

# Hypothesis Article: Food Proteins are a Potential Resource for Mining Cathepsin L Inhibitory Drugs to Combat SARS-CoV-2

Ashkan Madadlou<sup>1</sup>

<sup>1</sup>Wageningen Universiteit en Research

July 20, 2020

## Abstract

The entry of SARS-CoV-2 into host cells proceeds by a two-step proteolysis process, which involves the lysosomal peptidase cathepsin L. Inhibition of cathepsin L is therefore considered an effective method to prevent the virus internalization. Analysis from the perspective of structure-functionality elucidates that cathepsin L inhibitory proteins/peptides found in food share specific features: multiple disulfide crosslinks (buried in protein core), lack or low contents of  $\alpha$ -helix structures (small helices), and high surface hydrophobicity. Lactoferrin can inhibit cathepsin L, but not cathepsins B and H. This selective inhibition might be useful in fine targeting of cathepsin L. Molecular docking indicated that only the carboxyl-terminal lobe of lactoferrin interacts with cathepsin L and that the active site cleft of cathepsin L is heavily superposed by lactoferrin. Food protein-derived peptides might also show cathepsin L inhibitory activity.

## Abstract

The entry of SARS-CoV-2 into host cells proceeds by a two-step proteolysis process, which involves the lysosomal peptidase cathepsin L. Inhibition of cathepsin L is therefore considered an effective method to prevent the virus internalization. Analysis from the perspective of structure-functionality elucidates that cathepsin L inhibitory proteins/peptides found in food share specific features: multiple disulfide crosslinks (buried in protein core), lack or low contents of  $\alpha$ -helix structures (small helices), and high surface hydrophobicity.

Lactoferrin can inhibit cathepsin L, but not cathepsins B and H. This selective inhibition might be useful in fine targeting of cathepsin L. Molecular docking indicated that only the carboxyl-terminal lobe of lactoferrin interacts with cathepsin L and that the active site cleft of cathepsin L is heavily superposed by lactoferrin. Food protein-derived peptides might also show cathepsin L inhibitory activity.

The ongoing coronavirus disease 2019 (COVID-19) is caused by Severe Acute Respiratory Syndrome Corona Virus-2 (SARS-CoV-2). Coronavirus binds through its spike glycoprotein (S) to angiotensin converting enzyme 2 (hACE2) on the host cell membrane for endocytosis. The S protein is trimeric and each monomer contains two subunits, S1 which mediates the attachment, and S2 which enables membrane fusion. The fusion potential of S protein is activated by a two-step proteolysis process: priming cleavage between S1 and S2 (amino acids 682-685, RRAR), and activating cleavage on S2' site. In the infection of some coronaviruses, cathepsin B is involved; however, an exploration on a SARS-CoV-2 S protein pseudovirus system indicated that cathepsin L, not B, is critical for SARS-CoV-2 S protein activation, similar to SARS-CoV and human immunodeficiency virus (HIV) (Ou et al., 2020). After endocytosis of SARS-CoV, cathepsin L cleaves S to S1 and S2. The cleavage enables fusion of the viral membrane, which was attached to hACE2, with the endosomal membrane (Fig. 1A). Then the viral genome is released into the host cell for replication (Adedeji et al., 2013).

Cathepsin L (220 amino acid residues) (Fujishima et al., 1997) is a lysosomal cysteine peptidase and has a two-chain form (L and R). The L domain contains three  $\alpha$ -helices (one is the longest central helix) and the R

domain is a  $\beta$ -barrel, which is closed at the bottom by an  $\alpha$ -helix (Fig. 1B). The reactive site-cleft composes of histidine (His163) located at the top of the barrel and cysteine (Cys25) located at the N-terminus of the L domain central helix (Turk et al., 2012). Cathepsin L contributes to protein turnover, and apoptosis in cells. When a clinically proven serine protease inhibitor was used to inhibit cellular serine protease TMPRSS2, which is exploited by SARS-CoV-2 for S protein priming, viral infection was incompletely inhibited. Full inhibition was accomplished by the concurrent employment of two drugs, one for inhibition of TMPRSS2 and the other for cathepsin L/B, suggesting the necessity of cathepsin L inhibition for prevention of SARS-CoV-2 membrane fusion (Hoffmann et al., 2020).

The overexpression of cathepsin L and H in cancer cells, in contrast to cathepsins B, C, S, and X/Z has made cathepsin L a target for anticancer strategies (Lankelma et al., 2010). The enzyme is also implicated in other pathologies such as osteoporosis, and periodontal diseases (Fujishima et al., 1997). Consequently, a plethora of cathepsin L inhibitory drugs are present for assessment. Some of the drugs e.g. E-64 (Fujishima et al., 1997) inhibit several cathepsins including cathepsin L, whereas, a few such as CLIK-148 (an epoxysuccinyl derivate) and iCL (an aldehyde derivative) exclusively inhibit cathepsin L (Lankelma et al., 2010). It has been shown that the inhibition of cathepsin L, rather than cathepsin B, for instance by SSAA09E1  $\{[(Z)-1\text{-thiophen-2-ylethylideneamino}]thiourea\}$  holds premises for suppression of SARS-CoV (Adedeji et al., 2013), and SARS-CoV-2 endocytosis. Many of the cathepsin L inhibitory drugs, like dipeptide epoxyketones, which are used for SARS-CoV suppression (Adedeji et al., 2013) and iCL (napsul-Ile-Trp-CHO) (Lankelma et al., 2010) are essentially peptidomimetic (Adedeji et al., 2013).

Given that food contains naturally occurring bioactive peptides and proteins, recently we outlined how food protein-derived peptides (small fragments, usually  $<3$  kDa) that influence the renin-angiotensin system might have implications on SARS-CoV-2 endocytosis and pulmonary function in COVID-19 patients. The presumed mechanisms of action were over-expression of hACE2, and the receptor Mas, as well as blockage of the angiotensin type I receptor on cell surface (Goudarzi et al., 2020). However, food proteins/peptides afford further opportunities for inhibition of SARS-CoV-2. Certain foods contain protein/peptide-based cathepsin L inhibitory compounds. This characteristic confers a new vision to exploration and implementation of biomolecules that can inhibit SARS-CoV-2 entry into host cells.

Whereas some naturally occurring bioactive peptides in food such as carnosine (226 Da,  $\beta$ -alanyl-L-histidine, in meat and chicken) do not affect cathepsin activity (Bonner et al., 1995), several cathepsin inhibitory molecules have been identified in plant and animal foods. Oryzacystatin-I (11.4 kDa, residue count 102) in rice is known to inhibit cathepsin L (inhibition constant,  $K_i : 7.3 \times 10^{-10}$  M), but also cathepsin B ( $K_i : 7.9 \times 10^{-8}$  M) and cathepsin H ( $K_i : 1.0 \times 10^{-6}$ ) (Hellinger and Gruber, 2019). A hydrophobic cluster between the  $\alpha$ -helix and the five-stranded antiparallel  $\beta$ -sheet structures (Fig. 1C) stabilizes the helix architecture of oryzacystatin-I (Nagata et al., 2000) and the N-terminal 21 amino acid residues are most probably not essential for cathepsin inhibitory activity of oryzacystatin-I (Abe et al., 1988). It is known that hydrophobic interactions are crucial for the inhibition of cathepsins, likely associated with the hydrophobic wedge-shaped structure of their active-site cleft (Turk et al., 2012). Corn cystatin-I as well inhibits cathepsin L ( $K_i : 1.7 \times 10^{-8}$  M) and cathepsin H ( $K_i : 5.7 \times 10^{-9}$  M) (Hellinger and Gruber, 2019).

In contrast to cystatin, bromelain inhibitor VI [BI-VI; 5.89 kDa; heavy chain (H, 41 amino acid residues) and light chain (L, 11 amino acid residues)] (Fig. 1D) which is a peptide present in pineapple stem selectively inhibits cathepsin L ( $K_i 0.2 \times 10^{-6}$  M) and at a lesser extent trypsin (Polya, 2003). The primary and secondary structures of BI-VI are remarkably distinct from those of cystatin. Hen egg white cystatin consists of two  $\alpha$ -helices and a five-stranded antiparallel  $\beta$ -sheet, but the secondary structure of BI-VI lacks  $\alpha$ -helix structure. It is composed of two domains A and B, each of which formed by a three-stranded antiparallel  $\beta$ -sheet (Hatano et al., 1995). BI-VI contains ten cysteine residues, and five intra/inter-chain disulfide bonds between Cys3<sup>L</sup>-Cys7<sup>H</sup>, Cys6<sup>L</sup>-Cys39<sup>H</sup>, Cys8<sup>L</sup>-Cys5<sup>H</sup>, Cys14<sup>H</sup>-Cys21<sup>H</sup>, and Cys18<sup>H</sup>-Cys30<sup>H</sup> crosslink the protein. The disulfide bonds form the protein core, which is not common as a protein core is commonly occupied with some bulky hydrophobic side chains. This arrangement is homologous with Bowman-Birk trypsin/chymotrypsin inhibitor from soybean (BBI-I), a typical serine protease inhibitor (Hatano et al.,

1996). The heavily S-S crosslinked double chain conformation might be important for the inhibition selectivity towards cathepsin L. This feature is shared with lactoferrin, which is discussed later in the current communication. It is noteworthy that similar to cathepsin L but not essentially required, trypsin could efficaciously activate SARS-CoV-2 S protein, enabling formation of syncytium (Ou et al., 2020). Therefore, the co-inhibition of trypsin and cathepsin L by BI-VI is advantageous.

Comparable to plant resources, foods of animal origin contain cathepsin L inhibitors. Though the proteins and peptides with cathepsin L inhibitory property in foods of animal origin have been scarcely investigated, some evidences are available in the literature. Mammalian milk has significant contents of cysteine protease inhibitors, such as lactoferrin,  $\beta$ -casein and  $\beta$ -lactoglobulin. The inhibitory effect of  $\beta$ -lactoglobulin B, also at a lower extent that of  $\beta$ -lactoglobulin A, on cathepsins K and L have been observed (Ogawa et al., 2009). *Beta*-lactoglobulin is a member of lipocalin family and binds to hydrophobic molecules. In addition to a single cysteine residue buried in protein and protected by  $\alpha$ -helix,  $\beta$ -lactoglobulin has 2 internal disulfide linkages that stabilize the protein structure. At physiological pH,  $\beta$ -lactoglobulin exists as a supramolecular dimer. The secondary structure of  $\beta$ -lactoglobulin consists of nine antiparallel  $\beta$ -sheets, three helical turns and only one short  $\alpha$ -helix (Ragona et al., 2000). *Beta*-lactoglobulin A has an additional negative charge compared to  $\beta$ -lactoglobulin B. A shared feature between BI-VI and  $\beta$ -lactoglobulin B is the existence of internal disulfide bonds.

For  $\beta$ -casein, an allosteric-type inhibition mechanism is believed to cause cathepsin inhibitory (Sano et al., 2005). *Beta*-casein (23.98 kDa) is a single chain polypeptide and the most hydrophobic casein. Essentially all of  $\beta$ -casein net charge and  $\alpha$ -helix structures are positioned at the N-terminal portion of the molecule (Kumosinski et al., 1993). It lacks cysteine. Considerable hydrophobicity and low contents of  $\alpha$ -helix structure (<20%) are the shared features between  $\beta$ -casein and  $\beta$ -lactoglobulin B. Taking hydrophobicity into account, I speculate that comparable to oryzacystatin-I (Abe et al., 1988), the hydrophilic N-terminal portion of  $\beta$ -casein may not be essential for cathepsin inhibitory activity. Rather, the C-terminal hydrophobic portion which is also poor in  $\alpha$ -helix structure probably causes the inhibition. In fact, inhibition of cathepsin L may occur as a consequence of the interaction between the carbonyl (carboxyl) group of an inhibitor with the electrophilic oxyanion hole of cathepsin L, which consists of side chains of Gln16, Trp189, His163 and the main chain of Cys25 (Fujishima et al., 1997).

Lactoferrin is present at much higher contents in human milk than cow's milk. It strongly inhibits cathepsin L. Cathepsin L inhabitation by lactoferrin takes place at  $10^{-7}$  M, while, a synthetic peptide which targets cathepsin L active site inhibited the enzyme at  $10^{-5}$  M. Notably, lactoferrin does not inhibit cathepsin B and cathepsin H (Sano et al., 2005). This may enable fine targeting of cathepsin L for obstacle SARS-CoV-2 internalization, while avoiding possible jeopardizes to cells. Bovine lactoferrin is a single polypeptide consisting of 689 residues. It is folded into two lobes (N-terminal and C-terminal halves) joined by a small helix (3-turn) (Moore et al., 1997) (Fig. 1E). The crystal structures of lactoferrin and cathepsin L were taken from the RCSB protein data bank. Molecular docking using MDSrv (Tiemann et al., 2017) indicates that the C-lobe of lactoferrin interacts with cathepsin L (Fig. 1F). It also shows that the C-lobe superposes the active-site cleft of cathepsin L (Fig. 1F). The  $\alpha$ -helix content of bovine lactoferrin varies between 7% and 16% depending on the pH (Sreedhara et al., 2010). This protein contains 17 disulfide bonds (Wang et al., 2019) and the interactions between the two lobes are mostly hydrophobic caused by packing of nonpolar surfaces on the lobes (Moore et al., 1997). Low  $\alpha$ -helix content, small helices, and hydrophobic surfaces are the shared features between lactoferrin,  $\beta$ -lactoglobulin and  $\beta$ -casein. It is worthy to note that both lactoferrin and BI-VI, which show high selectivity towards cathepsin L, include several disulfide bonds, positioned in protein core. As mentioned in an earlier place, cathepsin L inhibition proceeds by thiol chemistry. Moreover, both lactoferrin and BI-VI possess a double chain/lobe conformation.

In addition to the naturally present protein and peptides with cathepsin L inhibitory activity in foods, some peptides encrypted in food proteins might be able to inhibit cathepsin L and hence assist COVID-19 prevention. Food proteins release diverse biologically active peptides once they are hydrolyzed by the enzymes present in the gastrointestinal tract, and from microbes. Presently evidences showing cathepsin L

inhibitory by food protein-derived peptides are scarce. Two cathepsin B inhibitory peptides derived from  $\beta$ -casein were identified in a pancreatic digest of casein (Lee and Lee, 2000), as well as peptides from the digestion of  $\beta$ -lactoglobulin B by an endopeptidase could inhibit cathepsin K. It remains to be explored whether food protein-derived peptides can inhibit cathepsin L. Isolation of peptides that exclusively (or much more preferentially) inhibit cathepsin L than the other cathepsins can be highly advantageous.

The structural and biofunctional characteristics of many food proteins are extensively characterized. This is especially true for milk proteins, but also egg white proteins. For example, it is known that hen egg white riboflavin-binding protein has a high degree of crosslinking by nine disulfide bonds. Besides, food proteins are generally recognized as safe (GRAS) and inexpensive. Hundreds of biologically active peptide sequences derived from food proteins are known in the literature. The hydrolysis conditions, separation and purification procedures of these peptides are established. *In silico* examination of the cathepsin L inhibitory property of the known peptide sequences can accelerate the drug discovery.

*The author declares not conflict of interest*

## References

- Abe, K., Emori, Y., Kondo, H., Arai, S., and Suzuki, K. (1988). The NH<sub>2</sub>-terminal 21 Amino Acid Residues Are Not Essential for the Papain-inhibitory Activity of Oryzacystatin, a Member of the Cystatin Superfamily. *J. Biol. Chem.* *263* : 7655–7659.
- Adedeji, A.O., Severson, W., Jonsson, C., Singh, K., Weiss, S.R., and Sarafianos, S.G. (2013). Novel Inhibitors of Severe Acute Respiratory Syndrome Coronavirus Entry That Act by Three Distinct Mechanisms. *J. Virol.* *87* : 8017–8028.
- Bonner, A.B., Swann, M.E., Marway, J.S., Heap, L.C., and Preedy, V.R. (1995). Lysosomal and nonlysosomal protease activities of the brain in response to ethanol feeding. *Alcohol* *12* : 505–509.
- Fujishima, A., Imai, Y., Nomura, T., Fujisawa, Y., Yamamoto, Y., and Sugawara, T. (1997). The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett.* *407* : 47–50.
- Goudarzi, M., Garavand, F., Madadlou, A., and Fogliano, V. (2020). Food protein-derived antihypertensive peptides in the COVID-19 pandemic: friends of foes? *J. Hypertens.* *38* : 1614–1616.
- Hatano, K. -i, Kojima, M., Tanokura, M., and Takahashi, K. (1995). Primary Structure, Sequence-Specific 1H-NMR Assignments and Secondary Structure in Solution of Bromelain Inhibitor VI from Pineapple Stem. *Eur. J. Biochem.* *232* : 335–343.
- Hatano, K.I., Kojima, M., Tanokura, M., and Takahashi, K. (1996). Solution structure of bromelain inhibitor VI from pineapple stem: Structural similarity with Bowman - Birk trypsin/chymotrypsin inhibitor from soybean. *Biochemistry* *35* : 5379–5384.
- Hellinger, R., and Gruber, C.W. (2019). Peptide-based protease inhibitors from plants. *Drug Discov. Today* *24* : 1877–1889.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* *181* : 271-280.e8.
- Kumosinski, T.F., Brown, E.M., and Farrell, H.M. (1993). Three-Dimensional Molecular Modeling of Bovine Caseins: An Energy-Minimized  $\beta$ -Casein Structure. *J. Dairy Sci.* *76* : 931–945.
- Lankelma, J.M., Voorend, D.M., Barwari, T., Koetsveld, J., Spek, A.H. Van der, Porto, A.P.N.A. De, et al. (2010). Cathepsin L, target in cancer treatment? *Life Sci.* *86* : 225–233.
- Lee, H.S., and Lee, K.J. (2000). Cathepsin B inhibitory peptides derived from  $\beta$ -casein. *Peptides* *21* : 807–809.

Moore, S.A., Anderson, B.F., Groom, C.R., Haridas, M., and Baker, E.N. (1997). Three-dimensional structure of diferric bovine lactoferrin at 2.8 Å resolution. *J. Mol. Biol.* *274* : 222–236.

Nagata, K., Kudo, N., Abe, K., Arai, S., and Tanokura, M. (2000). Three-dimensional solution structure of oryzacystatin-I, a cysteine proteinase inhibitor of the rice, *Oryza sativa* L.  *japonica*. *Biochemistry* *39* : 14753–14760.

OGAWA, N., TAKAHASHI, M., ISHIDOH, K., and KATUNUMA, N. (2009). Inhibition of Collagenolytic Cathepsins by  $\beta$ -Lactoglobulin in Milk and Its Suppressive Effect on Bone Resorption. *J. Nutr. Sci. Vitaminol. (Tokyo)*. *55* : 264–270.

Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., et al. (2020). Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat. Commun.* *11* : 1620.

Polya, G.M. (2003). *Protein and Non-Protein Protease Inhibitors from Plants*. pp 567–641.

Ragona, L., Fogolari, F., Zetta, L., Pérez, D.M., Puyol, P., Kruif, K. De, et al. (2000). Bovine  $\beta$ -lactoglobulin: Interaction studies with palmitic acid. *Protein Sci.* *9* :

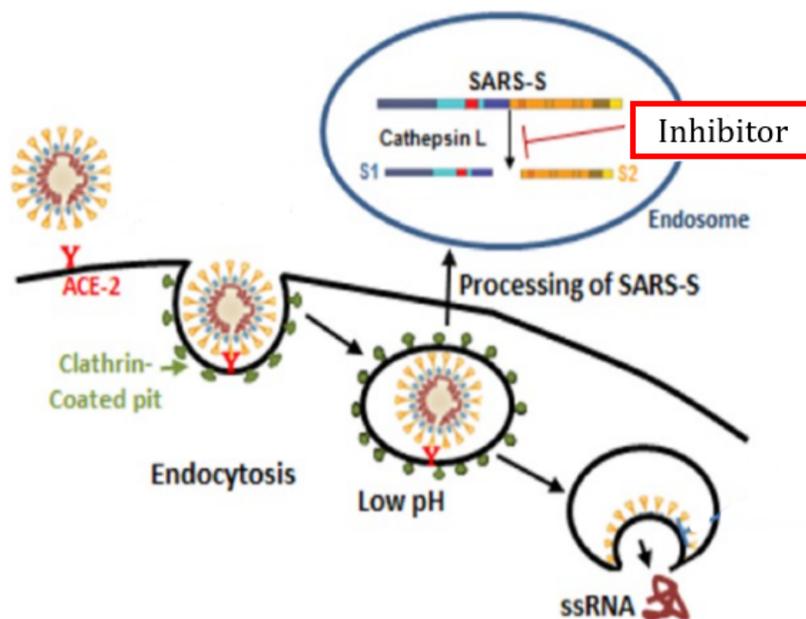
Sano, E., Miyauchi, R., Takakura, N., Yamauchi, K., Murata, E., Le, Q.T., et al. (2005). Cysteine protease inhibitors in various milk preparations and its importance as a food. *Food Res. Int.* *38* : 427–433.

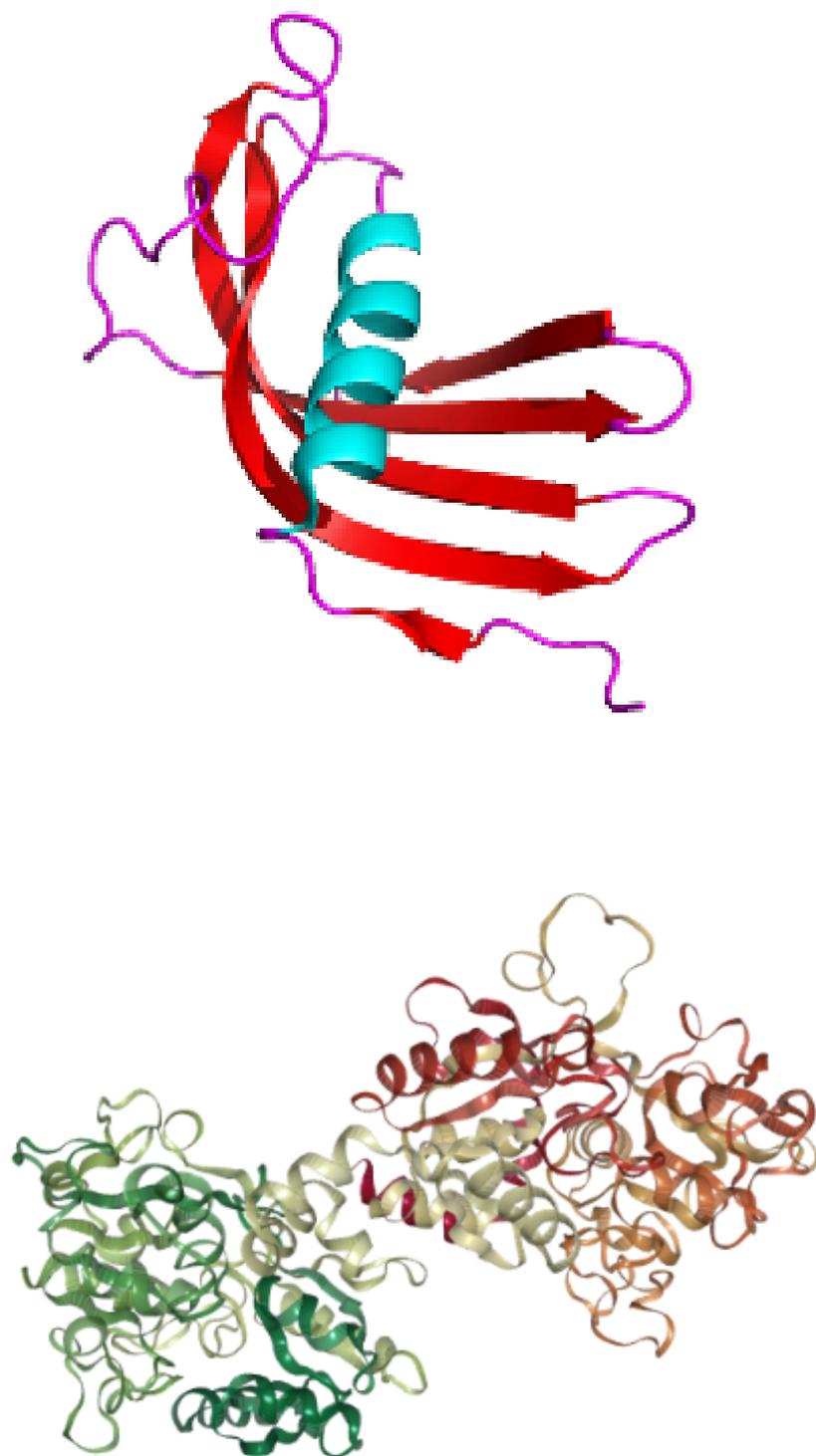
Sreedhara, A., Flengsrud, R., Prakash, V., Krowarsch, D., Langsrud, T., Kaul, P., et al. (2010). A comparison of effects of pH on the thermal stability and conformation of caprine and bovine lactoferrin. *Int. Dairy J.* *20* : 487–494.

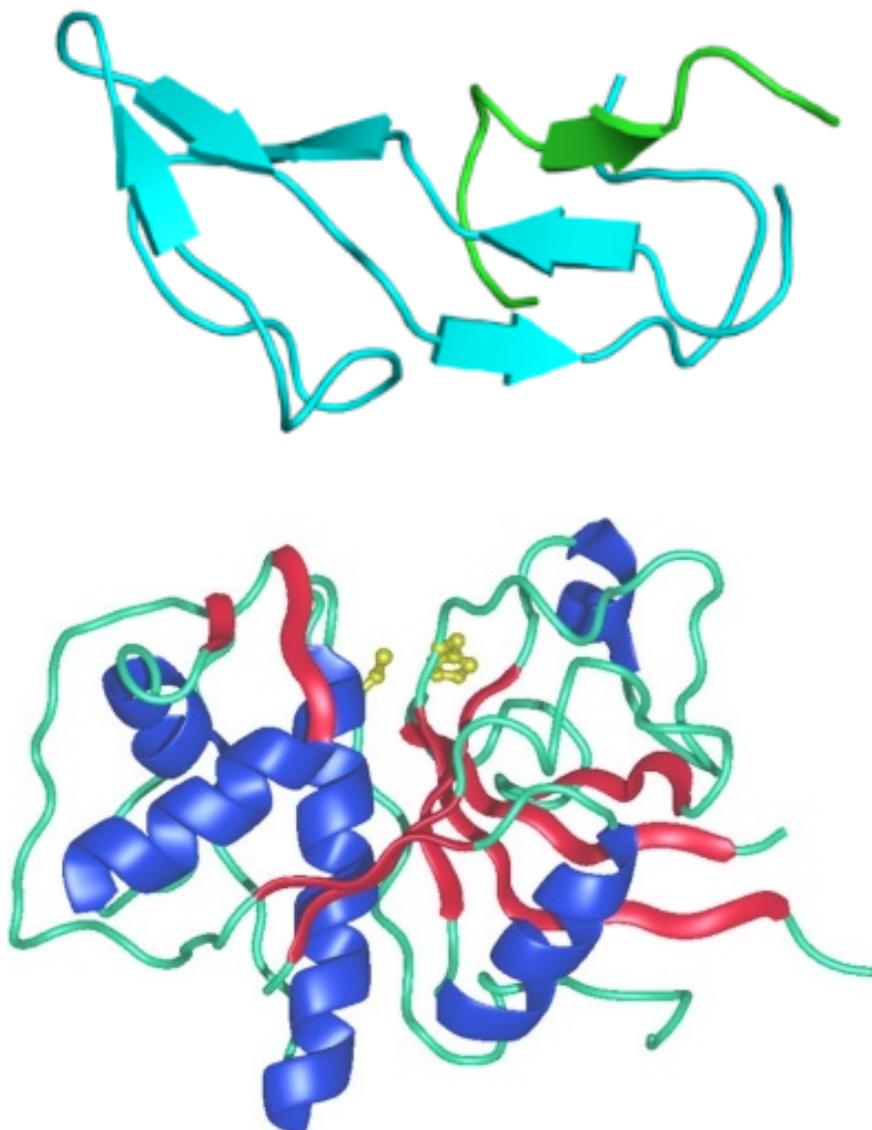
Tiemann, J.K.S., Guixà-González, R., Hildebrand, P.W., and Rose, A.S. (2017). MDsrv: viewing and sharing molecular dynamics simulations on the web. *Nat. Methods* *14* : 1123–1124.

Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., et al. (2012). Cysteine cathepsins: From structure, function and regulation to new frontiers. *Biochim. Biophys. Acta - Proteins Proteomics* *1824* : 68–88.

Wang, B., Timilsena, Y.P., Blanch, E., and Adhikari, B. (2019). Lactoferrin: Structure, function, denaturation and digestion. *Crit. Rev. Food Sci. Nutr.* *59* : 580–596.

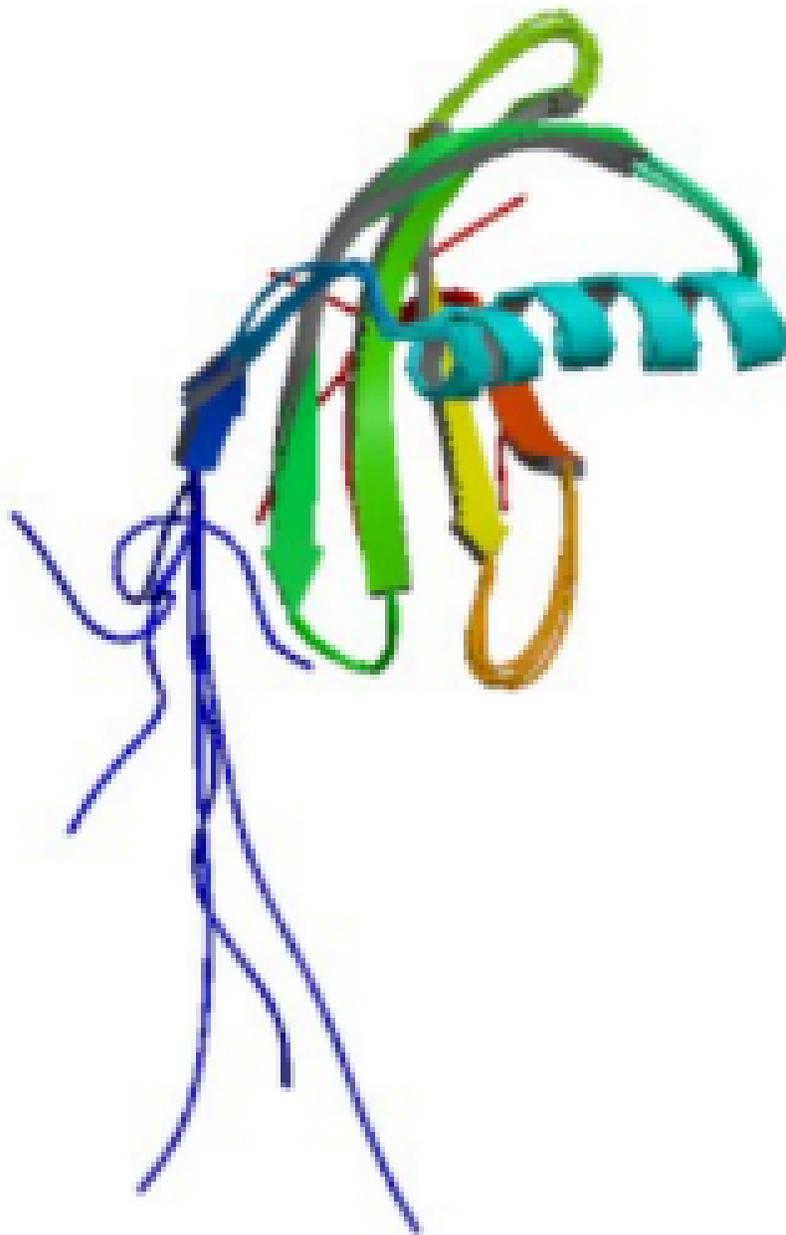






**B C D E**







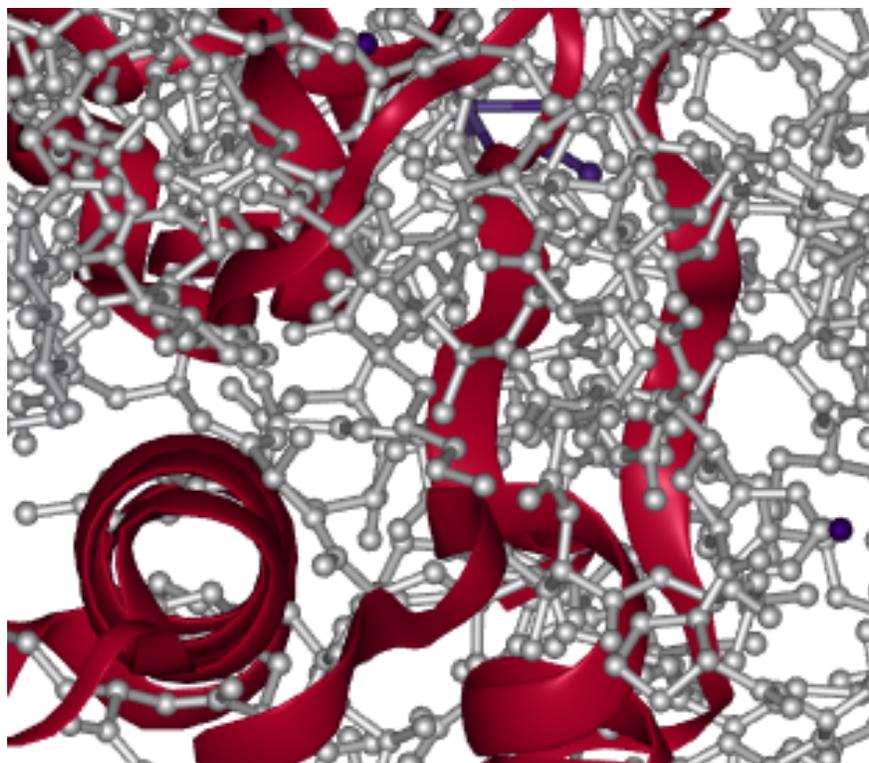


Fig. 1. **A** ) After endocytosis of SARS-CoV, cathepsin L cleaves SARS-S to S1 and S2, allowing subsequent fusion of the viral membrane with the endosomal membrane (adopted from Adedeji et al., 2013);**B**) Ribbon drawing structure of cathepsin L. Green ribbon: the fold of the two-chain form of native cathepsin L; Blue:  $\alpha$ -helices; Red:  $\beta$ -sheets; Ball and stick: the side chains of the reactive-site cysteine (Cys25) and histidine (His163) (Turk et al., 2012); Ribbon drawing structures of **C** ) cystatin and oryzacystatin, **D** ) bromelain inhibitor VI and **E** ) diferric bovine lactoferrin, which show cathepsin L inhibitory activity. Bromelain inhibitor VI and lactoferrin specifically inhibit cathepsin L (no other cathepsins);**F** ) Structural drawings of human cathepsin L superposed by bovine lactoferrin.