysis and resin binding by affinity
2. Washing cycle
3. Elution

After 3 hours the cells were collected and centrifuged at 4 ° C and 6000 RPM for 15 minutes. The pellet was then resuspended in 10 mL of lysis buffer (Tris-HCl 20 mM pH 8, NaCl 300 mM).

Cells were lysed by sonication: [PROGRAM 51] and the lysate was clarified by centrifugation at 6000g for 55 minutes at 4 ° C.

The supernatant was incubated for approximately 3 hours at 4 ° C, with 1 mL of pre-equilibrated Ni-NTA resin (Quiagen) to allow binding of His-labeled Mpro.

After three hours, the lysate and resin were transferred to Poly-Prep chromatographic columns (Bio-Rad, Hercules, CA, USA) and washed with two buffers (wash1 and wash2) at increasing concentrations of Imidazole.

Wash 1	Tris-HCl 20 mM pH 8, NaCl 300 mM
Wash 2	Tris-HCl 20 mM pH 8, NaCl 300 mM, Imidazole 10 mM

The protein was eluted in 5 mL of elution buffer (Tris-HCl 20 mM pH 8; 300 mM NaCl; 300 mM imidazole). After checking the presence of the protein in the eluate using an SDS-PAGE gel, followed by coomassie blue staining. The eluate was then dialyzed and concentrated in dialysis buffer (20 mM Tris-HCl pH 8, 150 mM NaCl) by using concentrators with a cutoff of 30 KDa (amicon ultra). The concentration of Mpro was determined with the A280 reading.

2. 5 Size exclusion chromatography

Size exclusion chromatography (SEC), also called gel filtration, is a technique that allows, thanks to the use of columns with different types of resins, to separate the proteins present in a solution based on their molecular weight. The instrument used (ÀKTA purifier GÈ

Healthcare) for this type of technique essentially consists of two pumps, a spectrophotometer (the instrument supplied has a lamp that emits at 280nm), a valve to regulate the flow path inside of the device, a capillary tube loop for the insertion of precise quantities of solution and a chromatographic column to be used.

The proteins were loaded into a 24 mL Superdex 75 10/300 GL Tricorn column (GE Life Science); this column is capable of separating proteins with dimensions between 3kDa and 70kDa. Standard molecular mass proteins (Sigma-Aldrich) BSA (Bovine serum albumin) (67KDa) and myoglobin (17KDa) were used to calibrate the column, each at a concentration of 1 mg / mL. The void volume of the column was calculated using Blue Dextran 2000.

The chromatographic buffer was subjected to filtration and contained 20mM of Tris-HCl pH 7.6, 150mM NaCl. The column was equilibrated with 3 volumes of buffer.

The solution containing the protein of interest was centrifuged at 10,000 g for 10 minutes, to exclude the presence of any precipitate that could interfere with the analysis.

The elution fractions were collected and those containing proteins, selected based on their absorbance at 280.

2.6 Prescission Cleavage

PreScission Protease is a glutathione S-transferase (GST) and human rhinovirus (HRV) protease 3C fusion protein type 14. Since the protease is fused with GST, it can be easily removed from the cleavage reactions using Glutathione Sepharose. The fusion proteins produced by pGEX-6P-1 will be cleaved by the PreScission protease between the GST fraction and the cloned fusion partner. The molecular weight of the PreScission protease is approximately 46 kDa.

Before being cut, the product given by the concentration of the protein eluate was dialyzed against the cleavage buffer (Tris-HCl 50 mM, pH 7.0 (at 25 ° C), NaCl 150 mM, EDTA 1 mM, dithiothreitol 1 mM). At the end of the dialysis, 1 µl (2 units) of PreScission Protease was added to the solution for every 100 µg of fusion protein in the eluate. Incubated at 4 ° C for 3 hours.

Once digestion is complete, apply the sample, consisting of the mixture of PreScission protease and Mpro fused to the GST, to the previously washed and equilibrated glutathione

sepharose to remove the GST portion of the fusion protein and the PreScission protease from the protein of interest.

collecting the eluate as it flows from the sepharose glutathione column. The eluate will contain the protein of interest, while the GST portion of the fusion protein and the PreScission protease will remain bound to the glutathione sepharose matrix.

At this point the Mpro was quantified again with the spectrophotometric reading A280.

2.7 Activity Assay

The kinetic activity of recombinant Sars-Cov2-mpro was evaluated using a fluorescently labeled self-cleavage sequence: Mca-AVLQ ↓ SGFRK (Dnp) K (purchased from GenScript, (excitation 320 nm, emission 405 nm).

The assay was started by mixing approximately 0.2 μM SARS-CoV-2Mpro to different amounts of substrate (10, 20, 40M) in order to set the best protein substrate concentration to detect Mpro activity.

The intensity of the fluorescence was measured by the DeNovix DS-11 FX + fluorometer.

The Mpro activity reported as reference for the tests was obtained by linear adaptation of the fluorescence curve in the presence of 40 M substrate concentration.

Subsequently, analyzes were performed at different temperatures.

Each reaction in a final volume of 200 mL was first incubated for 20 minutes at 16°C, 30°C and 37°C without substrate. Following the addition of the substrate, the fluorescence intensities were reported as relative fluorescence units (RFUs) and monitored every minute for a duration of 30 and 60 minutes at the above temperatures.

2. 8 Small-angle X-ray scattering (SAXS)

The recombinant SARS-COV-2-Mpro produced by E. coli were subsequently subjected to SAXS experiments in order to obtain structural details of equilibrium between the monomeric form and the dimeric form of the protein in solution.

The measurements were made on the B21 beamline of the Diamond Synchrotron (Didctot, UK), operating with a fixed camera length (4.014 m) at 12.4 keV (λ = 1.000 Å) and with a

flux of about 1012 photons per second. The samples were injected into the capillary (1.7 mm thick) by means of a robotic apparatus and measured 21 times with an exposure time of 1 minute. Sars-Cov-2-Mpro samples were measured at the molar concentration of 3, 10, 20 and 30 μM and at the temperature of 15 ° C, 25 ° C, 30 ° C, 37 ° C and 45 ° C.

All protein preparations were centrifuged for 10 minutes at 14000g prior to measurements, to remove aggregates or particles, and stored at 4 ° C.

During the analysis the samples were kept at room temperature.

3. Results & Discussion

3.1 Production and purification of SARSCoV2-Mpro in E. coli.

Once the recombinant plasmid PGEX-6-P1-Mpro was received and propagated, some tests were carried out to see what were the best conditions for its expression.

In particular, it was decided to study the effect of temperature on the expression yield.

The expression was carried out at different temperatures, both lower (16°C and 30°C) than the optimal one for the growth of *E.Coli* (37°C), and equal. This is because, for some proteins, low temperatures of expression are necessary for the correct folding of the amino acid chain and because they are not incorporated within inclusion bodies in order to be eliminated.

The three temperatures used in this study were specifically chosen after a careful review of the existing literature, because we wanted to investigate and compare the cell culture reaction.

The induction temperature of 16°C was chosen as it was used in the very first production protocols of SARSCoV-2 Mpro published since the beginning of 2020 (Zhenming Jin et al.; 2020).

The induction temperature of 30°C was chosen because it is the temperature at which the protein develops its active catalytic. (Yang, H. et al.; 2005).

Finally, the induction temperature of 37°C was chosen because it was recognized as the optimal temperature for the growth of *E. coli* but also because it is the representative temperature of the “human organism”, intended as a host of SARSCoV-2.

Cell cultures of BL21(DE3)*pLysS* containing PGEX-6P-Mpro were replicated at 37°C until reaching the OD^{600} of 0.6-0.8. Only at this point, when the IPTG was administered, did the cell cultures undergo a change in temperature.

A first analysis was conducted by taking samples at regular intervals to monitor the production of the protein of interest and understand if there was a relationship and what it was, between the induction time and the expression of the protein. Given that it was not possible to identify with certainty the representative bands of the protein from the individual total lysate samples loaded on SDS-PAGE (and later also on western blot membrane with anti-His antibody) o/n as regards the culture placed at 16°C (such a low temperature should slow down the protein production of *E.Coli* cells). While for the cultures placed at 30°C and 37°C, an induction for only three hours was chosen.

At the end of the set time, the purification process was studied step by step.

Already from the first steps of purification it was possible to notice a visible and substantial difference in the resuspension of the cell pellet (*Fig. 1*).

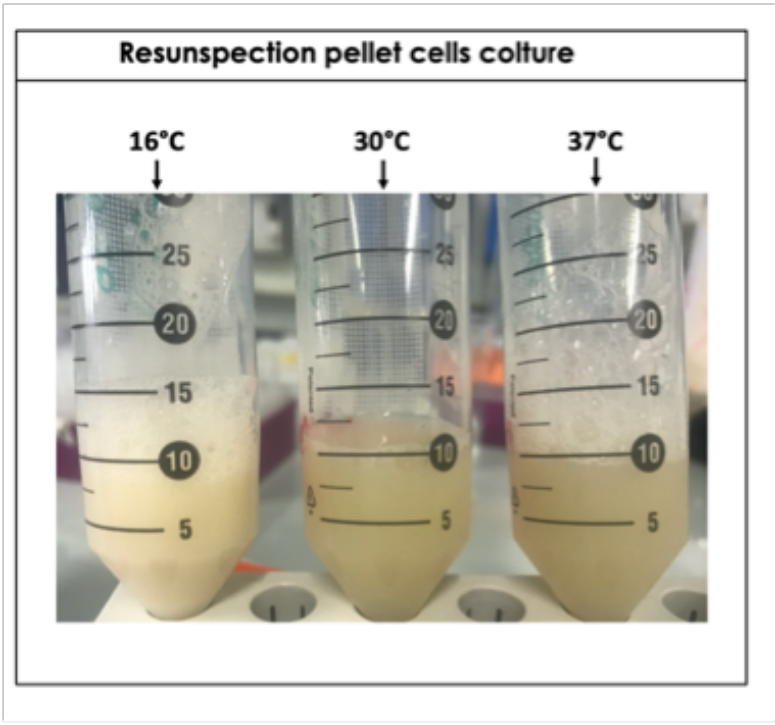


Figure 1.: Resuspension of cell culture pellets of E. coli expressing *SARSCoV-2 Mpro* at 16°C, 30°C and 37°C.

The total lysate shows different turbidity in the three different cultures. In fact, the culture that conducted an over night induction at 16°C looks like a rather milky solution. This data could be explained by a higher concentration of cells within the culture, which despite having been grown at a temperature considered "hostile" for the growth of this cell strain, having maintained it for a long time (over night) has, however, given way to them to replicate in a conspicuous way.

A difference in turbidity can also be found between the culture that conducted the induction at 37°C and the one that conducted it at 30°C. Although minimal, a slight difference can be seen: at 37°C the cells seem to have grown more, which can be explained by the fact that this is the temperature preferably used to replicate any cellular strain of *E.C*oli.

The purification steps were performed in parallel for all three cultures grown at the three different temperatures. No other detectable differences were noted in the course of purification.

At the end of the purification, the purity of the protein obtained was evaluated using an SDS-Page coloured with blue Coomassie.

From the detection of the SDS-Page electrophoresis, it was found that the expression at the three temperatures shows clearly different results (*Fig. 2*).

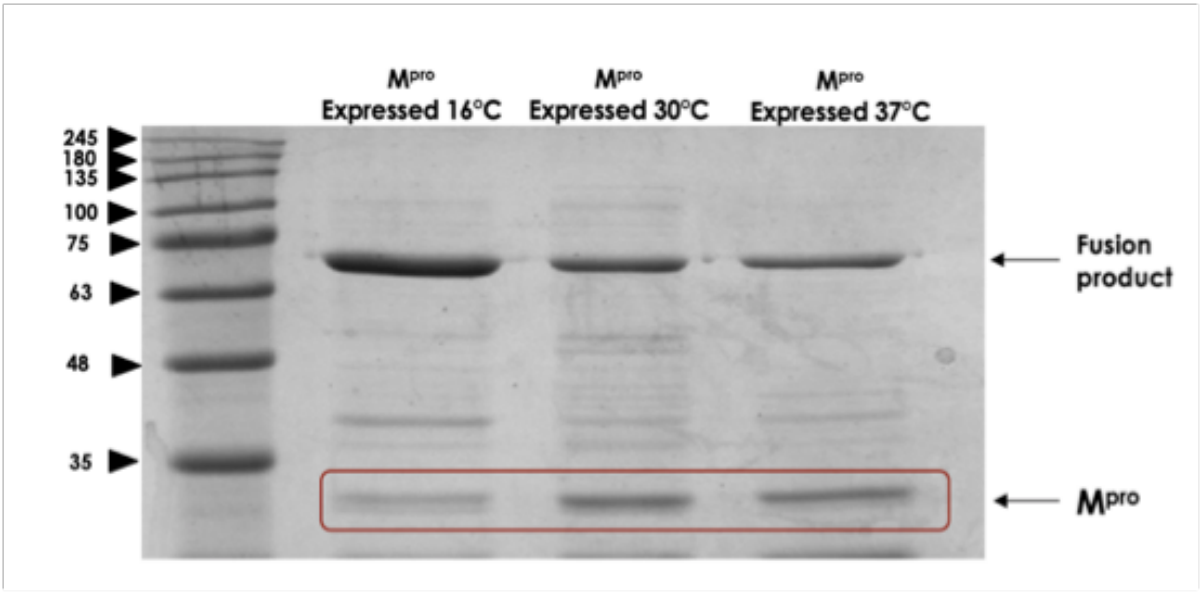


Figure 2: Recombinant *SARSCoV-2 Mpro* purified from different batch of E. coli and detected with Coomassie Blue Staining.

In all three samples, it is possible to see a basically similar profile: in fact there is the fusion product given by the expression of SARSCoV-2 Mpro fused with GST; we find the free SARSCoV-2 Mpro and finally, we find degradation products probably due to the instability of the protein itself and its tendency to change its conformation in solution passing from dimer to monomer continuously.

The fusion product between GST and SARSCoV-2 Mpro, should not be present, given the autocatalytic activity that characterizes this protein, allowed thanks to the positioning, immediately after the GST sequence and before that of the protein itself, of the signal peptide autocatalysis.

Note that, the representative band of this fusion product is particularly intense and relevant in the sample expressed at 16°C and contrasts with a rather weak band representative of the protein of interest. The explanation could lie in the fact that the choice of using such a low temperature for the expression of the protein could cause unexpected post-translational changes, responsible for unfolding events that would inhibit the autocatalytic activity of the protein itself, leading to an excessive presence of the fusion product of GST-Mpro, to the detriment of the yield of the free protein.

On the contrary, as regards the results obtained from the expressions carried out at higher temperatures (such as 30°C and 37°C), it can be seen that the presence of free SARSCoV-2 Mpro is clearly greater than that of the melting product, which in any case does not it is never absent. Furthermore, there are no substantial differences in the production yield of the protein, to the vision of this gel, between the culture induced at 30°C and that induced at 37°C. We can define them as comparable.

It has been proposed that the major proteases of the most common known Coronavirus exist as a mixture of monomer (65%) and dimer (35%) at a concentration of 1-2 mg/ml (Anand K. et al; 2003). Within this mixture, only the homodimeric form is able to operate the catalytic activities that distinguish the activity of the protein in question (Keqiang Fan et al; 2004).

For this reason, a Size Exclusion Cromathography analysis was performed using a Superdex 75 10/300 GL column in order to be able to separate, based on their molecular weight, knowing that one is twice the other, the two protein components (*Fig. 3*).

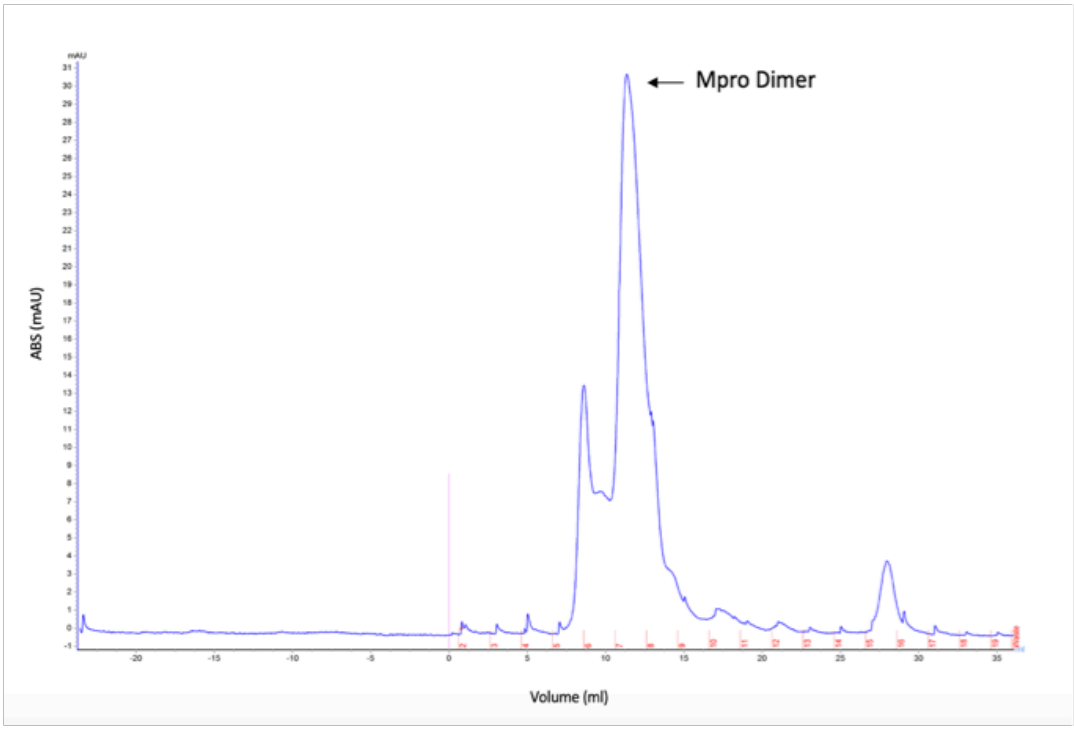


Figure 3: Size Esclusion Cromathography profile of Recombinant SARSCoV2-Mpro post purify.

The fraction corresponding to the dimeric form of the recombinant SARSCoV-2 Mpro was loaded on an SDS-PAGE gel to confirm the presence of the protein, recalling the denaturing activity of SDS (*Sodium Dodecyl Sulphate*), the reference lane for SARSCoV-2 Mpro shows a molecular weight corresponding to its monomer form (*Fig. 4*).

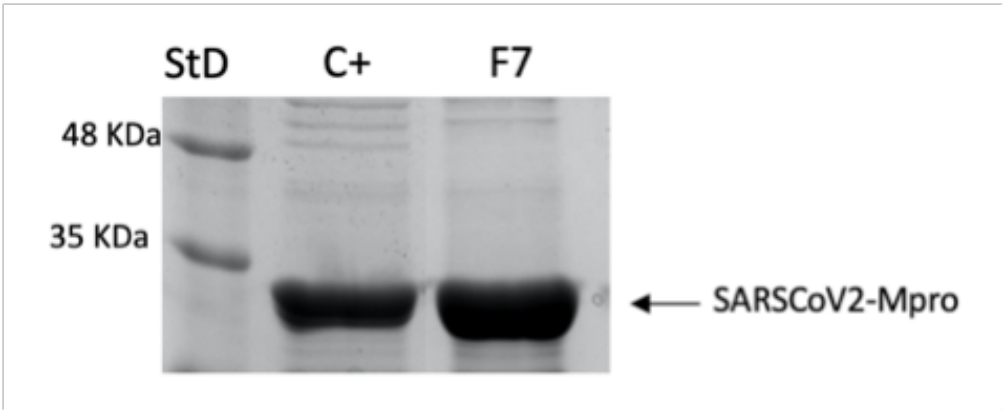


Figure 4: Detection SDS-PAGE electrophoresis of Size Esclusion Cromathogrphy fraction. StD:standard molecular weight; C+: positive control; F7: representative fraction of SARSCoV2-Mpro dimer.

3. 2 Activity Assay of Recombinant SARSCoV2-Mpro

To see if the expression of the Main protease, at different temperatures, also affected its activity, we went to investigate with fluorimetric investigations.

Knowing that the main protease of SARSCov2 recognizes a specific amino acid sequence to make cuts along the polyprotein produced by the Coronavirus, a synthetic substrate (purchased from GenScript) was selected that would emit fluorescence when the Moro recognizes the specific sequence of cut. In this way, it's possible to see if the recombinant protein produced, it is active.

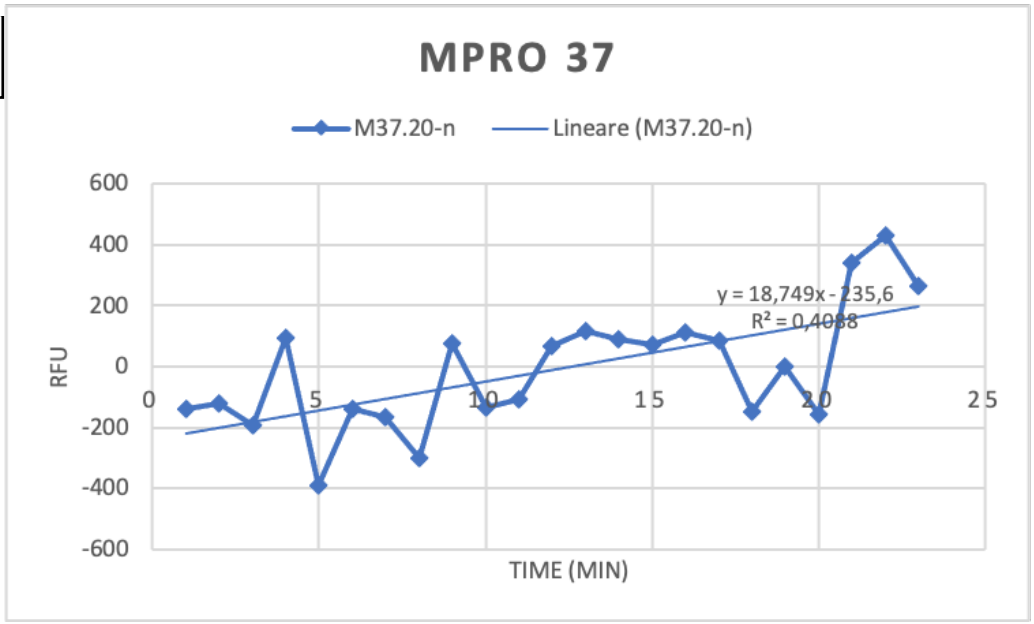
The test was carried out according to the protocol reported, adding the substrate, after about 20 minutes of incubation at 30°C, of the purified protein. At this point, the fluorescence emission of the sample was detected, maintained at 30°C, every 60 seconds, for a total 20 minutes.

The incubation temperature used was recognized as the temperature at which the Main protease is active (Yang, H. et al. 2005).

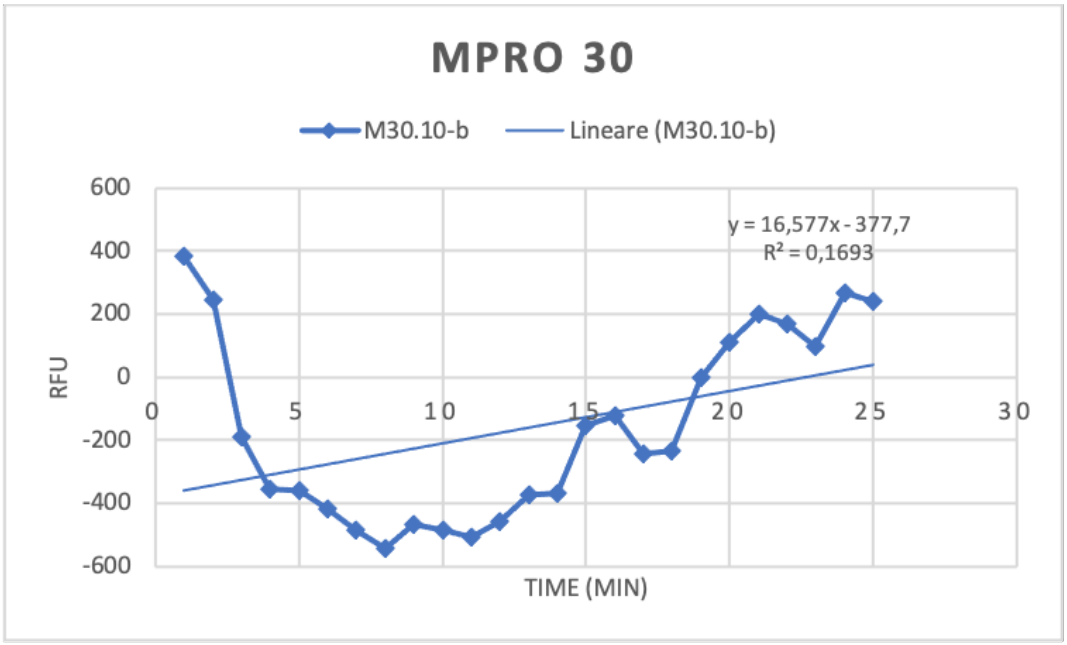
Each sample from the three cultures expressed at different temperatures was analyzed individually.

As clearly represented in the graphs below, the Main protease expressed at temperatures higher than or equal to 30°C, is active (*Fig. 5*).

A.



B.



C.

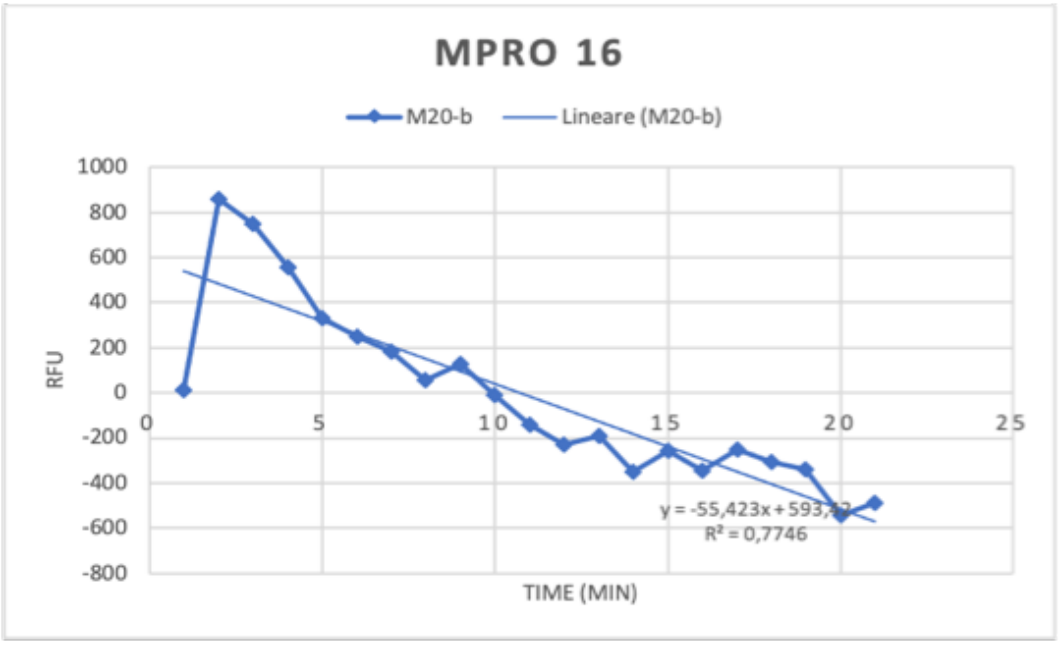


Figure 5 : Activity assay results. Representative activity curve for Recombinant SarSCoV2-Mpro expressed at 37°C; B.= SarSCoV2-Mpro expressed at 30°C;

On the other hand, it is curious to note that the same protein, expressed at 16°C, seems to show no activity, in disagreement with what is reported by Zhenming Jin et al; 2020

During the execution of the experiment, it was noted that the protein sample expressed at a temperature of 16°C seemed to begin to emit a remarkable and constant fluorescence after the canonical 20 minutes of analysis. It was therefore decided to continue analyzing the sample, minute by minute, up to a total of 60 minutes.

The SARSCoV-2 Main protease, expressed at 16°C, in addition to showing an expressionyield clearly lower than that of the other samples, seems to maintain inequalities also as regards the quantitative yield. This sample seems to have, in fact, a different kinetics than the proteins expressed at higher temperatures. In fact, its kinetics seem to be slower: after a first period of inactivity, the protein begins to cut the substrate and therefore to emit fluorescence allowing to detect a particularly positive signal for the next 40 minutes (*Fig. 6*).

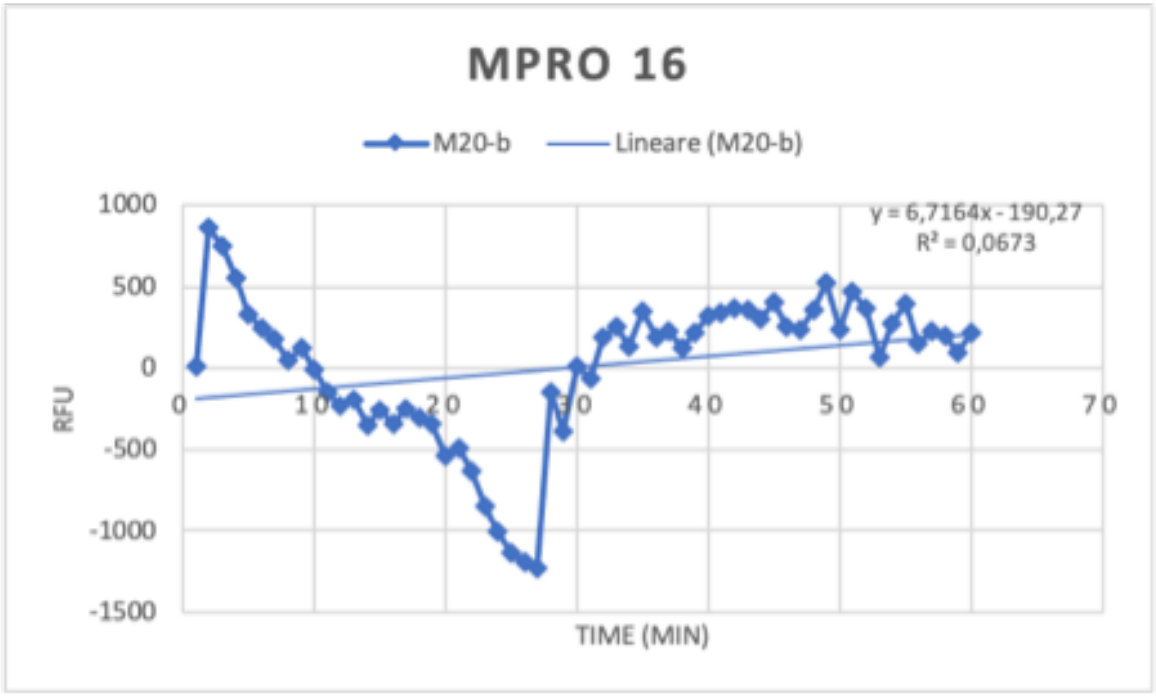


Figure 6 : Activity assay results. Representative activity curve for 60 minutes at 30°C for Recombinant SarSCoV2-Mpro expressed at 16°C.

The same analysis, with extended times up to 60 minutes, was also repeated for the main protease samples expressed at 30°C and 37°C, but the data were not considered relevant, the fluorescence emission, in fact, it tends to remain constant, or to decline, as we would expect from the data reported in the literature.

A second test on the activity of the protein was conducted by changing the sample analysis temperature. It was decided to test a higher temperature, namely 37°C. This temperature was chosen to see if at temperatures similar to those of the human body the protein was active or if it was prone to typical denaturation phenomena.

The samples were then, again, analyzed individually, changing only the temperature at which the analysis of the activity was carried out.

It was very interesting to note that at this temperature, the SARSCoV-2 Mpro expressed at 37°C is indisputably active.

Although minimal, even the SARSCoV-2 Mpro sample expressed at 16°C, shows a slight activity but it's not relevant.

On the contrary, the culture expressed at 30°C is unequivocally inactive at this temperature. (*Fig. 7*)

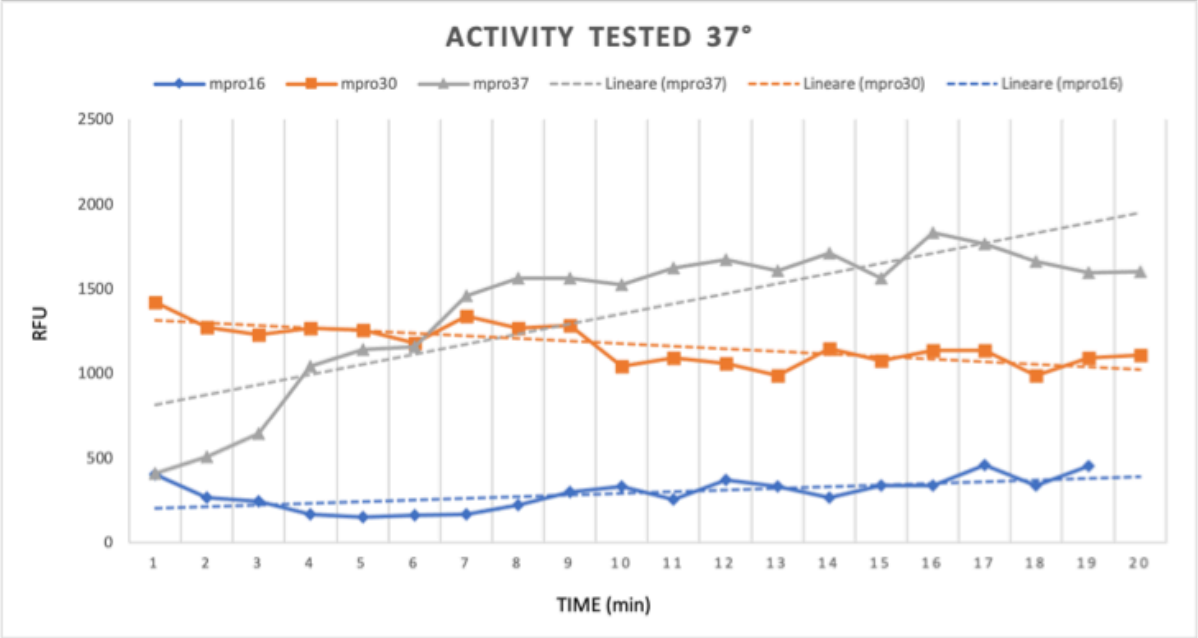


Figure 7 : Activity assay results. Comparison activity curve at 37°C for Recombinant SarSCoV2-Mpro expressed at 37°C, 30°C, and 16°C.

According to this study, we can therefore state that the best conditions for the expression of a recombinant SARSCoV-2 Mpro are represented by an expression temperature of 37°C.
In fact, this temperature allows an excellent yield of the protein both from a qualitative and quantitative point of view.
The yield of this purified recombinant protein is approximately 1mg/L of cell culture.
At the same time, the data obtained from the activity assay cannot but lead us to the reflection that if the biological activity of this protein is strictly dependent on its dimeric form (Keqiang Fan et al; 2004), it will therefore be directly connected to the balance of association/ dissociation of the monomer/dimer form present in solution.
Although the activity experiments were conducted with samples containing Maine protease fractions in dimeric form (*to see Fig. 3 for the SEC fraction*), it is not possible to be sure that the protein in solution does not tend to return to its monomeric form and consequently inactive.
Despite a large amount of experimental information already available regarding Coronavirus Main protease, there is wide disagreement on the SARSCoV-2 Mpro equilibrium monomer-dimer dissociation constant. For this reason, the small angle X-ray scattering technique (SAXS) was chosen to investigate the structural characteristics of the recombinant SARSCoV-2 Mpro monomer-dimer equilibrium, in order to reveal the corresponding equilibrium dissociation constant and associated thermodynamic parameters.

3. 3 Small angle X-Ray Scattering

The dimer-monomer balance of SARSCoV-2 Mpro was studied at different protein concentrations by performing SAXS experiments in solution in the temperature range between 15°C and 45°C.

SARSCoV-2 Mpro SAXS data were recorded on the B21 beamline of the Diamond Synchrotron (Didctot, UK) at different protein concentrations and also temperatures and are shown as log-log plots (*Fig. 8*).

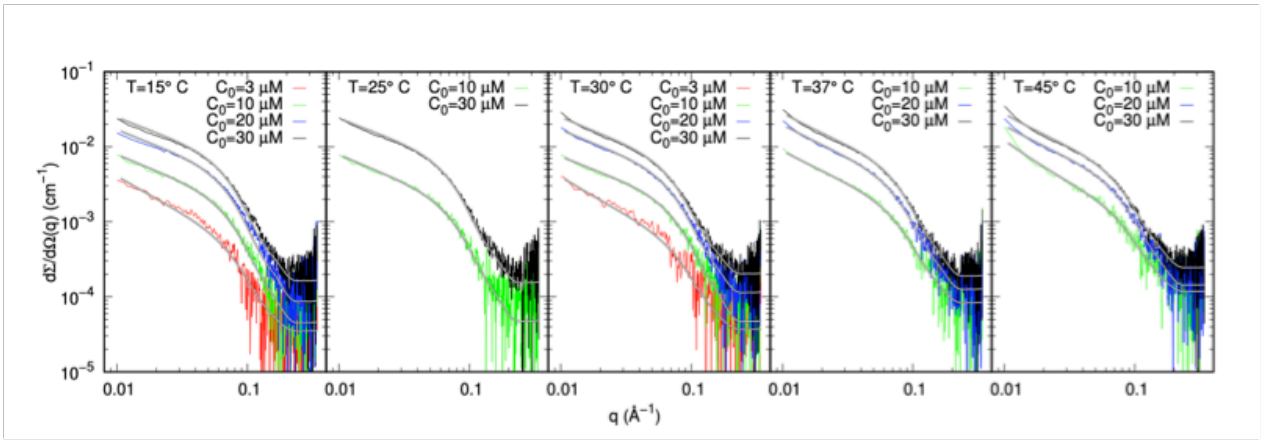


Figure 8 : SAXS experimental data of SARS-CoV-2 Mpro without inhibitors and best theoretical fits obtained by GENFIT software (solid black and white lines). Each panel reports a dataset obtained at the same temperature.

It has been hypothesized that the SAXS curves obtained from this analysis describe an interaction system between monomers and dimers of Main protease, according to the thermodynamic equilibrium dissociation process given by the relationship:

$$M_2^{pro} \rightleftharpoons 2M_1^{pro}$$

The corresponding equilibrium dissociation constant is :

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

Where: C is the total molar concentration of monomers; x_1 is the molar fraction of proteins that remain in the monomeric state; ΔG_D is the dissociation Gibbs free energy change; R is the universal gas constant and T the absolute temperature.

The most important parameter obtained from the simultaneous adaptation of the SAXS data is the dissociation constant $K°_D$, which resulted equal to 7 ± 1 μM, in good agreement with the value obtained by Graziano et al. on the very similar main protease of SARSCoV (Graziano et al.; 2006). The corresponding dissociation Gibbs free energy is $\Delta G°_D \simeq 30$ kJ mol - 1, a value quite similar to that observed for the dissociation of the β-lactoglobulin dimer at neutral pH (Apenten et al; 2002). It should be noticed that in a dissociation process, many factors other than translational and rotational motions contribute to a positive dissociation entropy and it is difficult to separate them. One such factor is, without doubts, the removal of about 200 hydration water molecules from the monomer-monomer interface when the dimer is formed (Apenten et al; 2002). The change of the heat capacity at constant pressure upon dissociation resulted positive and large. This parameter indirectly describes the monomer-monomer interface, as it can be attributed to the hydration and correlates with the interface size (Janin et al; 1995).

When dissociation heat capacities are positive and large, temperature meaningfully increases monomer-monomer affinities. This is our case: the Mpro large dissociation heat capacity might be directly correlated with the SARS-CoV-2 infective efficiency as a function of temperature.

However, a further investigation on the monomer-monomer interface area and its relationship with the dissociation heat capacity (Horton et al.; 1992) requires further calorimetric experiments in order to obtain lower estimation errors. Finally, the relative density of the hydration shell is slightly larger than one, in agreement with previous literature results on globular proteins (Svergun et al.; 1998; Sinibaldi et al.; 2007; Sinibaldi et al.; 2008).

Since the dimerization process is a fundamental key for the biological activation of SARSCoV-2 Mpro, several therapeutic strategies against covid-19 are relying on inhibitors that also (or only) act on the Mpro dimerization interface (6,12, 13).

In light of this, SAXS investigations were also used to investigate how the dissociation process of Mpro can be influenced by 7 small inhibitors selected through a combinatorial design.

The SAXS curves of SARSCoV-2 Mpro samples obtained in the presence of the seven selected potential inhibitors at two concentrations and at different temperatures are shown in Fig. 9.

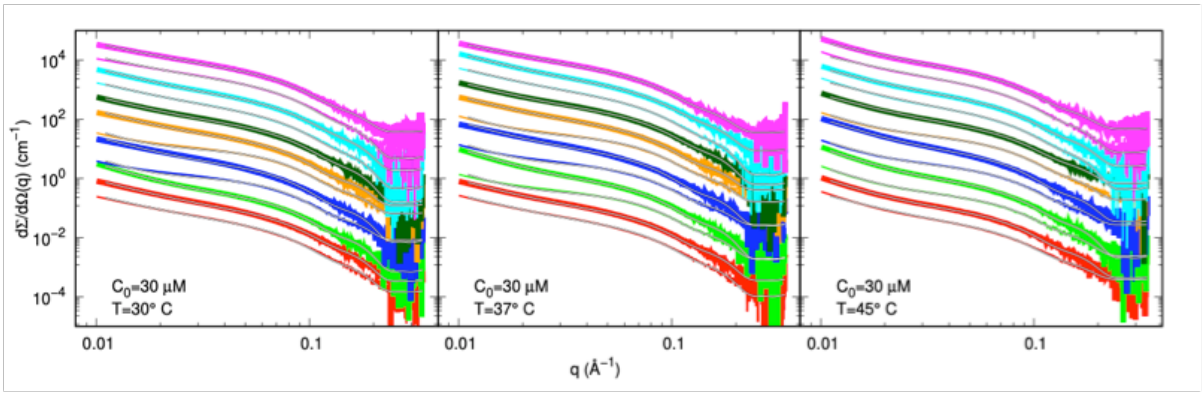


Figure 9 : SAXS data of SARSCoV2-Mpro with inhibitors at different concentrations and temperatures. Each panel reports curves at the same temperature. Red, green, blue, orange, dark-green, cyan and magenta refers to inhibitor 1, 2, 3, 4, 5, 6 and 7, respectively. Thin and thick lines refer to inhibitor concentrations of 30 and 60 μM, respectively. Subsequent curves are multiplied by a factor 3.0 for clarity. Solid black and white lines are the best fits obtained by GENFIT.

The SAXS data were analyzed with the same approach adopted for the data without inhibitors, with the further hypothesis that, for each compound, the thermodynamic parameters are linear functions of its C_{I} concentration.

the calculated SAXS curves are superimposed on the experimental curves and the resulting common thermodynamic fit parameters are shown in Table 1.

Inhibitor	α_G ($10^{-2}\ \mu M^{-1}$)	α_{C_p} ($10^{-2}\ \mu M^{-1}$)	α_S ($10^{-2}\ \mu M^{-1}$)
1	−0.18±0.02	−6±3	−4±1
2	−0.0±0.1	−1±3	−4±1
3	0.00±0.08	0±4	−8±4
4	0.0±0.1	−2±2	−1±5
5	−0.14±0.05	−1±4	−4±4
6	−0.20±0.05	−6±3	−1.7±0.8
7	−0.19±0.07	−3±3	−5±2

30 μM Inhibitor		**1**	**2**	**3**	**4**	**6**	**7**
K_D°	(μM)	14±1	8±3	7±2	6±2	15±3	14±3

60 μM Inhibitor		**1**	**2**	**3**	**4**	**5**	**6**	**7**
K_D°	(μM)	26±4	8±6	7±4	5±4	19±7	30±10	30±10

Table 1 Top panel: common thermodynamic fitting parameters of the analysis of SAXS data for SARSCoV-2 Mpro samples with inhibitors. Middle and bottom panels: dissociation constants derived by the analysis of SAXS data for SARS-CoV-2 Mpro samples with inhibitors.

The inhibitors with the lowest values of α_G are those that mostly favour dimer dissociation. Results reported in table1 suggest that compounds 1, 6, and 7 are, within the experimental error, mostly able to increase the dissociation equilibrium constant, which at C_I = 30 μM becomes as large as ≈ 15 μM and, at C_I = 60 μM almost doubles its value, reaching ≈ 30 μM. Inhibitor 5 is slightly less active: at C_I = 60 μM we found a dissociation equilibrium constant of ≈ 20 μ M. The other three compounds, 2, 3 and 4, do not show any statistically significant difference with respect to the results in the absence of inhibitors. Despite the high uncertainties on α_{Cp} and α_S, their negative values suggest that upon dissociation there are changes of heat capacity and of entropy smaller than those observed without inhibitors, indicating that inhibitors increase the monomer order.

Contrary to what has just been reported by the results obtained from the SAXS analyzes of SARSCoV2-Mpro in the presence of the 7 inhibitors, the activity assay with fluorescence inhibition of the same samples shows different results. In fact, as can be seen from figure 10 it seems that the same inhibitors, considered particularly effective in increasing the dissociation equilibrium constant, do not significantly change the biological activity of the analyzed protein.

In particular, compounds 2, 4, 5 and 7 induced irreversible inactivation of the enzyme, while compounds 1 and 6 were rather inactive (*Fig. 10*).

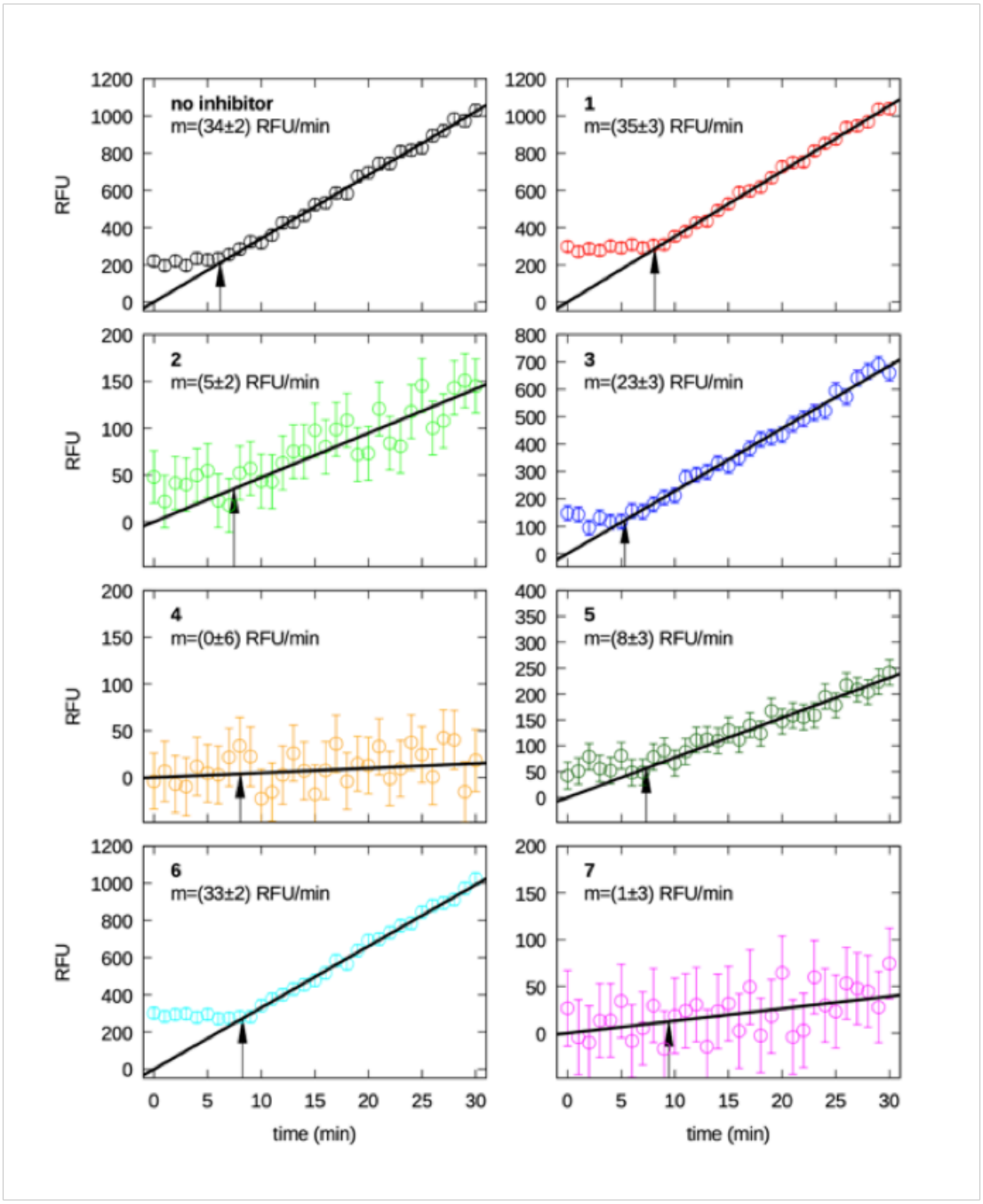


Figure10: Fluorescence inhibition curves of the selected compounds, as indicated in each frame. The straight lines are the best fitting lines obtained considering data points comprised between the time indicated by the arrow and 30 min. The slope of the straight line is reported in each frame.

4. Conclusion

In summary, the results obtained show that the optimal conditions for the production of recombinant SARSCoV2-Mpro in a bacterial expression system are a recovery of 3 hours, following the induction of protein expression with IPTG.

These conditions ensure an excellent quantitative and qualitative yield of the recombinant protein, which is perfectly active in a temperature range between 30 ° C and 37 ° C.

Furthermore, it has been established that the inhibitory effect of compounds designed to bind the catalytic site of SARSCoV2-Mpro, thus preventing its biological activity, does not necessarily change the dimer-monomer balance of the protein. On the other hand, we have also observed that compounds capable of causing dissociation of the dimeric form of SARSCoV2-mpro do not always show inhibitory effects, as evidently demonstrated in *Fig.1*.

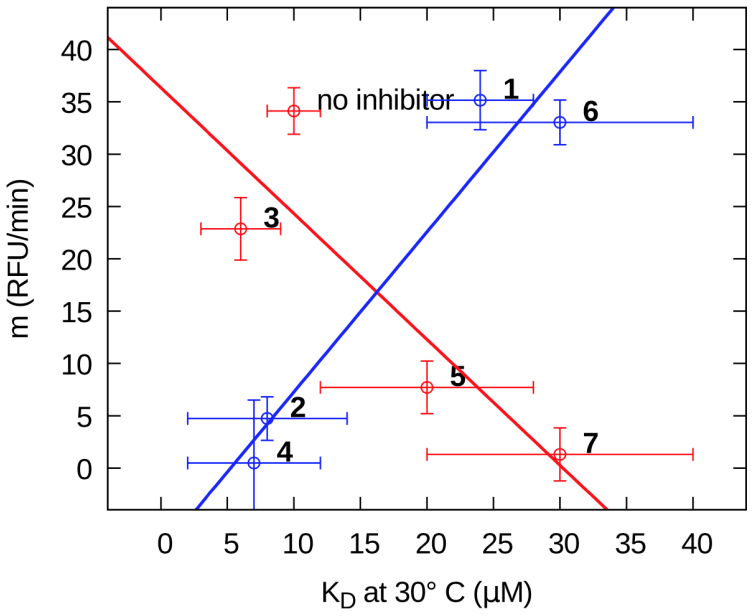


Figure 1: Correlation map between the catalytic activity, represented by the RFU slope *m*, and dimer dissociation capability, measured by the dissociation constant *KD* at 30○ C, of the seven SARS-CoV-2 Mpro inhibitors at *C*I =60µM.

The points on the map can be divided into two populations. In red the inhibitory compounds 3,5 and 7 are highlighted. These compounds show the expected behavior: the stronger their ability to induce the dissociation of Mpro, the more important is their inhibitory effect on the activity of the protein.

Therefore, the hypothesis is confirmed that the molecular mechanisms underlying the inhibition in the active site are directly linked to the ability of the aforementioned compounds, to cause the dissociation of the dimeric form.

On the contrary, in blue are highlighted inhibitory compounds 1, 2, 4 and 6 that show an opposite behavior: despite the obvious dissociation of the dimeric form, is not simultaneously detected an inhibition of the catalytic activity of Mpro.

This apparently contradictory result can be partly explained by considering that, in all reported cases, the dissociation equilibrium is weak.

Furthermore, considering a sample of SARSCoV2-Mpro in which both forms of protein (dimer and monomer) are present in solution, it can be stated that the presence of a compound that alters the dimer-monomer balance will not necessarily lead to an obstacle to the interaction of the protein with the substrate, as Mpro dimeric molecules are always present which can exert their enzymatic activity when a substrate is available.

This work provides further evidence of the complex interaction between inhibition of enzyme activity and dimeric dissociation. Furthermore, it shows for the first time the contribution of the SAXS technique, combined with advanced data analysis, to obtain structural information on SARSCoV-2 Mpro in solution and in the presence of promising inhibitors.

The reported results suggest that more experimental evidence on SARSCoV2-Mpro monomer and dimer impairment in the presence of inhibitors corroborated by computational results will be needed for a deeper understanding of the allosteric mechanism of Mpro.

Closing Remarks

The final comments from this thesis concern a few different points, that we think relevant. *First*, this study demonstrated once again that the production and the availability of recombinant proteins as pure as possible and in the required (large) quantities opens the way to structural and biophysical studies that may be necessary premise for technological applications of great interest to the entire community. *Second*, especially within eukaryotic expression systems, we observed that even small modifications of the production protocols could make large differences in protein yield, purity and stability. Indeed, the continuous demand for large quantities of different proteins for studies of all kinds has led to the development and research of increasingly efficient methods and reagents for the production of recombinant proteins. The case of the human hemoglobin receptor CD163 can be considered exemplar: we demonstrated in fact that the protein, which is human, exhibits a more efficient expression in a Baculovirus system than in human embryonic cells. The ease of culture, infection and high yields achievable by insect cells makes this system the best candidate for large-scale production of Cd163. This system, in fact, not only allows an increase in the production yield of the protein of interest with respect to mammalian cells, but also enables faster times and overall higher quality of Cd163 expression for future applications. *Third*, the subsequent biophysical and structural analyses of the produced proteins were in part not concluded and in part very successfully. At one side, the verification of the physiological activity of recombinant Cd163 did not produce the expected results: the binding of this receptor with the Haptoglobin-Hemoglobin complex resulted neither stable nor quantitatively relevant. The additional studies performed to verify the structural stability and folding of the produced protein and of its ligand confirmed the good quality of the different molecules. To date it is known that the link between CD163 and the Hp-Hb complex

does not occur suddenly and that small mutations, even a specific site of single amino acids, can prevent the recognition between the aforementioned molecules (Marianne JN et al.; 2013). Further analyzes are then necessary to better understand and solve the problems that limit the formation of the Cd163/Hp-Hb complex. It should be noticed that the structural characterization by means of cryo-electron microscopy, initially planned in collaboration with the group of Biochemistry of the Roma Sapienza University for the conclusion of this study, have been postponed due to the spread of the COVID-19 pandemic. On the other side, the spread of the COVID-19 pandemic, albeit in part limiting for the work presented in this thesis, has allowed us to contribute to the advancement of scientific discoveries regarding SARS-CoV-2. In fact, because the regulations that did not allow to move between different laboratories to complete the structural work on Cd163, we decided to quickly move to the investigation of the structural properties of the main protease (Mpro), a SARS-CoV-2 protein responsible of the cleavage the polyproteins encoded by the virus. The expression study was conducted in the E. coli host system. This system was chosen because it is inexpensive, offers fast culture times and the ability to obtain high biomass with high protein yields. The presented results show that functionally active Mpro can be produced and that an expression temperature of 37 °C allows to obtain the best qualitative / quantitative ratio. The same temperature does not seem to affect the biological autocleave activity of the protease. It is necessary to emphasize the importance of an extended expression study for the proteins of interest; in fact, biophysical techniques used for characterization often require large quantities of proteins. As Mpro is functionally active in the form of a homodimer, SAXS techniques were used to define the equilibrium dissociation constant between the monomeric and the dimeric forms and to evaluate how the dimerization process is affected by small inhibitors selected by virtual screening. We find that these inhibitors affect dimerization and enzymatic activity to a different extent and sometimes in an opposite way, likely due to the different molecular mechanisms underlying the two processes. This work, recently published in *Scientific Reports* and reported in the appendix, provides further evidence of the complex interaction between inhibition of enzyme activity and dimeric dissociation. Moreover, the analysis suggested a few Mpro amino acid residues that emerge as key to optimize both dissociation and enzymatic activity inhibition and that can be the basis for the development of future drug molecules.

Appendix

scientific reports

Check for updates

OPEN

The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors

Lucia Silvestrini[1], Norhan Belhaj[2], Lucia Comez[3], Yuri Gerelli[2], Antonino Lauria[4], Valeria Libera[5], Paolo Mariani[2], Paola Marzullo[4], Maria Grazia Ortore[2], Antonio Palumbo Piccionello[4], Caterina Petrillo[5], Lucrezia Savini[1], Alessandro Paciaroni[5] & Francesco Spinozzi[2]✉

The maturation of coronavirus SARS-CoV-2, which is the etiological agent at the origin of the COVID-19 pandemic, requires a main protease M^{pro} to cleave the virus-encoded polyproteins. Despite a wealth of experimental information already available, there is wide disagreement about the M^{pro} monomer-dimer equilibrium dissociation constant. Since the functional unit of M^{pro} is a homodimer, the detailed knowledge of the thermodynamics of this equilibrium is a key piece of information for possible therapeutic intervention, with small molecules interfering with dimerization being potential broad-spectrum antiviral drug leads. In the present study, we exploit Small Angle X-ray Scattering (SAXS) to investigate the structural features of SARS-CoV-2 M^{pro} in solution as a function of protein concentration and temperature. A detailed thermodynamic picture of the monomer-dimer equilibrium is derived, together with the temperature-dependent value of the dissociation constant. SAXS is also used to study how the M^{pro} dissociation process is affected by small inhibitors selected by virtual screening. We find that these inhibitors affect dimerization and enzymatic activity to a different extent and sometimes in an opposite way, likely due to the different molecular mechanisms underlying the two processes. The M^{pro} residues that emerge as key to optimize both dissociation and enzymatic activity inhibition are discussed.

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

Because of this mechanism of action, inhibiting M^{pro} might lead to an attenuation of the viral infection. Indeed, this enzyme is a very attractive target for anti-CoV drug design: the M^{pro} sequence is highly conserved among various CoVs[4], as mutations of M^{pro} turn out to be often fatal for the virus[5]. à us, the risk of mutation-mediated drug resistance is very low and inhibitors will display broad-spectrum antiviral activity. In addition, M^{pro} inhibitors are unlikely to be toxic because human proteases have different cleavage specificity. A second point should be however considered: the published X-ray structures of SARS-CoV-2 M^{pro}, obtained both in the presence and in the absence of inhibitors[6,7], revealed that two M^{pro} monomers form a functional active homodimer, as already detected in different coronaviruses[3], which share with SARS-CoV M^{pro} almost all the amino-acids involved in the dimerization. In such homodimer, the two monomers are arranged almost perpendicular

[1]Department of Life and Environmental Sciences, New York Marche Structural Biology Center (NY-MaSBiC), Marche Polytechnic University, 60131 Ancona, Italy. [2]Department of Life and Environmental Sciences, Marche Polytechnic University, 60131 Ancona, Italy. [3]Department of Physics and Geology, CNR-IOM c/o University of Perugia, 06123 Perugia, Italy. [4]STEBICEF Department, University of Palermo, 90128 Palermo, Italy. [5]Department of Physics and Geology, University of Perugia, 06123 Perugia, Italy. ✉email: f.spinozzi@univpm.it

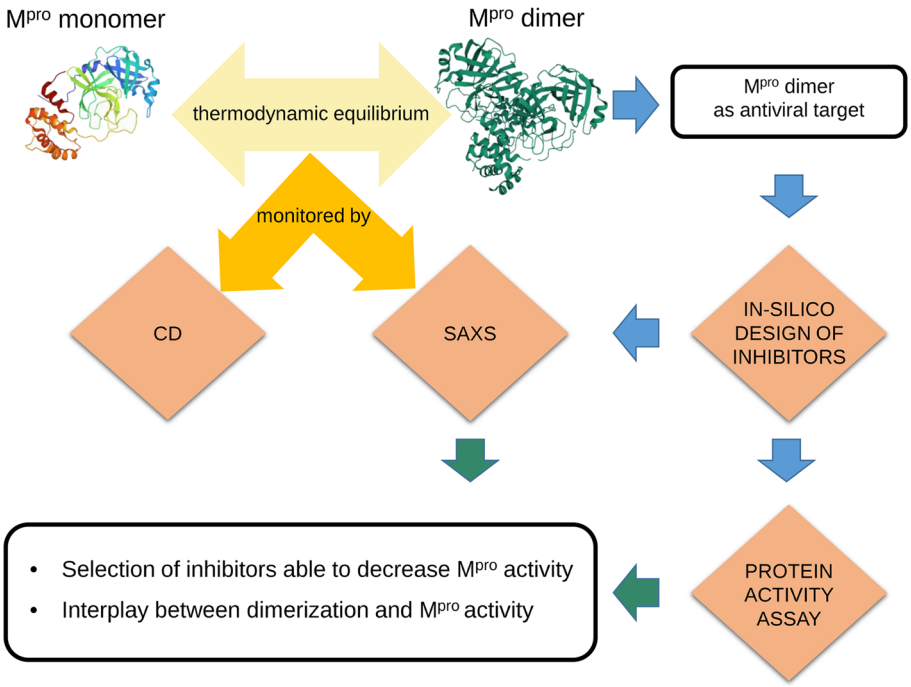


Figure 1. Graphic display of the flowchart of the biophysical method described in the text.

to each other[7] and each monomer comprises the catalytic dyad His41-Cys145 and the substrate-binding site located in a cleft between domains I and II. Domain III, which contains five α-helices arranged into a globular cluster, is directly involved in controlling the dimerization of M^{pro} mainly through a salt-bridge between Glu290 of one monomer and Arg4 of the other[8]. Quite remarkably, while individual monomers are enzymatically inactive, M^{pro} is active in the dimeric form. The structural reason behind the functionality of the dimer is probably due to the interaction of the N-finger of each of the two monomers with Glu166 of the other monomer, which establishes the shape of the so-called S1 pocket of the substrate-binding site[9]. To approach this interaction site, the N-terminal amino acid residues are squeezed in between domains II and III of the parent monomer and domain II of the other monomer[7].

According to these considerations, two different strategies have been considered for the development of therapeutic agents: first, direct inhibition of the catalytic site by using molecules targeting the substrate binding pocket; second, attenuation of the catalytic activity by using inhibitors targeting the dimerization site. The second alternative is strictly related to the M^{pro} equilibrium between dimers and monomers in solution. The thermodynamic equilibrium of M^{pro} dissociation process has been recently studied by analytical ultracentrifugation. Sedimentation velocity experiments provided a value of about 2.5 μM for the apparent dimer dissociation constant K_D[7]. However, a more recent estimate by mass-spectrometry based assays established for K_D a much lower value of 0.14 ± 0.03 μM, indicating that M^{pro} has a stronger preference to dimerize in solution than expected. In the case of SARS-CoV M^{pro}, an even wider discrepancy among the different estimates of the dimer-monomer dissociation constants has been observed, with the values of K_D provided by various experimental techniques falling in a range from 230 ± 30 μM[10] down to 0.19 ± 0.03 μM[11]. In this framework, synchrotron Small Angle X-ray Scattering (SAXS) can be a very sharp method to determine M^{pro} dimer-monomer equilibrium in solution. In fact, beyond pioneering SAXS studies performed also by some of us to investigate the thermodynamic features related to β-lactoglobulin dimerization[12–15], this approach has more recently provided noticeable information in many issues[16,17]. Among them it is worth citing the case of A3G, a key enzyme for HIV-1 infection[18], of LRRK2 protein, linked to Parkinson's disease[19]. Hence, given the above mentioned uncertainty on the K_D value that rules M^{pro} dimerization, we decided to take advantage of SAXS to provide new insights on the SARS-CoV-2 M^{pro} dimer-monomer equilibrium. The study was performed both in the absence and in the presence of a set of in-silico selected small inhibitors, whose activity was spectroscopically assayed, in order to simultaneously test their therapeutic potential with respect to dimerization inhibition. By measuring the large-scale structural features of SARS-CoV-2 M^{pro} as a function of temperature, protein concentration and in the presence of different amounts of inhibitors we provide an accurate thermodynamic picture of the SARS-CoV-2 M^{pro} inhibitor-dependent dimerization process.

Results

Our biophysical multi-technique approach, mainly based on SAXS, by which we studied the effects of potential inhibitors of the SARS-CoV-2 M^{pro} on the dimerization process and the connection with the catalytic activity, is reported in the flowchart shown in Fig. 1. We have first derived the thermodynamic parameters controlling

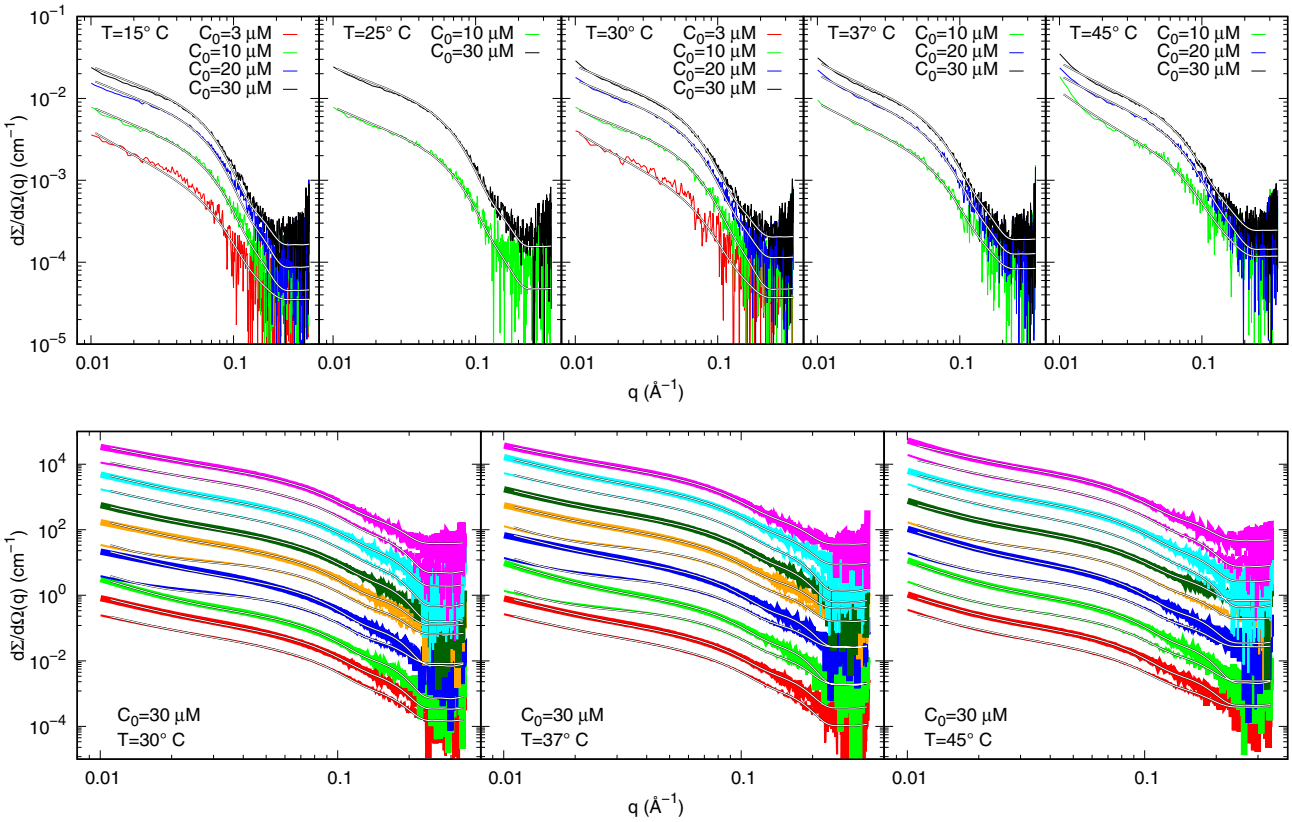


Figure 2. SAXS data and fits. **Top panels**: SAXS experimental data of SARS-CoV-2 M^{pro} without inhibitors and best theoretical fits obtained by GENFIT software[21,22] (solid black and white lines). Each panel reports a dataset obtained at the same temperature, as shown in the top left corner, and at different nominal protein concentration C_o. **Bottom panels**: SAXS data of M^{pro} at fixed concentration C_o=30 μM in the presence of inhibitors. Each panel reports the curves at the same temperature, shown in the bottom left corner. Red, green, blue, orange, dark-green, cyan and magenta refers to inhibitor **1**, **2**, **3**, **4**, **5**, **6** and **7**, respectively. Thin and thick lines refer to inhibitor concentrations C_I of 30 and 60 μM, respectively. Subsequent curves are multiplied by a factor 3.0 for clarity. Solid black and white lines are the best fits obtained by GENFIT[21].

the M^{pro} dimer-monomer equilibrium in solution by SAXS and CD spectroscopy techniques in the absence of inhibitors. Considering the results obtained by Graziano et al.[20] for the SARS-CoV M^{pro}, we have chosen to investigate a range of protein concentrations from 3 and 30 μM, the molarity being expressed in terms of M^{pro} monomers. It should be noted that, using these protein concentrations, one can discriminate between values of the dissociation constant that fall in the quite wide range of $\approx 0.2 - 200\,\mu$M (see Eq. 2).

Subsequently, we have studied by SAXS experiments the M^{pro} dimer-monomer equilibrium in the presence of a series of potential inhibitors, selected from an *in-house* database containing commercial and synthetic compounds. Just one protein concentration, 30 μM, and two concentrations of inhibitors, 30 and 60 μM, corresponding to an inhibitors-to-monomer M^{pro} molar ratio of 1 and 2, have been investigated. Activity assays were also performed and results are correlated with the M^{pro} dimerization inhibition.

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

SAXS. SAXS data of SARS-CoV-2 M^{pro} recorded at the B21 beam-line of the Diamond Synchrotron (Didcot, UK) at different protein concentrations and temperatures are shown as log-log plots in Fig. 2, top panels.

We have assumed that SAXS curves arise from both M^{pro} monomer and dimer species, according to the thermodynamic equilibrium dissociation process given by the relationship:

$$M_2^{pro} \rightleftharpoons 2M_1^{pro}, \tag{1}$$

The corresponding equilibrium dissociation constant is

K_D°	(μM)	7 ± 1
ΔC_{p_D}	(kJ mol^{-1}K^{-1})	1.7 ± 0.7
ΔS_D°	(J mol^{-1}K^{-1})	50 ± 20
d_h		1.100 ± 0.006

Table 1. Thermodynamic parameters resulting from the global fit of SAXS data for SARS-CoV-2 M^{pro} without inhibitors at different temperatures and concentrations.

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

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

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

According to classical thermodynamics, the temperature dependence of ΔG_D is

$$\Delta G_D = \Delta G_D^\circ + (\Delta C_{p_D} - \Delta S_D^\circ)(T - T_\circ) - \Delta C_{p_D} T \log \frac{T}{T_\circ} \tag{4}$$

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

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

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

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

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

where $P_j(q)$ stands for the form factor (which is the orientational average of the excess squared X-ray scattering amplitude) of the M^{pro} monomer ($j = 1$) or dimer ($j = 2$). We have calculated $P_j(q)$ from the the crystal structure of SARS-CoV-2 M^{pro} dimer recently determined[7] (PDB code 6y2e) considering one chain ($j = 1$) or both chains ($j = 2$) by means of the SASMOL method[23]. This method takes into account the contribution to the scattering due to the hydration water molecules around the protein, whose positions are found by embedding the atomic structure in a tetrahedral close packed lattice. For SARS-CoV-2 M^{pro} monomer and dimer, 726 and 1243 hydration water molecules have been respectively calculated, suggesting that for the dimer formation about 200 water molecules are removed from the hydration shell of both monomers. Hence, the water molecules attributed to each monomer decrease from 726 to 621 upon M^{pro} dimerization. This indicates that the dimerization process is accompanied with slight structural changes reducing the average area accessible to solvent. The $S_M(q)$ term in Eq. (5) is the so-called "measured" structure factor, which describes the long range intermolecular interactions among all the particles in solution. For sake of simplicity, here we consider a common effective structure factor that takes into account monomer-monomer, monomer-dimer and dimer-dimer interactions. Considering that at low q all the experimental scattering curves (Fig. 2 top panels) show a positive deviation from a Guinier trend, indicative of the prevalence of protein-protein attraction with respect to repulsion, we have approximated the structure factor by the one of fractal distribution of inhomogeneities developed by Teixeira[24], whose main parameters are D, the fractal dimension of the aggregates, r_0, the effective radius of the aggregating protein molecule and ξ, the correlation length, which can be interpreted as the average size of the aggregates (see Eqs. 10, 11 and 12).

The above described model, which combines SARS-CoV-2 M^{pro} thermodynamic and structural features, has been adopted to simultaneously analyze the whole set of the SAXS curves, recorded at different temperatures and concentrations, shown in Fig. 2, top panels. Fitting parameters shared by all the curves are K_D°, the dissociation equilibrium constant at $T_\circ$, ΔC_{p_D}, the constant pressure heat capacity upon dissociation, ΔS_D° and the dissociation entropy at $T_\circ$. Another parameter common to all the curves is the relative mass density of the hydration water (in general higher than 1), d_h, which is taken into account in the SASMOL method[23]. The shared fitted parameters are shown in Table 1, while all the other are reported in Supplementary Table S1.

The most important parameter obtained by the simultaneous fit of SAXS data is the dissociation constant K_D°, which turns out to be $7 \pm 1\,\mu$M, in good agreement with the value obtained by Graziano et al.[20] on the very similar main protease from SARS-CoV. The corresponding dissociation Gibbs free energy (calculated with

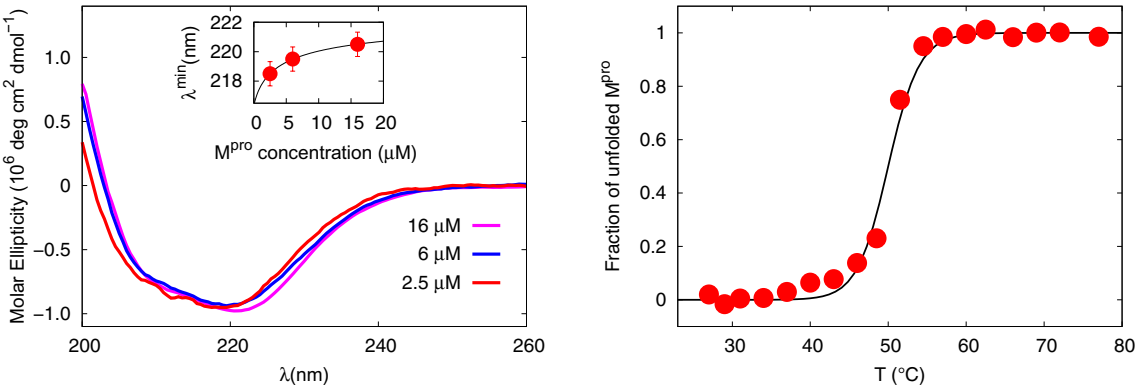


Figure 3. CD data and fits. **Left**: far-UV CD spectra of SARS-CoV-2 M^{pro} at three different concentrations. The CD data are represented in molar ellipticity units. Inset: position of the minimum of the spectra as a function of the concentration (red circles). The continuous line represents an estimate of the minimum position based on Eq. (7). Results from the fit are: $\lambda^{min}_{mon} = 216.4 \pm 0.1$ nm and $\lambda^{min}_{dim} = 222.9 \pm 0.1$ nm. **Right**: Thermal melting of the SARS-CoV-2 M^{pro} (16 μM concentration) followed by monitoring the far-UV CD signal at 221 nm. The continuous line results from the theoretical fitting model arising from Eq. (8).

Eq. 2) is $\Delta G^{\circ}_{D} \simeq 30$ kJ mol^{-1}, a value quite similar to the one observed for the β-lactoglobulin dimer dissociation at neutral pH[25]. Regarding the dissociation entropy, we have obtained a positive value, 50 ± 20 J mol^{-1}K^{-1}, meaningfully smaller with respect to the one derived for the above mentioned β-lactoglobulin case[25]. It should be noticed that in a dissociation process, many factors besides translational and rotational motions contribute to a positive dissociation entropy and it is difficult to separate them. One such factor is certainly the removal of about 200 hydration water molecules from the monomer-monomer interface when the dimer is formed. The change of the heat capacity at constant pressure upon dissociation resulted positive and large. This parameter indirectly describes the monomer-monomer interface, as it can be attributed to the hydration and correlates with the interface size[26]. The set of parameters reported in Table 1 allows to calculate the M^{pro} equilibrium dissociation constant, together with its standard deviation, at any temperature. Results are shown in Supplementary Fig. S1, top left panel. We notice a slight increase of K_D with T, an effect that is mainly due to the large increase of the constant pressure heat capacity upon dissociation. However, a further investigation on the monomer-monomer interface area and its relationship with the dissociation heat capacity[27] requires further calorimetric experiments in order to obtain lower estimation errors. Finally, the relative density of the hydration shell is little more than one, in agreement with previous literature results on globular proteins[28–30]. The determination of the thermodynamic features of the dimer-monomer equilibrium of M^{pro}, in conditions quite similar to those found *in vivo*, is a fundamental step to investigate the effects of drugs aimed to inhibit dimerization and underlines the importance to further investigate M^{pro} monomer-monomer interface by in-solution techniques.

Far-UV CD. To provide further insights on the dimer-monomer equilibrium, we have measured the far-UV CD spectra of M^{pro} at three different concentrations, as shown in Fig. 3, left panel.

At the higher concentration of 16 μM, the ellipticity shows a minimum wavelength λ^{min} at about 221 nm and a shoulder centered at about 208 nm, which are typical of proteins with α-helical and β-sheet content[31,32], fully consistent with the structural features of the SARS-CoV-2 M^{pro}[7], in agreement with CD measurements of the same enzyme[33] and of the very much similar SARS-CoV M^{pro}[34]. As concentration decreases, λ^{min} shifts towards lower values, thus reporting an increase of the β-sheet component at the expense of the α-helical content[31]. Quite interestingly, if we take the dissociation constant of $K_D = 7 \pm 1\,\mu$M, in the concentration range from 2.5 to 16 μM a decrease of about a factor 2 in the fraction of monomers is expected (from around 0.67 to 0.36, see Eq. 3). This large shift in the dimer-monomer populations can reasonably yield the changes occurring to the M^{pro} secondary structure and revealed by CD spectroscopy. Within this working hypothesis, we can describe the λ^{min} trend in terms of the dimer-monomer equilibrium through the following expression:

$$\lambda^{min} = x_1 \lambda^{min}_{mon} + (1 - x_1)\lambda^{min}_{dim} \tag{7}$$

where we take a fixed value of $K_D = 7\,\mu$M as estimated by SAXS, while λ^{min}_{mon} and λ^{min}_{dim} are the minimum wavelength parameters corresponding to the monomer and the dimer spectra, respectively. As shown in the inset of Fig. 3 (left panel), the trend of the λ^{min} values is fitted in an excellent way with Eq. (7).

The thermal stability of the M^{pro} has been characterized by monitoring the signal at 221 nm of the M^{pro} sample at 16 μM concentration, within the simplified hypothesis that the melting curve arises mainly from dimers. The rather sharp transition we have obtained is shown in Fig. 3 (right panel) and clearly suggests a two-state model, where the dimer unfolds and yields two random-coil monomeric chains:

$$M^{pro}_2 \rightleftharpoons 2M^{pro}_{1,unf}. \tag{8}$$

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

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

Considering the scheme 8, if we hypothesize that the dimer can unfold to two random-coil monomeric chains, we obtain an apparent melting temperature T_m of 50° C, with a melting Van't Hoff enthalpy of $\Delta H_v = 810 \pm 60$ kJ/mol. This value is in good agreement with the Van't Hoff enthalpy $\Delta H_v \sim 880$ kJ/mol estimated through the equation $\Delta H_v = 4RT_m^2 C_{p,max}/\Delta H_{cal}$ from DSC measurements[33]. It is also worth of note that, by taking $\Delta H_{cal} = 443$ kJ/mol[33], it turns out a ratio $\Delta H_v/\Delta H_{cal} \sim 1.8$: such a value larger than 1 is fully consistent with the unfolding transition coupled to the dimer dissociation. Quite interestingly, the thermal stability as revealed by CD measurements supports a view where about 90 the folded state in the temperature range investigated by SAXS experiments, i.e. up to 45° C, thus validating the model we used to interpret the corresponding scattering curves.

M^{pro} dimer-monomer equilibrium in presence of inhibitors. *In-silico inhibitor selection.* To identify new inhibitors of SARS-CoV-2 main protease from a large *in-house* database, we applied the *in silico* protocol, recently proposed by some of us[35]. The flowchart of the adopted protocol is depicted in Supplementary Fig. S2. As a first step, we performed molecular docking studies on the compounds present in the database to analyze their binding capability in the catalytic active site of the SARS-CoV-2 M^{pro} (PDB code 6y2f)[7], as detailed in the Materials and Methods section. Supplementary Fig. S3 shows the 3D binding active site of SARS-CoV-2 M^{pro} co-crystallized with the native inhibitor **13b**[7] covalently bonded to Cys145. The ligand binds to the enzymatic catalytic cleft of the protease located between domains I and II. The 3D binding site representation (Supplementary Fig. S3) highlights the interactions with the amino acid residues involved in the inhibition mechanism, such as Met49, Met165, Glu166, His164, Phe140, Gly143 and the catalytic Cys145. It is noteworthy the presence of hydrogen bonds between the pyridone moiety of ligand and Glu166, which rules the catalytic activity driving the SARS-CoV-2 M^{pro} to adopt an inactive conformation. The resulting best docked molecules have been selected based on a docking score cut-off of −6.5 kcal/mol and submitted to ligand based approaches, by taking advantage of the web-service DRUDIT (DRUgs Discovery Tools), an open access virtual screening platform recently developed[36], which represents the evolution of previous well-established protocols based on molecular descriptors[37,38].

DRUDIT implements the ligand based template of SARS-CoV-2 M^{pro}, available in the Biotarget Finder tool, which has been recently proposed as a useful mean in the identification of new SARS-CoV-2 M^{pro} modulators. Subsequently, the ligands selected by molecular docking were submitted to DRUDIT, as elsewhere reported[35], allowing the evaluation of their affinity to SARS-CoV-2 M^{pro} by the values of Drudit Affinity Score (DAS). The features of the ligand-based approaches based on molecular descriptors enabled us to assess topological, thermodynamic and charge-related characteristics of the ligands. Thus, two complementary standpoints in the evaluation of the binding capability (ligand- and structure-based) covered all the interaction aspects in the ligand-target complex. The top scored molecules (selected based on a DAS cut-off of 0.65) were processed by Induced Fit Docking (IFD) calculations to further screen the hits to submit to *in-wet* test. In Supplementary Fig. S4 and in Table 2 the seven best scored structures are reported. The analysis of the results in Table 2 shows as the selected compounds present similar overall scores (IFD_score). This confirms the robustness of the ligand-based approach exploited by DRUDIT, which is able to give an account of the receptor-ligand binding although it is based on molecular descriptors that, as known, do not take into consideration the 3D shape of the binding site.

Figure 4 reports the first two best scored molecules **3** and **7** (according to the IFD_score parameter) in the binding site (left panel) and their related amino-acid maps (right panel). The two molecules are deeply buried in the cleft of the substrate-binding pocket, but unlike the co-crystallized ligand **13b**, they interact with a somehow different pattern of amino-acids. This is evidence suggests that these compounds are not covalently bound to the SARS-CoV-2 M^{pro} catalytic site.

M^{pro} activity assays. The selected inhibitors have been tested for their efficacy to reduce the M^{pro} activity. As reported in Fig. 5, the time dependence of substrate fluorescence after hydrolysis indicates that the catalytic activity of M^{pro} changes in the presence of the selected compounds. In particular, compounds **2**, **4**, **5** and **7** induced an irreversible inactivation of the enzyme, while compounds **1** and **6** resulted rather inactive. For two of the most effective compounds (**2** and **7**) inhibition tests have been carried out as a function of the concentration. Unfortunately, we have not been able to perform this test for compound **4**, which shows the best inhibition efficacy, as it produces a fluorescence signal that partially obscures that of the substrate. Results are shown in Fig. 6 (left panel). Percent inhibition data have been fitted with the Hill equation, $p(C_I) = 100/(1 + (IC_{50}/C_I)^n)$, to

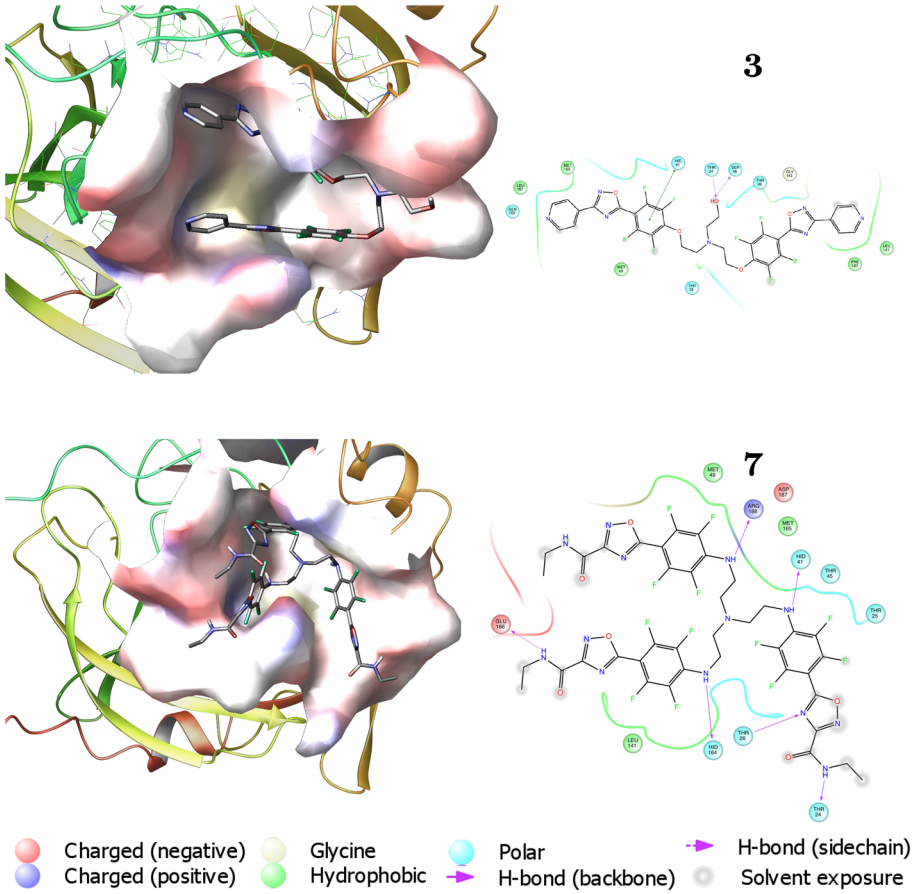


Figure 4. 3D binding modes of best scored compounds **3** and **7** into SARS-CoV-2 M^{pro} active site (**left**) and corresponding amino acid maps (**right**). The picture is elaborated by Maestro Schrödinger, version 10.2 (2017)[39].

get the half maximal effective concentration, IC_{50}, and the Hill slope n. We obtained $IC_{50} = 10.3 \pm 0.2\,\mu M$ for **2** and $15 \pm 2\,\mu M$ for **7**, with $n = 5 \pm 1$ and 3 ± 1, respectively. These values of n larger than one indicate that the binding is positively cooperative, in agreement with other recent experimental results[40].

SAXS. SAXS curves of SARS-CoV-2 M^{pro} at nominal concentration $C_\circ = 30\,\mu M$ and in the presence of each of the seven selected potential inhibitors, at concentrations $C_I = 30\,\mu M$ (thin lines) or $60\,\mu M$ (thick lines), are reported in the bottom panels of Fig. 2 as log-log plots. Each panel refers to a different temperature, as indicated. Unfortunately, some of the foreseen conditions are missing (e.g. SAXS curves of inhibitor 5 at $60\,\mu M$), due to an experimental problem with the sample injection in the beam-line capillary. SAXS data have been analysed with the same approach adopted for data without inhibitors, with the further assumption that, for each compound, the thermodynamic parameters are linear functions of its concentration C_I, namely $\Delta G^\circ_D = \Delta G^\circ_{D,0}(1 + \alpha_G C_I)$, $\Delta C_{PD} = \Delta C_{p\,D,0}(1 + \alpha_{C_p} C_I)$, and $\Delta S^\circ_D = \Delta S^\circ_{D,0}(1 + \alpha_S C_I)$. The three terms $\Delta G^\circ_{D,0} = -RT_\circ \log K^\circ_{D,0}$, $\Delta C_{P\,D,0}$ and $\Delta S^\circ_{D,0}$ are exactly the values already obtained from the analysis of SAXS data without inhibitors (reported in Table 1), and the three corresponding constant rates α_G, α_{C_p} and α_S are fitting parameters common to all the SAXS curves corresponding to the same inhibitor. The high quality of the fitting procedure can be appreciated in Fig. 2 (bottom panels), where the calculated SAXS curves are superposed to the experimental ones and the resulting thermodynamic common fitting parameters are shown in Table 3, first panel.

The inhibitors with the lowest negative values of α_G (Table 3, first panel) are those that mostly favour dimer dissociation. Results reported in Table 3 suggest that compounds **1**, **6**, and **7** are, within the experimental error, mostly able to increase the dissociation equilibrium constant, which at $C_I = 30\,\mu M$ becomes as large as $\approx 15\,\mu M$ and, at $C_I = 60\,\mu M$ almost doubles its value, reaching $\approx 30\,\mu M$. We indeed recall that, in the absence of inhibitors, the value of $K^\circ_{D,0}$ is $7 \pm 1\,\mu M$ (Table 1). Inhibitor **5** is slightly less active: at $C_I = 60\,\mu M$ we found a dissociation equilibrium constant of $\approx 20\,\mu M$. The other three compounds, namely **2**, **3** and **4**, show a value of α_G close to 0, indicating that they do not affect in a significant way the dimer-monomer equilibrium of M^{pro}. Despite the high uncertainties on α_{C_p} and α_S, their negative values suggest that upon dissociation there are changes of heat capacity and of entropy smaller than those observed without inhibitors, indicating that inhibitors increase the monomer order. The temperature dependence of the equilibrium dissociation constant K_D is reported, for each inhibitor, in Supplementary Fig. S1. The large uncertainties on the fitting parameters determines the presence of wide bands of uncertainty on the K_D trends. This is particularly evident for inhibitor **5**, since SAXS data have been

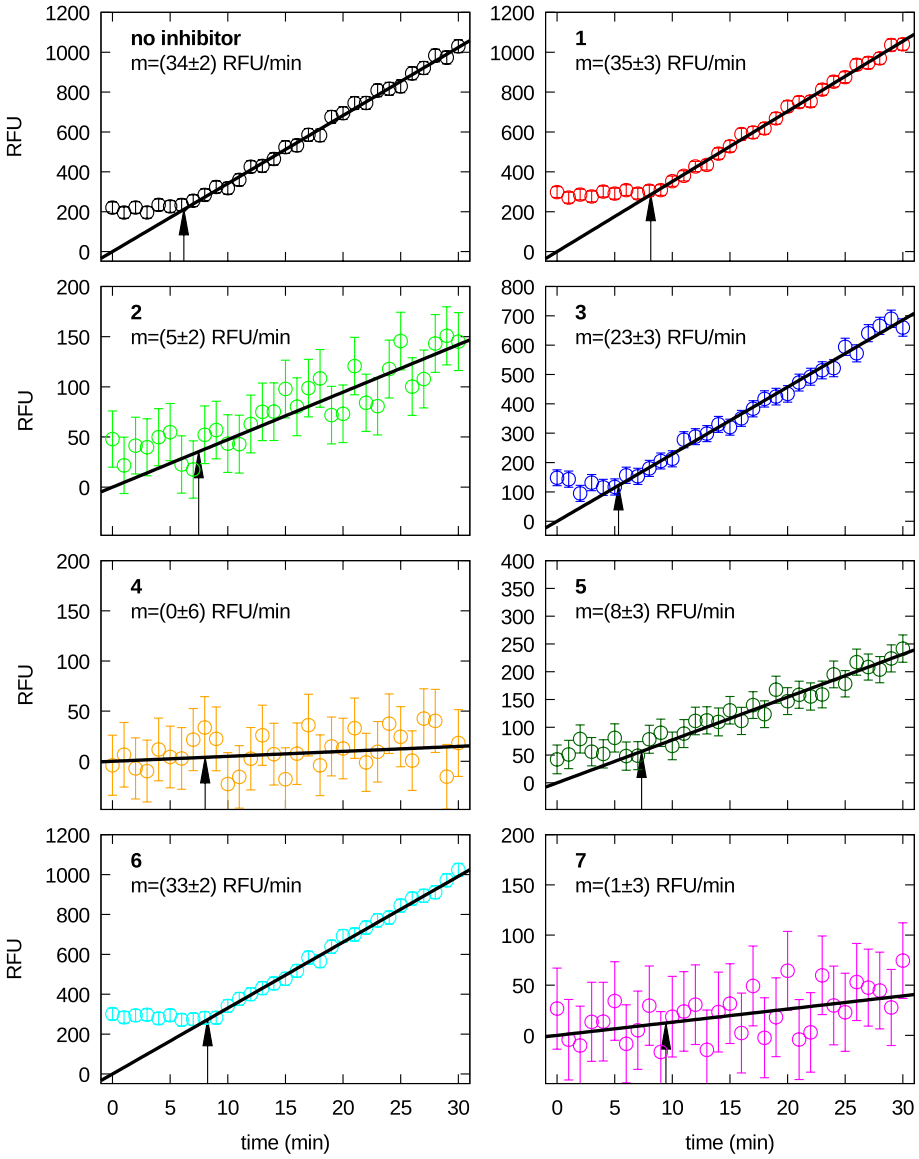


Figure 5. Fluorescence inhibition curves of the selected compounds, as indicated in each frame. à e straight lines are the best fitting lines obtained considering data points comprised between the time indicated by the arrow and 30 min. The slope of the straight line is reported in each frame.

recorded only for one inhibitor concentration. Hence, a word of caution is necessary regarding the temperature dependence of K_D in the presence of the seven inhibitors obtained by the SAXS analysis.

From a close inspection of the single curve parameters, reported in Supplementary Tables S1 and S2, we observe that the value of the correlation length ξ is in the range 3000-4500 Åand rather independent of the temperature and the presence of the inhibitors. à e fractal dimension is ≈ 2, suggesting a two-dimensional fractal growth of protein clusters in the presence of inhibitors.

Discussion

à e active site of M^{pro} monomer, which is highly conserved in different coronaviruses, is typically composed of four subsites, referred to as S1′, S1, S2, and S4[41–43]. à ey accommodate the corresponding domains P1′, P1, P2, and P4 of the substrate or the ones of the inhibitor compound mimicking the substrate[44]. à e S1′ subsite is constituted by the two residues àr 24 and à r25. à e S1 subsite (also referred to as the S1 pocket[43]) is formed by the side chains of residues Phe140, Asn142, Glu166, His163 and His172 and by the main chains of Phe140 and Leu141[44]. As discussed by Sacco et al.[43], S1 is considered a promising target for an inhibiting compound, as it can interact with both hydrophobic and hydrophilic groups. On the other hand, S2 is a hydrophobic subsite formed of the side chains of His41, Met49, Tyr54, Met165 and Asp187, while S4 is a small hydrophobic pocked that involves the side chains of Met165, Leu167, Phe185, Gln192 and Gln189[44]. An unusual catalytic dyad, His41-Cys145, acts in the active site, where His41 is a proton acceptor whereas Cys145 is attacked by the

Inhibitor	α_G ($10^{-2}\,\mu M^{-1}$)	α_{C_p} ($10^{-2}\,\mu M^{-1}$)	α_S ($10^{-2}\,\mu M^{-1}$)
1	-0.18 ± 0.02	-6 ± 3	-4 ± 1
2	-0.0 ± 0.1	-1 ± 3	-4 ± 1
3	0.00 ± 0.08	0 ± 4	-8 ± 4
4	0.0 ± 0.1	-2 ± 2	-1 ± 5
5	-0.14 ± 0.05	-1 ± 4	-4 ± 4
6	-0.20 ± 0.05	-6 ± 3	-1.7 ± 0.8
7	-0.19 ± 0.07	-3 ± 3	-5 ± 2

30 μM Inhibitor		**1**	**2**	**3**	**4**	**6**	**7**
K_D°	(μM)	14 ± 1	8 ± 3	7 ± 2	6 ± 2	15 ± 3	14 ± 3

60 μM Inhibitor		**1**	**2**	**3**	**4**	**5**	**6**	**7**
K_D°	(μM)	26 ± 4	8 ± 6	7 ± 4	5 ± 4	19 ± 7	30 ± 10	30 ± 10

Table 3. Top panel: common thermodynamic fitting parameters of the analysis of SAXS data for SARS-CoV-2 M^{pro} samples with inhibitors. **Middle and bottom panels**: dissociation constants derived by the analysis of SAXS data for SARS-CoV-2 M^{pro} samples with inhibitors.

carbonyl carbon of the substrate. Hence, a signature of the inhibiting power of a compound is its capability to form a covalent bond with Cys145[41], as very recently confirmed by Dai et al.[42], who have found two promising inhibitors **11a** and **11b**. à e importance of the protonation state of Cys145 as well as the network of hydrogen bonds between the catalytic site of M^{pro} and inhibiting compounds has also been recently discussed by Kneller et al.[45] by combining X-ray and neutron scattering data. On these grounds, the experimental results obtained in the present study, together with the structure of the seven inhibitors within the M^{pro} active site determined by the refined molecular docking, can be discussed.

à e interaction map of inhibitor **1** is shown in Supplementary Fig. S5. à ere are a total of 11 contacts with amino acids of M^{pro} monomer (Ser1, àr25, àr26, Ser46, Asn119, Leu141, Asn142, Cys145, Pro168, Arg188, Gln192), 2 of them (Ser46, Asn119) are hydrogen bonds. à e residues of the catalytic dyad and the four subsites in contact with **1** are: Cys145 (dyad, 1 of 2 (50 of 2 (50 5 (20 To note, these contacts involve only one of the residues of the catalytic dyad, Cys145, without a hydrogen bond, whereas for inhibitor **13b** there is a hydrogen bond with Cys145 (Supplementary Fig. S3). It is also worth to notice that Ser1 is among the residues in contact with inhibitor **1**: since the mutual interaction of Glu166 of one monomer and the N-finger residues of the other monomer, like Ser1, has been proven to shape the catalytic cleft[7], we argue that this compound could destabilize the dimer, consistently with the high value of $K_D^\circ = 26 \pm 4\,\mu M$ at $C_I = 60\,\mu M$. However, its enzymatic inhibition is very poor, as shown by the high similarity of the RFU slope with the one in the absence of inhibitors (Fig. 5). A possible explanation of this result could be the absence of any hydrogen bond with Cys145 as well as the absence of any contact with the residues of subsite S2.

Regarding inhibitor **2**, the map of contacts shown in Supplementary Fig. S6 reveals a total of 10 interactions with the monomer chain (Met49, Asn142, Gly143, Cys145, Asp187, His164, Met165, Glu166, Arg188, Gln189); 4 of them are hydrogen bonds (Asn142, His164, Glu166, Gln189) that do not involve the catalytic site. More in detail, the residues of the catalytic dyad and the four subsites in contact with **2** are: Cys145 (dyad, 1 of 2 (50 Asn142 and Glu166 (S1, 2 of 6 (33 of 5 (60 We also note that 2 residues of S1, Asn142 and Glu166, interact with this inhibitor via a hydrogen bond. ài s evidence, together with the high number of contacts with S2 and S4, could explain the experimentally observed inhibition effect ($m = 5 \pm 2$ RFU/min, Fig. 5). To note, this compound does not modify the dimer-monomer equilibrium, being the fitting parameter α_G almost 0 (Table 3) within the experimental error.

Turning now to inhibitor **3**, we have found that it does not alter the native dissociation equilibrium of M^{pro} ($\alpha_G \approx 0$, see Table 3) and also its inhibition effect, observed by fluorescence analysis, is week ($m = 23 \pm 3$ RFU/min, slightly lower that the value in absence of inhibitors). à e contact map of compound **3** (Fig. 4) shows a total of 12 interactions with the monomer chain (àr24, àr25, àr26, H is41, Ser46, Met49, Phe140, Leu141, Gly143, Met165, Leu167, Gln192), 2 of them being hydrogen bonds (à r24, Ser46) and one (His41) a $\pi - \pi$ stacking. à e residues of the catalytic dyad and the four subsites in contact with **3** are: His41 (dyad, 1 of 2 (50 (S1′, 2 of 2 (100 His41, Met49 and Met165 (S2, 3 of 5 (60 (S4, 3 of 5 (60 We notice that with respect to compound **2**, there are no hydrogen bonds with the five residues that stabilize the S1 pocket. ài s difference might be the reason for the weak inhibition effect.

Compound **4** is the most effective among the seven inhibitors ($m \approx 0$, Fig. 5) even if it shows at $C_I = 30\,\mu M$ a dissociation equilibrium constant $K_D^\circ = 6 \pm 2\,\mu M$ slightly lower than the one without inhibitors (see Table 1). Indeed the parameter α_G is positive but very close to 0 within the experimental error (Table 3), suggesting that compound **4** provokes a very weak stabilizing effect of the dimer. à e contact map (Supplementary Fig. S7) shows 11 interactions with the monomer chain (à r25, à r26, Leu27, His41, Ser46, Met165, Glu166, Pro168, Gln189, àr190, G ln192), including one hydrogen bond (Gln189). à e residues of the catalytic dyad and the four subsites in contact with **4** are: His41 (dyad, 1 of 2 (50 of 2 (50 (40 Only one of the five residues that stabilize the S1 pocket is among the ones in contact with this inhibitor, Glu166, which does not form a hydrogen bond.

On this ground, the high inhibition effect of compound **4** could be only justified by the contact with His41, one of the two residues of the catalytic dyad.

Results are different for compound **5**: at $C_I = 60\,\mu$M (unfortunately no data are available at 30 μM) it provokes a rather important increase of K_D° (Table 3) and shows a moderate inhibition effect ($m = 8 \pm 3$ RFU/min, see Fig. 5). Looking at the interaction map (Supplementary Fig. S8), we notice 11 interactions with monomer M^{pro} (ȧr 25, ȧ r26, Leu27, His41, Asn142, Gly143, Met165, Glu166, Pro168, Arg188, Gln192), one of them regards Glu166, involved in two hydrogen bonds, and the other one His41, involved in two a $\pi - \pi$ stacking interactions. ȧ e residues of the catalytic dyad and the four subsites in contact with **5** are: His41 (dyad, 1 of 2 (50 of 2 (50 Met165 (S2, 2 of 5 (40 ȧ ere are not hydrogen bonds involving the five residues that stabilize the S1 site. One could speculate that this inhibitor, probably due to its steric hindrance, provokes a modification of S1 that could interfere with the enzymatic activity of M^{pro}.

Compound **6** determines 11 contacts with the amino acid of the monomer (His41, Leu141, Ser144, Cys145, Met165, Glu166, Leu167, Arg188, Gln189, Ala191, Gln192, Supplementary Fig. S9), including two hydrogen bond (His41, Gln192). In particular, the residues of the catalytic dyad and the four subsites in contact with **6** are: His41 and Cys145 (dyad, 2 of 2 (100 2 of 5 (40 An almost absent inhibition effect is seen by fluorescence, being the slope of RFU (Fig. 5) very similar to the one determined in the absence of inhibitors. On the other side, compound **6** is able to modify the dimer-monomer dissociation, with one of the highest values of $K_D^\circ = 30 \pm 10\,\mu$M at $C_I = 60\,\mu$M (Table 3). To note, only one of the 6 amino acids that stabilize the S1 site are included in the list of residues interacting with compound **6**. Hence, the absence of its inhibition activity could be explained by the small size of its molecular structure, which might not be able to provoke important modifications of the S1 pocket and hence to modify the catalytic features of M^{pro}.

We finally turn to compound **7**. It shows an opposite behaviour with respect to compound **6**: it is capable to change the dimer-monomer equilibrium at the same extent ($K_D^\circ = 30 \pm 10\,\mu$M at $C_I = 60\,\mu$M, Table 3) and displays a promising inhibition effect, with $m \approx 1$. For this compound, the map of contacts shows 12 interactions (ȧr24, ȧr25, ȧr26, H is41, ȧr45, M et49, Leu141, His164, Met165, Glu166, Asp187, Arg188, Fig. 4) with a large number of hydrogen bonds (ȧr 24, ȧ r26, His41, His164, Glu166, Arg188). ȧ e residues of the catalytic dyad and the four subsites in contact with **7** are: His41 (dyad, 1 of 2 (50 (S1′, 2 of 2 (100 His41, Met49, Met165 and Asp187 (S2, 4 of 5 (80 5 (20 Only one of the interacting residues (Glu166) is involved in the stabilization of the S1 pocket. We can infer the high inhibition effect could be due to the high number of contact with S2 and to the presence of 6 hydrogen bonds. Another hypothesis, which needs further insights to be confirmed, is that the fluorinated groups, which are present in a high number in compound **7**, may originate a new reactive warhead able to form a covalent bond with Cys145. We may also consider that one of them involves a residue of the catalytic dyad, His41, suggesting a possible important modification of the enzymatic activity.

Conclusions

Considering the results that we have discussed above, a picture emerges where the selected compounds designed to bind the catalytic site of SARS-CoV-2 M^{pro} may affect dimerization and enzymatic activity processes to a different extent. Since the functional form of M^{pro} is a dimer, compounds that disrupt dimerization should be in principle also effective at diminishing its catalytic activity. However, the compound-induced shift in the dimer and monomer thermal equilibrium populations may not directly translate into a loss of enzymatic activity. Indeed, the latter strongly depends also on the local interactions occurring at the catalytic site, that in turn governs the competition between inhibitor and substrate. To better visualize the scenario presented by our findings, we report in Fig. 6 (right panel) the slope m of the fluorescence inhibition curve as a function of the dimer-monomer equilibrium constant K_D calculated from the set of fitting parameters, derived by the global fit of SAXS data of the seven compounds, by fixing the inhibitor concentration C_I at 60 μM and the temperature at 30° C, the same value employed in the enzymatic activity assays. ȧ e points in this map could be organized in two groups, as represented in blue and in red. ȧ e latter group refers to compounds **3**, **5** and **7**, which show the expected behaviour: the stronger is their capability to induce the dissociation of the M^{pro} dimer, the more important is their inhibiting effect. For these compounds the molecular mechanisms underlying inhibition at the active site are likely linked with their ability to provoke dimer dissociation. On the other hand, for compounds **1**, **2**, **4**, and **6**, displayed in blue, we find the opposite behavior: the increase of dimer dissociation does not determine an increase of inhibition, thus suggesting that the molecular mechanisms of inhibition at the active site play a major role with a marginal involvement of the monomer-monomer interface. ȧi s apparently contradictory result can be in part explained by considering that, in all cases, the dissociation equilibrium is weak so that, in the presence of a compound that alters the dimer-monomer equilibrium but that does not hamper the interaction with the substrate, there are always dimeric M^{pro} molecules that can exert their enzymatic activity when a substrate is available.

A further complementary explanation could be the existence of ligand binding sites alternative to the orthosteric active site located at the interface of dimeric M^{pro}. Indeed, the presence of two of such binding sites, not directly involved in enzymatic inhibition but probably interfering with dimerization, has been very recently revealed by a molecular dynamics simulation study in the SARS-CoV-2 M^{pro}[46].

We can provide a qualitative interpretation of the behavior exhibited by the two groups of compounds by looking at the contact maps of the residues grouped in the different sub-sites of the active site. We note that both compound **6** and **7**, which show the higher values of K_D, are in contact with two residues of the S1, Leu141 and Glu166. Besides, compound **1**, which has a slightly lower K_D, interacts with S1 through the contacts with Leu141 and Asn142. ȧ is suggests that binding with at least two of the residues Glu166, Leu141 and Asn142 is crucial to modify the dimer-monomer equilibrium. On the other hand, the interaction of compounds **7**, **2** and **5** with Glu166 via an hydrogen bond is likely linked to their high inhibiting action. On the contrary, there is

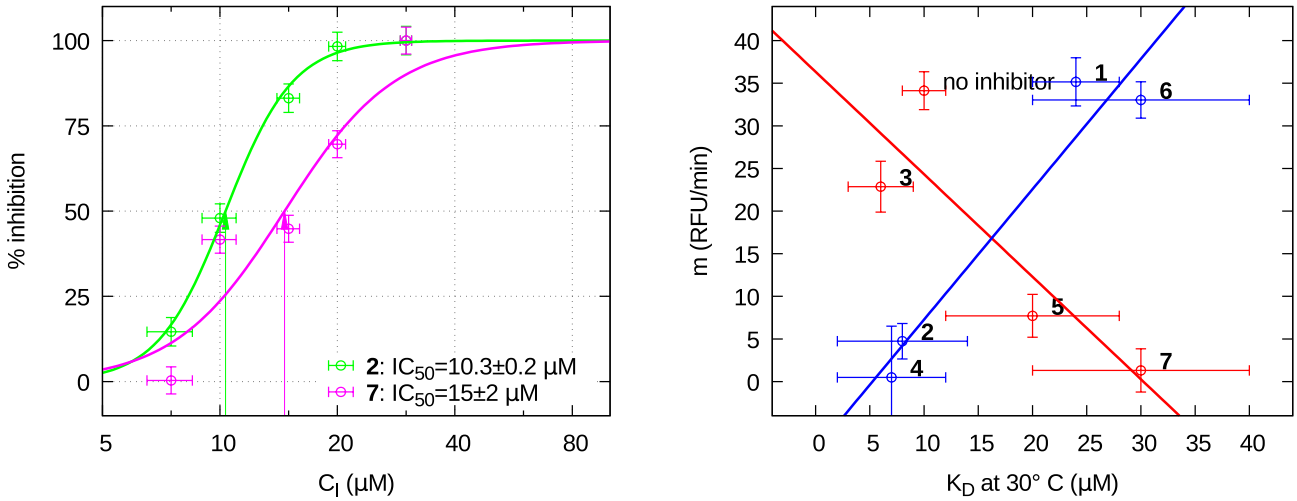


Figure 6. **Left**: percent inhibition data of SARS-CoV-2 M^{pro} as a function of the concentration of inhibitor **2** (green points) and **7** (magenta points). Best fits with the Hill equation are shown as solid lines. **Right**: correlation map between the catalytic activity, represented by the RFU slope m, and dimer dissociation capability, measured by the dissociation constant K_D at 30° C, of the seven SARS-CoV-2 M^{pro} inhibitors at $C_I = 60\,\mu$M.

no hydrogen bond between compound **6** and Glu166. Hence, on the basis of the results obtained for the seven selected compounds, the interplay between SAXS results, enzymatic activity assays and contact map analysis suggests a relevant clue: in order to promote both M^{pro} dimer dissociation and the inhibition of its catalytic activity, a small molecule should interact with at least two residues of the S1 sub-site and most likely form an hydrogen bond with Glu166. The key role of Glu166 residue, which is conserved among all human coronaviruses, for inhibition has been pointed out also very recently[47].

To note, according to Goyal and Goyal[6], Glu166 is among the residues that should be targeted to inhibit the dimerization of SARS-CoV M^{pro}. However, for a more detailed investigation of the dimerization process in stabilizing the catalytic activity of M^{pro}, it is also important to take into account the overall contribution of protein flexibility, as recently evidenced by Suárez et al.[48], through a 2 μs Molecular Dynamics simulation of M^{pro} with and without a model peptide mimicking the enzyme substrate.

In summary, the experimental work presented here brings basic information to decipher the complex interplay between enzymatic activity inhibition and dimer dissociation. To the best of our knowledge, we have shown for the first time how structural information about the SARS-CoV-2 M^{pro} in solution in the absence and in the presence of potential inhibitors and as a function of temperature can be obtained from an advanced analysis of SAXS data within an overall thermodynamic picture, complemented by more conventional approaches. Our results suggest that more experimental evidences about the impairment of monomer and dimer M^{pro} in the presence of inhibitors corroborated by computational information will be necessary for a deeper understanding of the M^{pro} allosteric mechanism.

Materials and methods

M^{pro} expression and western blot analysis. pGEX-6P-1 vector harboring the full length cDNA sequence encoding for SARS-CoV-2 Main Protease (M^{pro} NC_045512) was purchased from GenScript (clone ID_M16788F). The expressing vector was transformed into BL21DE3pLys *Escherichia coli* cells and the obtained clones were assayed both in small scale (5 mL) and medium scale (500 mL and 1 L) for the production of SARS-CoV-2 M^{pro}. Transformants were grown onto LB medium containing 100 μg/mL Ampicillin and 34 μg/mL Chloramphenicol as selective antibiotics. Cultures were grown up to OD600 of 0.6-0.8 at 37° C, 200 rpm and then M^{pro} expression was induced by addition of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Growth under induction was achieved both for 3 h at 37° C and 10 h at 16° C in order to test the best expressing condition. Cells were harvested by centrifugation at 6000 g. Cell pellets were resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 2 mM β-mercaptoethanol), and cell rupture was achieved by sonication (Sonics Vibra Cell sonicator) at 4° C. Cell debris was separated from the total protein extract by centrifugation at 6500 g for 1 h. Supernatant aliquotes were resuspended in Laemmli sample buffer, run onto 12 polyvinylidene difluoride membrane for Western blot analysis. M^{pro} was decorated by 6×-His tag monoclonal primary antibody (Invitrogen) and anti-mouse secondary antibody and detected by chemiluminescence (Clarity Western ICL Substrate, Biorad, Supplementary Fig. S11, panels A and B).

M^{pro} purification and His-tag cleavage. The total cell extract was loaded onto Ni-NTA affinity column (G-Biosciences) and washed by washing buffer (Tris-HCl 20 mM pH 7.6, NaCl 100 mM). M^{pro} was eluted by elution buffer (Tris-HCl 20 mM pH 7.6, NaCl 100 mM, 300 mM imidazole) in 5 fractions of 1 mL each. Aliquotes of elution fractions were loaded onto 12 acrylamide gel and imidazole was removed by dialysis against Prescission cleavage buffer (Tris-HCl 50 mM pH 8.0, NaCl 150 mM, dithiothreitol 1 mM, ethylenediaminetetraacetic

acid 1 mM) through Amicon Ultra-4 centrifugal filters 30K (Merck Millipore). For M^{pro} C-terminal His-tag removal, the Prescission (1 U for 100 μg of protein) cleavage reaction was performed at 4° C for 4 h and Prescission protease was then removed by GSTrap FF column (GE-Healthcare). The M^{pro} solution was further purified by FPLC size-exclusion chromatography on Superdex 75 10/300 GL column (Supplementary Fig. S10, panels A and B)[33,44].

M^{pro} activity assay and inhibition. The fluorescently labelled auto-cleavage sequence of SARS-CoV-2 M^{pro}, ((7-Methoxycoumarin-4-yl)acetyl)-AVLQ↓SGFRK(2,4-dinitrophenyl)K (purchased from GenScript), was utilized to monitor the recombinant M^{pro} kinetics (excitation 320 nm, emission 405 nm). The assay was started by mixing ≈ 0.2 μM SARS-CoV-2 M^{pro} to different amounts of substrate (10, 20, 40 μM) in order to set the best protein-substrate concentration to detect M^{pro} activity[41]. Fluorescence intensity was measured by DeNovix DS-11 FX+ fluorometer. The M^{pro} activity reported as reference for inhibition tests was obtained by linear fitting of the fluorescence curve in the presence of 40 μM of substrate concentration[33,41]. Seven inhibitors dissolved in dimethyl sulfoxide (DMSO) were tested at a final concentration of 30 μM[42] (Fig. 5). Each reaction in a final volume of 200 μL was firstly incubated for 20 min at 30° C without substrate. After substrate addition, fluorescence intensities were reported as relative fluorescence units (RFU) and monitored every minute for a duration of 30 min at 30° C.

Circular dichroism. In-house circular dichroism experiments were performed at room temperature using a JASCO J-810 spectropolarimeter (Physics and Geology Department, University of Perugia). Quartz cuvettes with path-length of 1 mm was used, in order to obtain the optimum signal-to-noise ratio for the M^{pro} samples with concentrations of 16, 6 and 2.5 μM respectively. Protein concentration was measured by performing absorption measurements on the same samples, with an extinction coefficient of 33640 M^{-1} cm^{-1} estimated from amino acid sequence Expasy online ProtParam tool[49]. Each spectrum was collected in the range from 200 to 260 nm with a scan speed of 50 nm/min, and repeated three times. The CD data are represented in molar extinction units, by using the formula $[\Theta] = m^{\circ}/(10\,C\,L)$, where m° is the ellipticity (in millidegree unit), C is the protein molar concentration and L is the path length of cell (in cm). The thermal stability has been studied at 16 μM M^{pro} concentration, by varying the temperature through a thermal bath from 27° to 77° C.

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

SAXS data analysis approach has been described in the main text, with the exception of some minor points. Since in all conditions the nominal molar protein concentration is lower than 1 mM, its temperature variations can be considered to be only determined by the dependency with T of the relative mass density of water, which, according to literature results[50] is written as

$$d_w = e^{-\alpha_w(T-T_\circ)-\beta_w(T-T_\circ)^2/2}, \tag{9}$$

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

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

$$S_M(q) = 1 + \beta(q)[S(q) - 1] \tag{10}$$

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

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

and $P^{(1)}(q)$ is the average of the protein excess scattering amplitude, a function provided, together with $P(q)$ by the SASMOL method. According to Ref.[24], $S(q)$ has been written as

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

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

In-silico design. *Ligand preparation.* The default setting of the LigPrep tool implemented in Schrödinger's software (version 2017-1) was used to prepare the ligands for docking[51]. All possible tautomers and combination of stereoisomers were generated for pH 7.0 ± 0.4, using the Epik ionization method[52]. Energy minimization was subsequently performed using the integrated OPLS 2005 force field[53].

Protein preparation. The crystal structure of SARS-CoV-2 M^{pro} in complex with ligand **13b** (PDB code 6y2f)[7] was downloaded from the Protein Data Bank[54]. The cocrystal ligand, covalently bonded to Cys145, was treated by breaking the covalent bond and filling in open valence. Protein Preparation Wizard of Schrödinger software was subsequently employed for further preparations of the protein structure using the default settings[55]. Bond orders were assigned, and hydrogen atoms as well as protonation of the heteroatom states were added using the Epik-tool (with the pH set at biologically relevant values, i.e. at 7.0 ± 0.4). The H-bond network was then optimized. The structure was subjected to a restrained energy minimization step (RMSD of the atom displacement for terminating the minimization was 0.3 Å), using the Optimized Potentials for Liquid Simulations (OPLS) 2005 force field[53].

Docking validation. Molecular Docking was performed by the Glide program[37,56,57]. The receptor grid preparation was performed by assigning the original ligand (**13b**) as the centroid of the grid box. The generated 3D conformers were docked into the receptor model using the Standard Precision (XP) mode as the scoring function. A total of 5 poses per ligand conformer were included in the post-docking minimization step, and a maximum of 2 docking poses were generated for each ligand conformer. The proposed docking procedure was validated by the re-dock of the crystallized **13b** within the receptor-binding pockets of 6y2f by Glide covalent docking. The results obtained were in good agreement of the experimental poses, showing a RMSD of 0.75 Å.

Biotarget finder module (DRUDIT). The refined selection of suitable SARS-CoV-2 M^{pro} inhibitors was performed through the module Biotarget Finder as available in the www.drudit.com webserver[36]. The tool allows to predict the binding affinity of candidate molecules versus the selected biological target. The template of the biological target was built as previously reported. Thus, the in-house database was submitted to the Biological Predictor module by setting the DRUDIT parameters, *N*, *Z*, and *G*, using the crystallized structure of **13b**, as previously reported[35].

Induced fit docking. Induced fit docking simulation was performed using the IFD application as available[38,58] in the Schrödinger software suite[39], which has been demonstrated to be an accurate and robust method to account for both ligand and receptor flexibility[59]. The IFD protocol was performed as follows[60,61]: the ligands were docked into the rigid receptor models with scaled down van der Waals (vdW) radii. The Glide Standard Precision (XP) mode was used for the docking and 20 ligand poses were retained for protein structural refinements. The docking boxes were defined to include all amino acid residues within the dimensions of 25 Å× 25 Å×25 Å from the centre of the original ligands. The induced-fit protein-ligand complexes were generated using Prime software[39,62,63]. The 20 structures from the previous step were submitted to side chain and backbone refinements. All residues with at least one atom located within 5.0 Å of each corresponding ligand pose were included in the refinement by Prime. All the poses generated were then hierarchically classified, refined and further minimized into the active site grid before being finally scored using the proprietary GlideScore function defined as follows: XPG_score = 0.065 vdW + 0.130 Coul + Lipo + Hbond + Metal + BuryP + RotB + Site, where vdW is the van der Waals energy term, Coul is the Coulomb energy, Lipo is a lipophilic contact term that rewards favourable hydrophobic interactions, Hbond is an H-bonding term, Metal is a metal-binding term (where applicable), BuryP is a penalty term applied to buried polar groups, RotB is a penalty for freezing rotatable bonds and Site is a term used to describe favourable polar interactions in the active site. Finally, IFD_score (IFD_score = XPG_score + 0.05 Prime_Energy), which accounts for both protein-ligand interaction energy and total energy of the system, was calculated and used to rank the IFD poses. More negative IFD_score values indicated more favourable binding. Results are shown in Table 2.

Chemical synthesis of inhibitors. Inhibitors **1**[64], **3**[64], **5**[64] and **6**[65] have been prepared as previously reported. Inhibitor **2** is commercial. Inhibitors **4**[64] and **7**[64] have been synthesized as described in detail in the next paragraphs. All solvent and reagents were used as received, unless otherwise stated. Melting points were determined on a hot-stage apparatus. ^{1}H-NMR and ^{13}C-NMR spectra were recorded at indicated frequencies, residual solvent peak was used as reference. Chromatography was performed by using silica gel (0.040-0.063 mm) and mixtures of ethyl acetate and petroleum ether (fraction boiling in the range of 40-60° C) in various ratios (v/v). Compounds **8**[64] and **9**[66], used in the synthesis of inhibitors **4** and **7**, have been prepared as previously reported.

Synthesis of inhibitor 4. Inhibitor **4** was synthesized through a nucleophilic aromatic substitution (S_NAr) of 5-pentafluorophenyl-1,2,4-oxadiazole **8** with 1-Aza-18-crown-6 in para position (Supplementary Fig. S12). Oxadiazole **8** (312 mg, 1 mmol) was dissolved in acetonitrile (5 mL). 1-Aza-18-crown-6 (289 mg, 1.1 mmol) and potassium carbonate (152 mg, 1.1 mmol) were added and the suspension was stirred at room temperature for 24 h. The reaction was monitored by TLC. The reaction mixture was dried under vacuum and treated with H_2O (50 mL) before extraction three times with EtOAc (50 mL each). The combined organic layers were dried with Na_2SO_4 and then concentrated in vacuo to give the crude product, which was recrystallized from EtOH. **16-(2,3,5,6-tetrafluoro-4-(3-phenyl-1,2,4-oxadiazol-5-yl)phenyl)-1,4,7,10,13-pentaoxa-16-azacyclooctadecane**: yield: 73 *δ*: 3.68-3.80 (m, 24H, overlapped $-CH_2-$ signals), 7.51-7.54 (m, 3H, Ar), 8.16-8.20 (m, 3H, Ar). FTIR (Nujol) 1647, 1529, 1518 cm^{-1}; HRMS-ESI [$(M+H)^+$]: *m/z* calculated for $(C_{26}H_{30}F_4N_2O_6)^+$: 556.2060; found, 556.2046.

Synthesis of inhibitor 7. Tripodal oxadiazolylamide (inhibitor **7**) was easily obtained by means of nucleophilic displacement with ethylamine from tripodal ester **9**, which was previously reported as heavy metal fluores-

cent sensor (Supplementary Fig. S13)[66]. Tripodal **9** (101 mg, 0.1 mmol) was dissolved in acetonitrile (3 mL). Ethylamine (2 M in MeOH, 150 μL, 0.3 mmol) was added and the solution was stirred at room temperature for 24 h. The reaction was monitored by TLC. The reaction mixture was dried under vacuum and treated with H_2O (50 mL) before extraction three times with EtOAc (50 mL each). The combined organic layers were dried with Na_2SO_4 and then concentrated in vacuo to give the crude product, which was purified by chromatography. **5,5',5"-(((nitrilotris(ethane-2,1-diyl))tris(azanediyl))tris(2,3,5,6-tetrafluorobenzene-4,1-diyl))tris(N-ethyl-1,2,4-oxadiazole-3-carboxamide)**: yield: 66 d $_6$) δ: 1.18 (t, 9H, J = 6.9 Hz, CH_3), 2.85 (bs, 6H, NCH_2CH_2 NH-), 3.34-3.48 (m, 6H, CH_3CH_2NH-), 3.56 (bs, 6H, NCH_2CH_2NH-), 6.91 (bs, 3H, NCH_2CH_2NH-), 9.02 (t, 3H, J = 5.7 Hz, $NHCH_2CH_3$). FTIR (Nujol) 3325, 1701, 1680, 1647 cm^{-1}; HRMS-ESI [$(M+H)^+$]: m/z calculated for $(C_{39}H_{34}F_{12}N_{13}O_6)^+$: 1007.2485; found, 1007.2518.

Received: 24 December 2020; Accepted: 14 April 2021
Published online: 29 April 2021

References

1. Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. Lancet 395, 470–473 (2020).
2. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270+, https://doi.org/10.1038/s41586-020-2012-7 (2020).
3. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. & Hilgenfeld, R. Coronavirus main proteinase (3CL(pro)) structure: Basis for design of anti-SARS drugs. Science 300, 1763–1767, 2003, https://doi.org/10.1126/science.1085658.
4. Morse, J. S., Lalonde, T., Xu, S. & Liu, W. R. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. ChemBioChem 21, 730–738 (2020).
5. Zhang, D. et al. Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. Infection genetics and evolution 10, 84–88 (2010).
6. Goyal, B. & Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum therapeutic Strategy. ACS Combinatorial Science 22, 297–305 (2020).
7. Zhang, L. et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science 368, 409–412 (2020).
8. Shi, J. & Song, J. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. FEBS Journal 273, 1035–1045 (2006).
9. Anand, K. et al. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. Embo Journal 21, 3213–3224, 2002, https://doi.org/10.1093/emboj/cdf327.
10. Chen, S. et al. Severe acute respiratory syndrome coronavirus 3C-like proteinase n terminus is indispensable for proteolytic activity but not for enzyme dimerization - Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. Journal of Biological Chemistry 280, 164–173 (2005).
11. Chou, C. et al. Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. Biochemistry 43, 14958–14970 (2004).
12. Spinozzi, F., Gazzillo, D., Giacometti, A., Mariani, P. & Carsughi, F. Interaction of proteins in solution from small angle scattering: a perturbative approach. Biophys. J. 82, 2165–2175 (2002).
13. Gottschalk, M., Nilsson, H., Roos, H. & Halle, B. Protein self-association in solution: the bovine β-lactoglobulin dimer and octamer. Protein Science 12, 2404–2411 (2003).
14. Ortore, M. G. et al. High pressure small-angle neutron scattering study of the aggregation state of β-lactoglobulin in water and water/ethylene glycol solutions. Chem. Phys. Lett. 418, 338–342 (2005).
15. Russo, D. *et al.* The impact of high hydrostatic pressure on structure and dynamics of beta-lactoglobulin. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1830**(10), 4974–4980 (2013).
16. Molodenskiy, D. et al. Thermally induced conformational changes and protein-protein interactions of bovine serum albumin in aqueous solution under different ph and ionic strengths as revealed by saxs measurements. Phys. Chem. Chem. Phys. 19, 17143–17155 (2017).
17. Sauter, A. et al. Structural evolution of metastable protein aggregates in the presence of trivalent salt studied by (v)sans and saxs. The Journal of Physical Chemistry B 120, 5564–5571 (2016).
18. Salter, J. D., Krucinska, J., Raina, J., Smith, H. C. & Wedekind, J. E. A hydrodynamic analysis of apobec3g reveals a monomer-dimer-tetramer self-association that has implications for anti-hiv function. Biochemistry 48, 10685–10687 (2009).
19. Deyaert, E. et al. A homologue of the Parkinson's disease-associated protein LRRK2 undergoes a monomer-dimer transition during GTP turnover. Nat. Commun. 8, 1008 (2017).
20. Graziano, V., McGrath, W. J., Yang, L. & Mangel, W. F. SARS CoV Main Proteinase: The Monomer-Dimer Equilibrium Dissociation Constant. Biochemistry 45, 14632–14641 (2006).
21. Spinozzi, F. Genfit, version 2020. (2020).
22. Spinozzi, F., Ferrero, C., Ortore, M. G., Antolinos, A. D. M. & Mariani, P. GENFIT: software for the analysis of small-angle X-ray and neutron scattering data of macromolecules in-solution. J. App. Cryst. 47, 1132–1139 (2014).
23. Ortore, M. G. et al. Combining structure and dynamics: non-denaturing high-pressure effect on lysozyme in solution. J. R. Soc. Interface 6, S619–S634 (2009).
24. Teixeira, J. Small-angle scattering by fractal systems. J. Appl. Cryst. 21, 781–785 (1988).
25. Apenten, R., Khokhar, S. & Galani, D. Stability parameters for β-lactoglobulin thermal dissociation and unfolding in phosphate buffer at pH 7.0. *Food Hydrocolloids* **16**, 95 – 103 (2002).
26. Janin, J. Elusive affinities. *Proteins: Struct. Funct. Bioinform.* **21**, 30–39 (1995).
27. Horton, N. & Lewis, M. Calculation of the free energy of association for protein complexes. Protein Science 1, 169–181 (1992).
28. Svergun, D. et al. Protein hydration in solution: experimental observation by X-ray, neutron scattering. Proc. Natl. Acad. Sci. USA 95, 2267–2272 (1998).
29. Sinibaldi, R. et al. Preferential hydration of lysozyme in water/glycerol mixtures: A small-angle neutron scattering study. J. Chem. Phys. 126, 235101–235109 (2007).
30. Sinibaldi, R. et al. Sans/saxs study of the bsa solvation properties in aqueous urea solutions via a global fit approach. Eur. Biophys. J. 37, 673–681 (2008).
31. Greenfield, N. J. & Fasman, G. D. Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. Biochemistry 8, 4108–4116 (1969).

32. Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nature Protocols 1, 2527–2535 (2006).
33. Abian, O. et al. Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experimental screening. International Journal of Biological Macromolecules 164, 1693–1703 (2020).
34. Shi, J., Wei, Z. & Song, J. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme - Defining the extra domain as a new target for design of highly specific protease inhibitors. Journal of Biological Chemistry 279, 24765–24773 (2004).
35. Martorana, A., Gentile, C. & Lauria, A. In Silico Insights into the SARS CoV-2 Main Protease Suggest NADH Endogenous Defences in the Control of the Pandemic Coronavirus Infection. Viruses 12, 805 (2020).
36. Lauria, A. et al. DRUDIT: web-based DRUgs DIscovery Tools to design small molecules as modulators of biological targets. Bioinformatics 36, 1562–1569 (2019).
37. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
38. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. Journal of Medicinal Chemistry 49, 534–553 (2006).
39. Schrödinger, N. Y., LLC. Maestro, version 10.2 (2017).
40. Lee, J. *et al.* Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *NATURE COMMUNICATIONS* **11**, https://doi.org/10.1038/s41467-020-19662-4 (2020).
41. Yang, H. et al. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLOS Biology 3, e324 (2005).
42. Dai, W. et al. Structure-based design of antiviral drug candidates targeting the sars-cov-2 main protease. Science 368, 1331–1335 (2020).
43. Sacco, M. D. *et al.* Structure and inhibition of the sars-cov-2 main protease reveal strategy for developing dual inhibitors against m^{pro} and cathepsin l. *Sci. Adv.* **6**, (2020).
44. Jin, Z. et al. Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. Nature 582, 289–293 (2020).
45. Kneller, D. W., Phillips, G., Kovalevsky, A. & Coates, L. Room-temperature neutron and X-ray data collection of 3CL M^{pro} from SARS-CoV-2. Acta Crystallographica Section F 76, 483–487 (2020).
46. Komatsu, T. S. *et al.* Drug binding dynamics of the dimeric SARS-CoV-2 main protease, determined by molecular dynamics simulation. *Sci. Rep.* **10**, https://doi.org/10.1038/s41598-020-74099-5 (2020).
47. Shitrit, A. *et al.* Conserved interactions required for inhibition of the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Scientific Reports* **10**, https://doi.org/10.1038/s41598-020-77794-5 (2020).
48. Suárez, D. & Díaz, N. Sars-cov-2 main protease: a molecular dynamics study. *JCIM* **0**, null (2020).
49. Gasteiger, E. et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research 31, 3784–3788, 2003, https://doi.org/10.1093/nar/gkg563.
50. Kell, G. S. Density, thermal expansivity, and compressibility of liquid water from 0° C to 150° C. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data* **20**, 97–105 (1975).
51. Schrödinger Release 2017, N. Y., LLC. LigPrep (2017).
52. Schrödinger Suite 2017-2, N. Y., LLC. Epik, Protein Preparation Wizard (2017).
53. Banks, J. L. et al. Integrated Modeling Program, Applied Chemical à eory (IMPACT). Journal of Computational Chemistry 26, 1752–1780 (2005).
54. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. Nature Structural Biology 10, 980–980 (2003).
55. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **27**, 221–234 (2013).
56. Friesner, R. A. et al. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. Journal of Medicinal Chemistry 49, 6177–6196 (2006).
57. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–1759 (2004).
58. Sherman, W., Beard, H. S. & Farid, R. Use of an Induced Fit Receptor Structure in Virtual Screening. Chemical Biology Drug Design 67, 83–84 (2006).
59. Zhong, H., Tran, L. M. & Stang, J. L. Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. Journal of Molecular Graphics and Modelling 28, 336–346 (2009).
60. Wang, H., Aslanian, R. & Madison, V. S. Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17α pocket flexibility. Journal of Molecular Graphics and Modelling 27, 512–521 (2008).
61. Luo, H.-J., Wang, J.-Z., Deng, W.-Q. & Zou, K. Induced-fit docking and binding free energy calculation on furostanol saponins from Tupistra chinensis as epidermal growth factor receptor inhibitors. Medicinal Chemistry Research 22, 4970–4979 (2013).
62. Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins: Struct. Funct. Bioinform.* **55**, 351–367 (2004).
63. Jacobson, M. P., Friesner, R. A., Xiang, Z. & Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. Journal of Molecular Biology 320, 597–608 (2002).
64. Buscemi, S., Pace, A., Palumbo Piccionello, A. & Vivona, N. Synthesis of fluorinated first generation starburst molecules containing a triethanolamine core and 1,2,4-oxadiazoles. *J. Fluorine Chem.* **127**, 1601–1605 (2006).
65. Martorana, A., Palumbo Piccionello, A., Buscemi, S., Giorgi, G. & Pace, A. Synthesis of 4(5)-phenacyl-imidazoles from isoxazole side-chain rearrangements. *Organic Biomol. Chem.* **9**, 491–496 (2011).
66. Pibiri, I. et al. Fluorescent Hg^{2+} sensors: Synthesis and evaluation of a tren-based starburst molecule containing fluorinated 1,2,4-oxadiazoles. European Journal of Organic Chemistry 2010, 4549–4553 (2010).

Acknowledgements

We thank Diamond Light Source for access to beam-line B21 (SM27078) and for the assistance of Charlotte Edwards-Gayle during the SAXS experiments. à is work has been partially funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823780.

Author contributions

P.Mariani, L.Si., A.P. and F.S. conceived and designed the work. L.Si., L.Sa. and N.B. synthetised M^{pro} and performed enzymatic activity assays. A.P., C.P., L.C. and V.L. collected and analysed CD data. A.L. performed *in-silico* studies. F.S., Y.G., P.Mariani and M.G.O. collected and analyzed SAXS data. A.P.P. and P.Marzullo synthetised inhibitors. A.P. and F.S. took overall responsibility for the manuscript. All authors approved the final version of the manuscript.

Competing interests
à e authors declare no competing interests.

Additional information
Supplementary Information à e online version contains supplementary material available at https://doi.org/10.1038/s41598-021-88630-9.

Correspondence and requests for materials should be addressed to F.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCE

Cd163

Andersen M.N.,Etzerodt A., Graversen J.H., Holthof L.C., Moestrup S.K., Hokland M.,Møller H.J.,(2019): STAT3 inhibition specifically in human monocytes and macrophages by CD163-targeted corosolic acid-containing liposomes. *Cancer Immunol Immunother.* **68**, 489-502

Andersen C.B., Torvund-Jensen M., Nielsen M.J., de Oliveira C.L., Hersleth H.P., Andersen N.H., Pedersen J.S., Andersen G.R., Moestrup S.K.(2012): Structure of the haptoglobin–haemoglobin complex. *Nature*, **489**(7416), 456-459.

Aruffo A., Bowen M.A., Patel D.D., Haynes B.F., Starling G.C., Gebe J.A., and Bajorath J. (1997): CD6-ligand interactions: a paradigm for SRCR domain function? *Immunol Today* **18,** 498-504.

Bikker F.J., Ligtenberg A.J., End C., Renner M., Blaich S., Lyer S., Wittig R., van't Hof W., Veerman E.C., Nazmi K., et al. (2004a): Bacteria binding by DMBT1/SAG/gp-340 is confined to the VEVLXXXXW motif in its scavenger receptor cysteine-rich domains. *J Biol Chem* **279,** 47699-4770

Bikker F.J., van der Wal J.E., Ligtenberg A.J., Mollenhauer J., de Blieck-Hogervorst J.M., van der Waal I., Poustka A., and Nieuw Amerongen AV (2004b):Salivary agglutinin/ DMBT1SAG expression is up-regulated in the presence of salivary gland tumors. *J Dent Res* **83**:567-571.

Bowen M.A., Bajorath J., Siadak A.W., Modrell B., Malacko A.R., Marquardt H., Nadler S.G., and Aruffo A. (1996): The amino-terminal immunoglobulin-like domain of activated leukocyte cell adhesion molecule binds specifically to the membrane- proximal scavenger receptor cysteine-rich domain of CD6 with a 1:1 stoichiometry. *J Biol Chem* **271**:17390-17396.

Brown M.S. and Goldstein J.L. (1983) Lipoprotein metabolism in the macrophage: implications for cholesterol deposition in atherosclerosis. *Annu Rev Biochem* **52**: 223-261.

Burkard C., Opriessnig T., Mileham A: J., Stadejek T., Ait-Ali T., Lillico S.G., Whitelaw C.B.A., Archibald A.L. (2018): Pigs Lacking the Scavenger Receptor Cysteine-Rich Domain

5 of CD163 Are Resistant to Porcine Reproductive and Respiratory Syndrome Virus 1 Infection. *Journal of Virology* **92** (16).

Calvert J.G., Slade D.E., Shields S.L., Jolie R., Mannan R.M., Ankenbauer R.G., and Welch S.K. (2007): CD163 expression confers susceptibility to porcine reproductive and respiratory syndrome viruses. *J Virol* **81**:7371-7379.

Chakraborty A.K., de Freitas Sousa J., Espreafico E.M., Pawelek J.M.,(2001):Human monocyte x mouse melanoma fusion hybrids express human gene. *Gene,* **275**, 103-106.

Droste A., Sorg C., and Högger P. (1999): Shedding of CD163, a novel regulatory mechanism for a member of the scavenger receptor cysteine-rich family. *Biochem Biophys Res Commun* **256,** 110-113.

Etzerodt A., Moestrup S.K., (2013): CD163 and inflammation: Biological, diagnostic, and therapeutic aspects. *Antioxid. Redox Signal.*, **18**, 2352-2363.

Etzerodt A., Maniecki M.B., Møller K., Møller H.J., and Moestrup S.K. (2010): Tumor necrosis factor alfa-converting enzyme (TACE/ADAM17) mediates ectodomain shedding of the scavenger receptor CD163. *J Leukoc Biol* **88**:1201-1205.

Fabriek B.O., van Bruggen R., Deng D.M., Ligtenberg A.J., Nazmi K., Schornagel K., Vloet R.P., Dijkstra C.D., and van den Berg T.K. (2009) : The macrophage scavenger receptor CD163 functions as an innate immune sensor for bacteria. *Blood* **113**:887- 892.

Fabriek B.O:, Dijkstra C.D., van den Berg T.K. (2005):The macrophage scavenger receptor CD163. *Immunobiology*, **210**(2-4):153-60.

Freeman M., Ashkenas J., Rees D.J., Kingsley D.M., Copeland N.G., Jenkins N.A., and Krieger M. (1990): An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. *Proc Natl Acad Sci USA* **87,** 8810-8814.

Garby L., Noyes W.D.: Studies on hemoglobin metabolism. I. The kinetic properties of the plasma hemoglobin pool in normal man (1959). *J. Clin. Investig.*, **38**, 1479–1483.

Garby L., Noyes W.D.: Studies on hemoglobin metabolism. II. Pathways of hemoglobin iron metabolism in normal man(1959) *J. Clin. Investig.*, **38**, 1484-1486.

Gough P.J. and Gordon S. (2000): The role of scavenger receptors in the innate immune system. *Microbes Infect* **2**:305-311.

Graversen J.H., Moestrup S.K. (2015): Drug trafficking into macrophages via the endocytotic receptor CD163. *Membranes* (Basel), **5**(2):228-52.

Graversen J.H., Madsen M., Moestrup S.K.(2002): CD163: A signal receptor scavenging haptoglobin-hemoglobin complexes from plasma. *Int. J. Biochem. Cell Biol.,* **34**, 309-314.

Grochot-Przeczek A., Dulak J., Jozkowicz A. (2012): Haem oxygenase-1: Non-canonical roles in physiology and pathology. *Clin. Sci.* (Lond.) **122**, 93-103.

Grønlund J., Vitved L., Lausen M., Skjødt K., and Holmskov U. (2000):Cloning of a novel scavenger receptor cysteine-rich type I transmembrane molecule (M160) expressed by human macrophages. *J Immunol* **165**:6406-6415.

Högger P. and Sorg C. (2001): Soluble CD163 inhibits phorbol ester-induced lymphocyte proliferation. *Biochem Biophys Res Commun* **288**:841-843.

Holmgren A. and Bränden C.I. (1989): Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* **342**:248-251.

Huang H.J., Jones N.H., Strominger J.L., and Herzenberg L.A. (1987): Molecular cloning of Ly-1, a membrane glycoprotein of mouse T lymphocytes and a subset of B cells: molecular homology to its human counterpart Leu-1/T1 (CD5). *Proc Natl Acad Sci USA* **84**:204-208.

Huysentruyt L.C., Mukherjee P., Banerjee D., Shelton L.M., Seyfried T.N.(2008): Metastatic cancer cells with macrophage properties: Evidence from a new murine tumor model. *Int. J. Cancer*, **123**, 73-84.

Kneidl J., Loffler B., Erat M.C., Kalinka J., Peters G., Roth J., Barczyk K.(2012): Soluble CD163 promotes recognition, phagocytosis and killing of Staphylococcus aureus via binding of specific fibronectin peptides. *Cell. Microbiol.*, **14**, 914-936.

Kodelja V., Goerdt S.(1994):Dissection of macrophage differentiation pathways in cutaneous macrophage disorders and in vitro. *Exp. Dermatol.*, **3**, 257-268.

Kristiansen M., Graversen J.H., Jacobsen C., Sonne O., Hoffman H.J., Law S.K., and Moestrup S.K. (2001): Identification of the haemoglobin scavenger receptor. *Nature* **409**:198 –201.

Larizza L., Schirrmacher V., Plugger E. (1984):Acquisition of high metastatic capacity after in vitro fusion of a nonmetastatic tumor line with a bone marrow-derived macrophage. *J. Exp. Med.*, **160**, 1579-1584.

Law S.K., Micklem K.J., Shaw J.M., Zhang X.P., Dong Y., Willis A.C., Mason, D.Y. (1993): A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily. *Eur. J. Immunol.*, **23**, 2320-2325.

Lee Y.J. and Lee C. (2010): Deletion of the cytoplasmic domain of CD163 enhances porcine reproductive and respiratory syndrome virus replication. *Arch Virol* **155**: 1319 -1323.

Li R., Ma H., Jiang L., Qiao S., Zhi Y., Huang M., Deng R., and Zhanga G. , 2018 :The CD163 long-range scavenger receptor cysteine-rich repeat: expression, purification and X-ray crystallographic characterization. *Acta Crystallogr F Struct Biol Commun*. **74**(Pt 5): 322-326.

Madsen M, Møller H.J., Nielsen M.J., Jacobsen C., Graversen J.H., van den Berg T., and Moestrup S.K. (2004) Molecular characterization of the haptoglobinhemoglobin receptor CD163. Ligand binding properties of the scavenger receptor cysteine-rich domain region. *J Biol Chem* **279,** 51561-51567.

Maniecki M.B., Etzerodt A., Ulhoi B.P., Steiniche T., Borre M., Dyrskjot L., Orntoft T.F., Moestrup S.K., Møller H.J. (2012):Tumor-promoting macrophages induce the expression of the macrophage-specific receptor CD163 in malignant cells. *Int. J. Cancer*, **131,** 2320-2331.

Maniecki M.B., Etzerodt A., Moestrup S.K., Møller H.J., Graversen J.H. (2011): Comparative assessment of the recognition of domain-specific CD163 monoclonal antibodies in human monocytes explains wide discrepancy in reported levels of cellular surface CD163 expression. *Immunobiology*, **216**, 882-890.

Maniecki M.B. Møller H.J., Moestrup S.K. Møller B.K. (2006): CD163 positive subsets of blood dendritic cells: The scavenging macrophage receptors CD163 and CD91 are coexpressed on human dendritic cells and monocytes. *Immunobiology*, **211**, 407–417.

Martinez V.G., Moestrup S.K., Holmskov U., Mollenhauer J., Lozano F. (2011): The Conserved Scavenger Receptor Cysteine- Rich Superfamily in Therapy and Diagnosis. *Pharmacol. Rev.*, **63**, 967-1000.

Møller H.J. (2012):Soluble CD163. *Scand. J. Clin. Lab. Investig.*, **72**, 1-13.

Møller HJ, Nielsen MJ, Maniecki MB, Madsen M, and Moestrup SK (2010) Soluble macrophage-derived CD163: a homogenous ectodomain protein with a dissociable haptoglobin-hemoglobin binding. *Immunobiology* **215**:406-412.

Møller H.J., Peterslund N.A., Graversen J.H., and Moestrup S.K. (2002) Identification of the hemoglobin scavenger receptor/CD163 as a natural soluble protein in plasma. *Blood* **99**:378–380.

Munzarova M., Lauerova L., Capkova J. (1992): Are advanced malignant melanoma cells hybrids between melanocytes and macrophages? *Melanoma Res.*, **2**, 127-129.

Nielsen M.J., Andersen C.B., Moestrup S.K.: CD163 binding to haptoglobin-hemoglobin complexes involves a dual-point electrostatic receptor-ligand pairing. (2013), *J. Biol. Chem.*, **288:** 18834-18841.

Nielsen M.J., Møller H.J., and Moestrup S.K. (2010) Hemoglobin and heme scavenger receptors. *Antioxid Redox Signal* **12**:261-273.

Pancione M., Giordano G., Remo A., Febbraro A., Sabatino L., Manfrin E., Ceccarelli M., Colantuoni V.,(2014): Immune escape mechanisms in colorectal cancer pathogenesis and liver metastasis. *J. Immunol. Res.*, 2014, 686879.

Patton J.B., Rowland R.R., Yoo D., Chang K.O. (2009): Modulation of CD163 receptor expression and replication of porcine reproductive and respiratory syndrome virus in porcine macrophages. *Virus Res* **140**:161–171.

Pey P., Pearce R.K., Kalaitzakis M.E., Griffin W.S., Gentleman S.M. (2014): Phenotypic profile of alternative activation marker CD163 is different in Alzheimer's and Parkinson's disease. *Acta Neuropathol. Commun.*, **2,** 21.

Philippidis P., Mason J.C., Evans B.J., Nadra I., Taylor K.M., Haskard D.O., Landis R.C. (2004):Hemoglobin scavenger receptor CD163 mediates interleukin-10 release and heme oxygenase-1 synthesis: Antiinflammatory monocyte-macrophage responses in vitro, in resolving skin blisters in vivo and after cardiopulmonary bypass surgery. *Circ. Res.*, **94**, 119–126.

Pulford K., Micklem K., Law S., Mason D.(1998) : CD163 (M130 antigen) workshop panel report. In Kishimoto T., Kikutani H., von dem Borne A, Goyert S., Miyasaka M., Moretta L., Okumura K., Shaw S., Springer T., Sugamura K., et al., Leukocyte Typing VI: White Cell Differentiation Antigens; (Eds).; *Garland Publishing*: New York, NY, USA, pp. 1089–1091.

Pulford K., Micklem K., McCarthy S., Cordell J., Jones M., Mason D.Y., (1992) :A monocyte/macrophage antigen recognized by the four antibodies GHI/61, Ber-MAC3, Ki-M8 and SM4. *Immunology*, **75**, 588–595.

Resnick D., Chatterton J.E., Schwartz K., Slayter H. and Krieger M. (1996): Structures of class A macrophage scavenger receptors. Electron microscopic study of flexible, multidomain, fibrous proteins and determination of the disulfide bond pattern of the scavenger receptor cysteine-rich domain. *J Biol Chem* , **271**:26924–26930.

Resnick D., Pearson A. and Krieger M. (1994) :The SRCR superfamily: a family reminiscent of the Ig superfamily. *Trends Biochem Sci* **19,** 5-8.

Ritter M., Buechler C., Langmann T., Orso E., Klucken J., and Schmitz G. (1999) The scavenger receptor CD163: regulation, promoter structure and genomic organization. *Pathobiology* **67,** 257-261.

Sæderup K.L., Stødkilde K., Graversen J.H., Dickson C.F., Etzerodt A., Hansen S.K., Fago A., Gell D., Andersen C.B.F., Moestrup S.K.(2016)The Staphylococcus aureus Protein IsdH Inhibits Host Hemoglobin Scavenging to Promote Heme Acquisition by the Pathogen. *J Biol Chem*. **291**(46):23989-23998

Sanchez-Torres C., Gomez-Puertas P., Gomez-del-Moral M., Alonso F., Escribano J.M., Ezquerra A., Dominguez J.(2003) : Expression of porcine CD163 on monocytes/macrophages correlates with permissiveness to African swine fever infection. *Arch. Virol.* **148**, 2307–2323.

Sarrias M.R., Grønlund J., Padilla O., Madsen J., Holmskov U. and Lozano F. (2004a): The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. *Crit Rev Immunol* **24**:1–37.

Schaer C.A., Schoedon G., Imhof A., Kurrer M.O., Schaer D.J. (2006): Constitutive endocytosis of CD163 mediates hemoglobin-heme uptake and determines the

noninflammatory and protective transcriptional response of macrophages to hemoglobin. *Circ. Res.* **99**, 943–950.

Schaer D.J., Boretti F.S., Hongegger A., Poehler, D., Linnscheid P., Staege H., Müller C., Schoedon G., Schaffner A.(2001): Molecular cloning and characterization of the mouse CD163 homologue, a highly glucocorticoid-inducible member of the scavenger receptor cysteine-rich family. *Immunogenetics,* **53**(2):170-177

Schafer D.J., Schaer C.A., Bueller P.W., Boykins R.A., Schoedon G., Alayash A.I., Schaffner A. (2006): CD163 is the macrophage scavenger receptor for native and chemically modified hemoglobins in the absence of haptoglobin. *Blood*, **107**, 373-380.

Shabo I., Olson H., Elkarim R., Sun X.F., Svanvik J.(2014): Macrophage Infiltration in TumorStroma is Related to Tumor Cell Expression of CD163 in Colorectal Cancer., *Cancer Microenviron*, **7**, 61–69.

Shabo I., Svanvik J.(2011): Expression of macrophage antigens by tumor cells. *Adv. Exp. Med. Biol.*, **714**, 141–150.

Shabo,I., Olson H., Sun X.F., Svanvik,J.(2009): Expression of the macrophage antigen CD163 in rectal cancer cells is associated with early local recurrence and reduced survival time *Int. J. Cancer*, **125**, 1826–1831.

Shabo I., Stal O., Olsson H., Dore S., Svanvik J. (2008):Breast cancer expression of CD163, a macrophage scavenger receptor, is related to early distant recurrence and reduced patient survival. *Int. J. Cancer*, **123**, 780–786.

Stover C.M., Schleypen J., Grønlund J., Speicher M.R., Schwaeble W.J., and Holmskov U. (2000): Assignment of CD163B, the gene encoding M160, a novel scavenger receptor, to human chromosome 12p13.3 by in situ hybridization and somatic cell hybrid analysis. *Cytogenet Cell Genet* **90**:246–247.

Tavazzi E., Morrison D., Sullivan P., Morgello S., Fischer T. (2014):Brain inflammation is a common feature of HIV-infected patients without HIV encephalitis or productive brain infection. *Curr. HIV Res.*, **12**, 97–110.

Thomsen J.H., Etzerodt A., Svendsen P., Moestrup S.K.,(2013): The haptoglobin-CD163-heme oxygenase-1 pathway for hemoglobin scavenging. *Oxidative Med. Cell. Longev.*, 2013, 523652.

Van Gorp H., Van Breedam W., Van Doorsselaere J., Delputte P.L., and Nauwynck H.J. (2010) Identification of the CD163 protein domains involved in infection of the porcine reproductive and respiratory syndrome virus. *J Virol* **84**:3101–3105.

Weaver L.K., Hintz-Goldstein K.A., Pioli P.A., Wardwell K., Qureshi N., Vogel S.N., and Guyre P.M. (2006): Pivotal advance: activation of cell surface Toll-like receptors causes shedding of the hemoglobin scavenger receptor CD163. *J Leukoc Biol,* **80**:26 –35.

Wheeler G.L,. Miranda-Saavedra D., and Barton G.J. (2008): Genome analysis of the unicellular green alga Chlamydomonas reinhardtii Indicates an ancient evolutionary origin for key pattern recognition and cell-signaling protein families. *Genetics,* **179**:193–197.

Zwadlo G., Voegeli R., Schulze Osthoff K., Sorg, C. A.(1987): Monoclonal antibody to a novel differentiation antigen on human macrophages associated with the down-regulatory phase of the inflammatory process. *Exp. Cell Biol.*, **55**, 295–304.

MPro

Abian O. et al., (2020): Structural stability of SARS-CoV-2 3CLpro and identification of quercetin as an inhibitor by experiments screening. *International Journal of Biological Macromolecules* **164**, 1693 -1703.

Anand K., Ziebuhr J.,Wadhwani P., Mesters J.R. , Hilgenfeld R., (2003): Coronavirus main protease (3CLpro) structure: basis for design of anti-SARS drugs, *Science*, **300** (5626) 1763-1767.

Anand K. et al. (2002): Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *Embo Journal* **21**, 3213-3224

Apenten R., Khokhar S. & Galani D. (2002): Stability parameters for b-lactoglobulin thermal dissociation and unfolding in phosphate buffer at pH 7.0. *Food Hydrocolloids* **16**, 95-103

Bailey O.T., Pappenheimer A.M., Cheever F.S., Daniels J.B.(1949):Murine Virus (JHM)Causing Disseminated Encephalomyelitis with Extensive Destruction of Myelin: II. Pathology. *J. Exp. Med.*, **90**,(3), 195-212.

Banks J.L. et al. (2005): Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of Computational Chemistry*. **26**(16), 1752-1780

Berman H., Henrick K. & Nakamura H. (2003): Announcing the worldwide Protein Data Bank. *Nature Structural & Molecular Biology*, **10**, 980.

Buscemi S., Pace A., Palumbo Piccionello A. & Vivona N. (2006): Synthesis of fluorinated first generation starburst molecules containing a triethanolamine core and 1, 2, 4-oxadiazoles. *Journal of Fluorine Chemistry* **127**, 1601-1605.

Cascella M., Rajnik M., Cuomo A., Dulebohn S,C, Di Napoli R. (2020): Features, evaluation and treatment coronavirus(COVID-19). In: *Stat Pearls Treasure Island* (FL): StatPearls Publishing.

Cheever F.S., Daniels J.B., Pappenheimer A.M., Bailey O.T. (1949): A Murine Virus (JHM) Causing Disseminated Encephalomyelitis with Extensive Destruction of Myelin: I. Isolation and Biological Properties of the Virus. *J. Exp. Med.* **90**, (3)181-194.

Chen S. et al. (2005): Severe acute respiratory syndrome coronavirus 3C-like proteinase n terminus is indispensable for proteolytic activity but not for enzyme dimerization - Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *Journal of Biological Chemistry* **280**, 164-173.

Chou C. et al. (2004): Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease. *Biochemistry* **43**, 14958-14970

Dai W. et al.(2020): Structure-based design of antiviral drug candidates targeting the sars-cov-2 main protease. *Science* **368,** 1331-1335.

de Wit E., van Doremalen N., Falzarano D., Munster V. J. (2016): SARS and MERS: Recent Insights into Emerging Coronaviruses. *Nat. Rev. Microbiol.* **14**, (8), 523-534.

Fan K., Wei P., Feng Q., Chen S., Huang C., Ma L., Lai B., Pei J., Liu Y., Chen J., and Lai L. (2004) :Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase .*J Biol Chem*, **279**(3),1637-1642.

Friesner R.A. et al. (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry* **49**, 6177-6196 .

Friesner R. A. et al. (2004): Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **47**, 1739-1749.

Gorbalenya A.E., Baker S.C., Baric R.S., de Groot R.J., Drosten C., Gulyaeva A.A., Haagmans B.L., Lauber C., Leontovich A.M., Neuman B.W., Penzar D., Perlman S., Poon L.L.M., Samborskiy D., Sidorov I.A., Sola I., Ziebuhr J. (2020): Severe Acute Respiratory Syndrome-Related Coronavirus: The Species and its Viruses–A Statement of the Coronavirus Study Group. *Nat. Microbiol.* **5**, 536–544.

Goyal B., & Goyal D. (2020). Targeting the dimerization of the main protease of coronaviruses: A potential broad-spectrum therapeutic strategy. *ACS Combinatorial Science*, **22**(6), 297-305.

Graziano V., McGrath W. J., Yang L. & Mangel W.F. (2006): SARS CoV Main Proteinase: The Monomer-Dimer Equilibrium Dissociation Constant. *Biochemistry* **45**, 14632-14641.

Greenfield N. J. & Fasman G. D.(1969): Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry,* **8,** 4108-4116.

Greenfield N. J. (2006): Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nature Protocols* **1**, 2527-2535

Halgren T. A. et al. (2004): Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in DatabaseScreening. *Journal of Medicinal Chemistry* **47**(7), 1750-175947.

Hegyi A. and Ziebuhr J. (2002) :Conservation of substrate specificities among coronavirus main proteases. *Journal of General Virology*, **83**, 595-599.

Horton N. & Lewis M. (1992): Calculation of the free energy of association for protein complexes. *Protein Science* **1,** 169-181.

Jacobson M. P. et al. (2004): A hierarchical approach to all-atom protein loop prediction. *Proteins,* **55**, 351-367.

Jacobson M. P., Friesner R. A., Xiang Z. & Honig B. (2002): On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *Journal of Molecular Biology* **320**, 597-608.

Janin, J. (1995): Elusive affinities . *Proteins:* **21**, 30-39

Jin Z. et al. (2020):Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289-293.

Kell G.S. (1975):Density, thermal expansivity, and compressibility of liquid water from 0 C to 150 C. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data* **20**, 97-105.

Kneller D.W., Phillips G., Kovalevsky A. & Coates L. (2020): Room-temperature neutron and X-ray data collection of 3CL Mpro from SARS-CoV-2. *Acta Crystallographica* Section F **76**, 483-487.

Lee J. et al. (2020): Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. *Nature Communications* **11** 5877.

Luo H.-J., Wang J.-Z., Deng W.-Q. & Zou K. (2013): Induced-fit docking and binding free energy calculation on furostanol saponins from Tupistra chinensis as epidermal growth factor receptor inhibitors. *Medicinal Chemistry Research*, **22**, 4970-4979.

Madhavi Sastry G., Adzhigirey M., Day T., Annabhimoju R. & Sherman W. (2013): Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design,* **27**, 221-234.

Martorana A., Gentile C. & Laura A. (2020) In Silico Insights into the SARS CoV-2 Main Protease Suggest NADH Endogenous Defences in the Control of the Pandemic Coronavirus Infection. **Viruses 12**, 805.

Martorana A., Palumbo.Piccionello A., Buscemi S., Giorgi G. & Pace A. (2011):Synthesis of 4(5)-phenacyl-imidazoles from isoxazole side-chain rearrangements. *Organic and Biomolecular Chemistry* **9**, 491-496.

McKee D.L., Sternberg A., Stange U., Laufer S., Naujokat C. (2020):Candidate drugs against SARS-CoV-2 and COVID-19. *Pharmacol Res* **157**:104859.

Morse J. S., Lalonde T., Xu S. & Liu W. R. (2020): Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. *ChemBioChem*, **21**, 730–738

Ortore M.G. et al. (2009): Combining structure and dynamics: non-denaturing high-pressure effect on lysozyme in solution. *J. R. Soc. Interface* v 6(5), S619–S634.

Pibiri I. et al. (2010): Fluorescent Hg2+ sensors: Synthesis and evaluation of a tren-based starburst molecule containing fluorinated 1,2,4-oxadiazoles. *European Journal of Organic Chemistry* **20**(24), 4549-4553.

Pillaiyar T., Manickam M., Namasivayam V., Hayashi Y. & Jung S.- H. (2016): An Overview of Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy..*Journal of Medicinal Chemistry* **59,** 6595-6628

Reiner Ž, Hatamipour M., Banach M., Pirro M., Al-Rasadi K.,Jamialahmadi T., Radenkovic D., Montecucco F., Sahebkar A.(2020): Statins and the Covid-19 main protease: in silico evidence on direct interaction. *Arch Med Sci* **16** (3): 490-496.

Sacco M.D. et al. (2020): Structure and inhibition of the sars-cov-2 main protease reveal strategy for developing dual inhibitors against mpro and cathepsin l. *Science Advances* **6**.

Sherman W., Beard H. S. & Farid R. (2006): Use of an Induced Fit Receptor Structure in Virtual Screening. *Chemical Biology Drug Design,* **67**, 83–84.

Sherman W., Day T., Jacobson M. P., Friesner R. A. & Farid R. (2006): Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. J*ournal of Medicinal Chemistry* **49**, 534-553

Shi J. & Song J.(2006): The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. *FEBS Journal ,* **273**(5), 1035-1045

Shi J., Wei Z. & Song J. (2004): Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme - Defining the extra domain as a new target for design of highly specific protease inhibitors. *Journal of Biological Chemistry,* **279**, 24765-24773.

Silvestrini L., Belhaj N., Comez L., Gerelli Y., Lauria A., Libera V., Mariani P., Marzullo P., Ortore M.G., Palumbo Piccionello A., Petrillo C., Savini L., Paciaroni A. & Spinozzi F. (2021):The dimer-monomer equilibrium of SARS-CoV-2 main protease is affected by small molecule inhibitors. *Nature*, *Scientific Reports volume* **11**, Article number: 9283

Sinibaldi R. et al. (2008): Sans/saxs study of the bsa solvation properties in aqueous urea solutions via a global fit approach. *Eur.Biophys. J.* **37**, 673–681.

Sinibaldi R. et al. (2007): Preferential hydration of lysozyme in water/glycerol mixtures: A small-angle neutron scattering study. *J. Chem. Phys.* **126**, 235101-235109.

Suárez D. & Díaz N. (2020): Sars-cov-2 main protease: A molecular dynamics study. *J Chem Inf Model*, **60**(12):5815-5831.

Svergun, D. et al. (1998): Protein hydration in solution: experimental observation by X-ray, neutron scattering. *Proc. Natl. Acad. Sci.* USA **95**, 2267-2272.

Teixeira J. (1988): Small-angle scattering by fractal systems. *J. Appl. Cryst.* **21,** 781-785.

Wang C., Horby P. W., Hayden F. G. & Gao G. F. (2020): A novel coronavirus outbreak of global health concern. *Lancet,* **395,** 470- 473.

Wang H., Aslanian R. & Madison V. S. (2008): Induced-fit docking of mometasone furoate and further evidence for glucocorticoid receptor 17a pocket flexibility. *Journal of Molecular Graphics and Modelling* **27**, 512-521.

Wei P. et al. (2006): The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase. *Biochemical and Biophysical Research Communications* **339**, 865-872.

Wu C.G., Cheng S-C, Chen S-C, Li J-Y, Fang Y-H, Chen Y-H and Chou C-Y.(2013): Mechanism for controlling the monomer–dimer conversion of SARS coronavirus main protease. *Acta Crystallogr D Biol Crystallogr* **69** (Pt 5):747-755.

Wu F., Zhao S., Yu B., Chen Y.-M., Wang W., Song Z.-G., Hu Y., Tao Z.-W., Tian J.-H., Pei Y.-Y., Yuan M.-L., Zhang Y.-L., Dai F.-H., Liu Y., Wang Q.-M., Zheng J.-J., Xu L., Holmes E.C., Zhang Y.-Z.(2020): A New Coronavirus Associated with Human Respiratory Disease in China. *Nature,* **579,** 265-269.

Xue X., Yang H., Shen W., Zhao Q., Li J., Yang K., Chen C., Jin Y., Bartlam M. and Rao Z. (2007):Production of Authentic SARS-CoV Mpro with Enhanced Activity: Application as a Novel Tag-cleavage Endopeptidase for Protein Overproduction, *J. Mol. Biol.* **366**(3), 965–975.

Yang H. et al. (2005): Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.,* **3**(10)**,** e324.

Yang H., Yang M., Ding Y., Liu Y., Lou Z., Zhou Z., Sun L., Mo L., Ye S., Pang H., Gao G.F., Anand K., Bartlam M., Hilgenfeld R., and Zihe Rao(2003):The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci.* U. S. A., **100**, 13190 -13195.

Zhang L. et al. (2020): Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors. Science **368**, 409-412

Zhang D. et al. (2010): Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. *Infection genetics and evolution* **10**, 84-887.

Zhong H., Tran L. M. & Stang J. L. (2009): Induced-fit docking studies of the active and inactive states of protein tyrosine kinases. *Journal of Molecular Graphics and Modelling* **28**, 336-346.

Zhou P., Yang X.-L., Wang X.-G., Hu B., Zhang L., Zhang W., Si H.-R., Zhu Y., Li B., Huang C.-L., Chen H.-D., Chen J., Luo Y., Guo H., Jiang R.-D., Liu M.-Q., Chen Y., Shen X.-R.,Wang X., Zheng X.-S., Zhao K., Chen Q.-J., Deng F., Liu L.-L., Yan B., Zhan F.-X., Wang Y.-Y., Xiao G.-F., She Z.-L. (2020): A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin. *Nature*, **579**, 270-273.

Ziebuhr J., Heusipp G., and Siddell S.G.(1997):Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J Virol*. **71**(5): 3992-3997.

Zumla A., Chan J. F., Azhar E. I., Hui D. S., Yuen K.-Y. (2016): Coronaviruses-Drug Discovery and Therapeutic Options. *Nat. Rev. Drug Discovery* , **15**(5), 327–347.