Crystal structure of inhibitor-bound human MSPL that can activate high pathogenic avian influenza

Ayako Ohno[1,10], Nobuo Maita[2,10,11], Takanori Tabata[3], Hikaru Nagano[4], Kyohei Arita[5], Mariko Ariyoshi[6], Takayuki Uchida[1], Reiko Nakao[1], Anayt Ulla[1], Kosuke Sugiura[1,7], Koji Kishimoto[8], Shigetada Teshima-Kondo[4], Takeshi Nikawa[1] and Yuushi Okumura[9,12]

1. Department of Nutritional Physiology, Institute of Medical Nutrition, Tokushima University Graduate School, 3-18-15 Kuramoto, Tokushima, Tokushima, 770-8503,
Japan
2. Division of Disease Proteomics, Institute of Advanced Medical Sciences, Tokushima University, 3-18-15 Kuramoto, Tokushima, Tokushima, 770-8503, Japan
3. Laboratory for Pharmacology, Pharmaceutical Research Center, Asahikasei Pharma, 632-1 Mifuku, Izunokuni, Shizuoka 410-2321, Japan
4. Department of Nutrition, Graduate School of Comprehensive Rehabilitation, Osaka Prefecture University, 3-7-30 Habikino, Habikino, Osaka 583-8555, Japan
5. Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro, Tsurumi, Yokohama, 230-0045, Japan
6. Graduate School of Frontier Biosciences, Osaka University, 1-3 Yamadaoka, Suita, Osaka, 565-0871, Japan
7. Department of Orthopedics, Institute of Biomedical Sciences, Tokushima University Graduate School, 3-18-15 Kuramoto, Tokushima, Tokushima, 770-8503, Japan
8. Graduate School of Technology, Industrial and Social Sciences, Tokushima University, 1-1 Minamijyosanjima, Tokushima, Tokushima, 770-8502, Japan
9. Department of Nutrition and Health, Faculty of Nutritional Science, Sagami Women's University, 2-1-1 Bunkyo, Minami, Sagamihara, Kanagawa, 252-0383,
Japan
10. These authors contributed equally.
11. Present address: Institute for Quantum Life Science, National Institute for Quantum and Radiological Science and Technology, 4-9-1, Anagawa, Inage-ku, Chiba 263-8555, Japan
12. Correspondence: okumura_yushi@isc.sagami-wu.ac.jp (Y. O.)

Abstract

Viral infection is triggered when a surface envelope glycoprotein, hemagglutinin (HA), is cleaved by host cell proteins of the transmembrane protease serine (TMPRSS) family. The extracellular region of TMPRSS-2, -3, -4, and MSPL are composed of LDLA, SRCR, and SPD domains. MSPL can cleave the consensus multibasic (R-X-X/R-R) and monobasic (Q(E)-T/X-R) motifs on HA, while TMPRSS2 or -4 only cleaves monobasic motifs. To better understand HA cleavage mediated by MSPL, we solved the crystal structure of the extracellular region of human MSPL in complex with the furin inhibitor. The structure revealed that three domains are gathered around the C-terminal α-helix of the SPD domain. The furin inhibitor structure shows that the side chain of P1-Arg inserts into the highly conserved S1 pocket, whereas the side chain of P2-Lys interacts with the Asp/Glu-rich 99 loop that is unique to MSPL. We also constructed a homology model of TMPRSS2, which is identified as an initiator of SARS-CoV-2 infection. The model suggests that TMPRSS2 is more suitable for Ala/Val residues at the P2 site than Lys/Arg residues.

Key Words: transmembrane protease serine, crystal structure, MSPL, viral infection

Subject Categories: Structural biology

Introduction

Mosaic serine protease large form (MSPL), also known as TMPRSS13, was originally identified from a human lung cDNA library and is a member of the type II transmembrane serine proteases (TTSPs) (1, 2). TTSPs comprise a transmembrane domain near the N-terminus and a catalytic serine protease domain at the C-terminus. Human MSPL is expressed in lung, placenta, pancreas and prostate (1). Little is known about the biological function of human MSPL, although there are reports it can cleave the spike protein of porcine epidemic diarrhea virus (PEDV) (3), MERS- and SARS-CoV (4), influenza virus hemagglutinin (5), and pro-hepatocyte growth factor (6). TTSPs share a similar overall organization comprising an N-terminal cytoplasmic domain, transmembrane region, and stem/catalytic domains at the C-terminus (7). All TTSPs are synthesized as single-chain zymogens and are subsequently activated into the two-chain active forms by cleavage within the highly conserved activation motif. The two chains are linked by a disulfide bridge, so that TTSPs remain bound to the cell membrane (8). The catalytic domain contains a highly conserved 'catalytic triad' of three amino acids (His, Asp, and Ser). The conserved Asp lies on the bottom of the S1 substrate-binding pocket. Substrate specificity results from Arg or Lys residues in the P1 position. Based on similarities in domain structure, the serine protease domain and chromosomal location, TTSPs are classified into four subfamilies: hepsin/TMPRSS, matriptase, HAT/DESC and corin (7, 9). MSPL belongs to the hepsin/TMPRSS subfamily. In this subfamily, hepsin and spinesin contain a single scavenger receptor cysteine-rich repeat (SRCR) domain in the stem region, while MSPL, TMPRSS2, -3, and -4 contains a low-density lipoprotein receptor A (LDLA) domain near the single SRCR domain in the stem region (9). The SRCR domain contains approximately 100–110 amino acids that

adopt a compact fold consisting of a curved β-sheet wrapped around an α-helix and is stabilized by 2-4 disulfide bonds. Depending on the number and the position of the cysteine residues, the SRCR domain has been divided into three subclasses (group A, B and C) (10). However, the canonical LDLA domain contains approximately 40 amino acids and contains six conserved cysteine residues that are involved in the formation of disulfide bonds. The LDLA domain also contains a calcium ion coordinated with six highly conserved residues near the C-terminus. Together, the disulfide bonds and calcium-binding stabilize the overall structure of the LDLA domain (11).

It was recently reported that TMPRSS2, -4, and MSPL are involved in infections by influenza virus by cleaving the glycoprotein hemagglutinin (HA) on the influenza viral surface (4, 5, 12, 13, 14). Specifically, HA is cleaved into HA1 and HA2 subunits by TMPRSS2, -4, and MSPL. Proteolytic cleavage of HA is essential for influenza virus infection, where HA1 mediates host cell binding as well as initiation of endocytosis and HA2 controls viral-endosomal fusion (15). To date, two main HA processing consensus motifs in the influenza virus have been identified. One is the single basic HA processing motif (Q(E)-T/X/-R) in human seasonal influenza viruses, which contain a single arginine at the cleavage site. The other is a multiple-basic-residue motif (R-X-X/R-R and K-K/R-K/T-R) found in highly pathogenic avian influenza viruses, which contain several basic amino acids at the cleavage site. TMPRSS2 and -4 recognize the single basic HA processing motif, while MSPL recognizes both the single basic and multiple basic residue motifs (5). However, it is not clear why only MSPL is able to recognize the multiple-basic-residue motif. The multibasic motif was also known to be recognized by ubiquitously expressed furin and proprotein convertases (PCs)5/6 in the trans-Golgi network (16). A previous study showed that the enzyme activity of MSPL was inhibited

by decanoyl-RVKR-cmk that mimics the substrate for the furin (5). To date, only one structure of the extracellular region of hepsin has been reported among the hepsin/TMPRSS family of proteins (17). The crystal structure of hepsin revealed that the SRCR domain is located at the opposite side of the active site of SPD, and these domains are splayed apart. Because hepsin lacks the LDLA domain, the relative orientation of the LDLA, SRCR and SPD domains in other members of the hepsin/TMPRSS family, such as MSPL, is still unknown. To elucidate the spatial arrangement of the three domains and substrate specificity, we determined the crystal structure of the extracellular region of MSPL in complex with the decanoyl-RVKR-cmk peptide at 2.6 Å resolution. To our surprise, the overall structure of MSPL reveals that the spatial arrangement of the SRCR and SPD domains in MSPL is markedly different from that of hepsin. The complex structure explains how MSPL is able to recognize both the single- and multiple-basic-residue motifs. In addition, we constructed a homology model of TMPRSS2, which is a key protease in SARS-CoV-2 infections. The model was used to investigate the target sequence preference to S1/S2 site of SARS-CoV-2 spike protein.

Results and Discussion

Overall structure of the MSPL extracellular domain

The extracellular region of hMSPL is composed of an LDLA domain (residues 198-221), an SRCR domain (residues 222-313) and a serine protease domain (residues 321-556) (Fig. 1A). We expressed and purified the extracellular region (residues 187-586) of hMSPL and crystallized the protein with decanoyl-RVKR-cmk, which is known as a furin inhibitor. Diffraction data was collected at the Photon Factory AR-NE3a (Tsukuba, Japan) and the structure was solved to a resolution of 2.6 Å (Fig. 1B). This is the first published structure of an LDLA-containing hepsin/TMPRSS subfamily protein. The refined model contains the hMSPL with the residue range of 188-558, except 319 and 320, decanoyl-RVKR-cmk, and a calcium ion. Glycans attached to residues Asn250 and Asn400 were observed, but no phosphorylated residues found (18).

The extracellular region of hMSPL is composed of the non-catalytic portion of the N-terminal region (LDLA domain and SRCR domain) and the catalytic part at the C-terminus (Fig. 1B). The three domains are linked to each other by disulfide bonds. The hMSPL is activated by hydrolytic cleavage at Arg320-Ile321 and residues in the 321-581 region are converted to the mature SPD (5). Ile321 is located in a pocket where the N atom interacts with Asp505 (Fig. S1A). Therefore, this structure could represent a mature form in which hMSPL is processed by an intrinsic protease during expression in the cell. The LDLA domain of hMSPL is 24 residues in length and composed of two turns and a short α-helical region. A canonical LDLA domain has an N-terminal antiparallel β-sheet and three disulfide bonds (11). Therefore, LDLA of MSPL lacks half of the canonical N-terminal region. Since the SRCR domain of MSPL has only two disulfide bonds, it does not belongs to either the group A or B (19). Intriguingly, the 3D

structures of the SRCR domains of MSPL and hepsin are very similar despite their low level of sequence homology (23% sequence identity), suggesting that the SRCR domain of MSPL belongs to group C (10).

To date, hepsin (PDB entry: 1P57) is the only protein in the same TTSP subfamily of proteins with an available 3D structure. However, hepsin lacks the LDLA domain. Here, we compared the structures of hMSPL and hepsin (Fig. 2). The root mean square deviation of the two SPDs (r.m.s.d. of Cα atoms = 0.637 Å), as well as the SRCR domains (r.m.s.d. of Cα atoms = 0.988 Å), are quite small. Although the SPD and SRCR domains of hMSPL and hepsin are almost identical, the arrangement of each domain with respect to each other is markedly different (Fig. 2B). Specifically, when the 3D structures of SPD from hepsin and MSPL are fitted, the SRCR domain of MSPL is rotated by ~80 degrees relative to that of hepsin. The difference may be caused by the presence of the LDLA domain in MSPL. The LDLA domain, SRCR and SPD domains of MSPL are more tightly packed than in hepsin, where these domains are splayed apart. Accordingly, a short parallel β-sheet between the N-terminal segment and the SPD domain was observed in MSPL, whereas the C-terminal end of hepsin forms an antiparallel β-sheet (Fig. 2A).

There are only three residues between the transmembrane domain and the N-terminal Thr188 residue of our structural model. Hence, the extracellular region of MSPL must be located very close to the plasma membrane. Indeed, the region that was predicted to be close to the plasma membrane is enriched in basic residues, such as Arg191, Lys193, Lys 213, Lys 215, and Arg556 (Fig. 2C). The extracellular region of hepsin is also thought to lie flat against the plasma membrane (17). Hence, both MSPL and hepsin may bind substrate in close proximity to the transmembrane region. However, the

extracellular region of MSPL is oriented in the opposite way with respect to that of hepsin.

Interaction of the furin inhibitor (decanoyl-RVKR-cmk) with the MSPL active site

As expected, the SPD of MSPL displays the conserved architecture of the trypsin- and chymotrypsin-like (S1 family) serine proteases (Fig. 1B). In the activated MSPL, Ile321 at the cleavage site forms a salt bridge with the conserved Asp505 residue located immediately prior to the catalytic Ser506 residue (Fig. S1A). This interaction might be generated by the activating cleavage. Formation of the S1 pocket and oxyanion hole comes about via a conformational change in the nearby hairpin loop (Fig. 3). This salt bridge was also observed in other proteases such as plasma kallikrein (20) (PDB entry: 1Z8G) and hepsin. A furin inhibitor peptide binds to the SPD of MSPL with P1-Arg, P2-Lys, C-terminal cmk (chloromethylketone; an active site-direct group) and N-terminal decanoyl group (Fig. 1C, 3). Covalent interaction between the furin inhibitor and catalytic residues (His361, Ser506) occurs via nucleophilic attack on the cmk moiety. P1-Arg inserts into the deep S1 pocket, and its carbonyl oxygen atom directly binds to the backbone amides of the oxyanion hole (Gly504 and Ser506). In addition, the guanidino group of P1-Arg forms a salt bridge with the side chain of Asp500, as well as a hydrogen bond with the side chain of Ser501 and the backbone carbonyl of Gly529. Asp500 is located in the bottom of S1 pocket. These residues are highly conserved among the hepsin/TMPRSS subfamily (Fig. 4). The interaction between P1-Arg and MSPL is characteristic of trypsin- and chymotrypsin-like serine proteases. However, P2-Lys interacts with residues at the so-called 99-loop (chymotrypsinogen numbering) that contains the catalytic residue Asp409. The Nζ of P2-Lys forms five hydrogen bonds with

the backbones of Asp403 and Glu405, the side chains of Tyr401 and Asp406 and a water molecule. This water molecule also mediates hydrogen bonds with the side chains of Asp406 and the catalytic Asp409 residue. Interestingly, with the exception of catalytic Asp409, residues that interact with the side chain of P2-Lys are not conserved among the hepsin/TMPRSS subfamily (Fig. 4, cyan dot). Indeed, this may explain why other TMPRSSs and hepsin recognize the single basic motif but not the di-basic motif. Compared with P1-Arg and P2-Lys, there is no hydrogen bond between the side chains of P3- Val/P4-Lys and the MSPL. The backbone carbonyl of P3-Val forms the hydrogen bond with the backbone amide of Gly527. The side chain of P3-Val makes van der Waals interactions with Trp526 and Gly527. By contrast, the backbone of P4-Arg forms no hydrogen bond with the MSPL and its side chain extends into the bulk solvent. The N-terminal decanoyl moiety makes van der Waals contacts with Thr528 and Gln532 at the 220-loop (chymotrypsinogen numbering). One ordered sulfate ion is located in close proximity to both P3-Val and P4-Arg where it forms hydrogen bonds with the backbone amides of P2 Lys and P3-Val and van der Waals contacts with P3-Val and P4-Arg.

Comparison of the binding mechanisms of furin inhibitor peptide to MSPL and furin

The crystal structure of the furin inhibitor in complex with mouse furin has been determined (21). Although furin also has the same Ser-His-Asp catalytic triad as MSPL, its catalytic domain belongs to the superfamily of subtilisin-like serine proteases (22). The catalytic domain of furin has a different overall fold from that of MSPL, which belongs to the trypsin- and chymotrypsin-like (S1 family) serine protease family. Despite the different overall fold of MSPL and furin, the inhibitor peptide (decanoyl-

RVKR-cmk) can bind to both enzymes. Therefore, we compared the structure of the MSPL-bound furin inhibitor with that of the furin-bound inhibitor (Fig. 5). Except for the P1-Arg, they are not superimposed. In the MSPL:furin inhibitor complex structure, the inhibitor exhibits a bend at the P3-Val. By contrast, in the furin:furin inhibitor complex structure, the inhibitor adopts an extended conformation. As a consequence, the P1, P2, and P4 site contacts with furin, whereas the P3 site is directed away from it. Nonetheless, structural differences between furin and MSPL do not prevent the inhibitor from binding to both proteins.

Homology model analysis of TMPRSS2

In 2020, the SARS-CoV-2 pandemic killed over 0.5 million people (https://ourworldindata.org/covid-deaths) and resulted in a worldwide recession as people were forced to socially distance. In the early stage of infection, the spike protein of SARS-CoV-2 is cleaved by human TMPRSS2, and converted to the infectious form (23, 24). To date, the structure of TMPRSS2 has not been reported. To investigate the structural features of TMPRSS2, we constructed a homology model (Fig. 6) using MSPL as template. Eight out of nine disulfide bonds are conserved (Fig. 4), and the relative domain alignment of TMPRSS2 is similar to that of MSPL. However, the SPD domain, specifically the β12-β13 loop region, displays significant differences (Fig. 6). These structural changes result in a wide substrate-binding groove, so that TMPRSS2 may more readily capture the target peptide. Furthermore, Glu404, an important residue for P2-Lys recognition in MSPL, is replaced by Lys225 in TMPRSS2 (Fig. 4, 6B). As mentioned earlier, this substitution leads to a preference for the monobasic target of TMPRSS2. In fact, the S1/S2 cleavage site of SARS-CoV-2 spike protein is reported as

P2-Ala instead of a basic residue (25, 26, 27). In summary, our homology model reflects the features of TMPRSS2 target peptide recognition.

Implications for autosomal recessive nonsyndromic deafness

Our structure also helps to predict the tertiary structure of TMPRSS3, the gene responsible for autosomal recessive nonsyndromic deafness. Mutations identified in patients with this syndrome were mapped onto a homology model of TMPRSS3 to better understand the disease. Seven missense TMPRSS3 mutants (D103G, R109W, C194F, R216L, W251C, P404L and C407R) associated with deafness in humans were unable to activate the ENaC (28, 29). One of seven missense mutants associated with the loss of hearing, D103G, was found in the LDLA domain of TMPRSS3 (28, 30). Because Asp103 in TMPRSS3 corresponds to Asp221 in MSPL, the LDLA structure stabilized by calcium-binding may be important for the function of the protein. Indeed, the mutations in LDLA and SRCR (D103G, R109W and C194F) as well as the SPD domains of TMPRSS3 affect its autoactivation by proteolytic cleavage at the junction site between the SRCR and the SPD domains (30).

Conclusion

In this study, we have elucidated the structure of the extracellular domain of MSPL and its spatial arrangement of three (LDLA, SRCR, and SPD) domains, as well as the substrate sequence specificity of MSPL. These findings will be useful in designing novel anti-influenza drugs that prevent HAPI virus uptake into the host cell. MSPL also contributes to the cleavage and activation of severe acute respiratory syndrome coronavirus (SARS-CoV) Middle East respiratory syndrome coronavirus (MERS-CoV)

spike proteins (4).

Materials and Methods

Cloning, expression, and purification

Soluble recombinant hMSPL was generated using a previously established stable cell line expressing hMSPL (5), which accumulated in serum-free culture medium (SFCM). Approximately 10 L of SFCM was concentrated by ultrafiltration using a Pellicon XL 50 (Merck-Millipore, Billerica, MA). The resulting SFCM was applied to an Anti-FLAG M2 agarose gel equilibrated in 50 mM Tris-HCl, 150 mM NaCl, pH 7.4 (TBS). Bound protein was subsequently eluted in 0.1 M glycine-HCl, pH 3.5 and fractions containing recombinant hMSPL pooled and dialyzed into phosphate-buffered saline (PBS).

Complex formation, crystallization, and data collection

The peptide inhibitor (decanoyl-RVKR-cmk) was purchased from Merck-Millipore and reconstituted in dimethyl sulfoxide (DMSO). MSPL-inhibitor complex was formed by incubating purified MSPL (6.1 mg/mL) with a 4-fold molar excess of decanoyl-RVKR-cmk at 4 °C for 5 min and then centrifuged (25,000 g) at 4 °C, for 5 min to remove the precipitate. Crystallization screening was performed by mixing 1 µL of the MSPL-inhibitor solution with 1 µL of reservoir solution using the hanging-drop vapor-diffusion method. The MSPL-inhibitor complex was crystallized at 15 °C with a reservoir solution composed of 0.1 M HEPES (pH 7.5), 2.4 M ammonium sulfate. Prior to data collection, the single crystal was transferred to the cryoprotectants [20% of glycerol and 80% of the reservoir] for five seconds, and then flash frozen in liquid nitrogen. The diffraction dataset for the MSPL:furin inhibitor complex crystal was collected at beamline NE3A in the Photon Factory Advanced Ring (PF-AR). The crystal belonged to space group $P2_12_12_1$ with unit cell parameters a = 55.84, b = 62.40, and c =

171.63 Å. Diffraction data were processed using the program iMosflm (31), followed by Aimless (32). Data collection statistics are summarized in Table S1.

Structure determination and refinement of the MSPL-inhibitor peptide complex

The structure of the complex was solved by the molecular replacement method using the program MolRep (33), with SPD of human plasma kallikrein (PDB code: 2ANY), which shows the highest sequence identity score (46.1%), as a search model. The model of SPD was manually fixed with COOT (34) and refined with Refmac5 (35). Once the SPD of MSPL was well refined, interpretable electron density of the unmodeled region was evident. The model of the LDLA and SRCR domains was then manually built. The final model contained one MSPL, one furin inhibitor, four sugars, 80 ions, and 65 water molecules, with *R*-work and *R*-free values of 18.5% and 25.1%, respectively. The refinement statistics are summarized in Table S1. In the MSPL-peptide inhibitor complex, some residues (N-terminal 3xFLAG-tag and His187, Gly319, Arg320, and C-terminal Thr559-Val 581) are missing due to disorder. All the structures in the figures were prepared using PyMOL (http://www.pymol.org/). The MSPL/peptide inhibitor interfaces were analyzed using LIGPLOT (36).

Homology modelling of TMPRSS2

The sequence alignment of the extracellular region of MSPL and TMPRSS2 was obtained using the BLAST webserver (https://www.uniprot.org/blast/). The amino acid identity between MSPL and TMPRSS2 was 39.8% with a score of 704, and E-value of 1.1e-86. The homology model of TMPRSS2 was build using MODELLER (37). Electrostatic surface potentials were calculated using the APBS server

(http://server.poissonboltzmann.org/).

Data availability

The coordinates and structure factors of the MSPL-peptide inhibitor complex have been deposited to the RCSB Protein Data Bank (PDB code: 6KD5).

Author Contributions

The authors have jointly contributed to project design, data analysis, and manuscript preparation. Y. O. performed initial construct design, purification experiments and evaluation of novel inhibitors. A. O. performed the crystallization and design of novel inhibitors. N. M. performed structure solution, model building, and structural analysis. H. N. performed evaluation of novel inhibitors. A. O. and N. M. wrote the manuscript with help from the other co-authors.

Acknowledgments

We thank the beamline staff at the PF-AR and SPring-8 BL44XU for supporting data collection under the proposal number 2013G075 and 20156537, respectively. This work was supported by JSPS KAKENHI grant number 15K09585 (Y. O.), 15J40096 (A. O.), 15K13747 (N. M.), and 19K05696 (N. M.).

Conflict of interest

The authors declare that they have no conflict of interest.

Reference

1. Kim D. R., Sharmin S., Inoue M., Kido H. (2001). Cloning and expression of novel mosaic serine proteases with and without a transmembrane domain from human lung. *Biochim. Biophys. Acta.*, 1518, 204-209.
2. Kido H., Okumura Y., (2008). MSPL/TMPRSS13. *Front. Biosci.*, 13, 754-758.
3. Shi W., Fan W., Bai J., Tang Y., Wang L., Jiang Y., Tang L., Liu M., Cui W., Xu Y., Li Y. (2017). TMPRSS2 and MSPL Facilitate Trypsin-Independent Porcine Epidemic Diarrhea Virus Replication in Vero Cells. *Viruses*, 18, 1-17.
4. Zmora P., Blazejewska P., Moldenhauer A. S., Welsch K., Nehlmeier I., Wu Q., Schneider H., Pöhlmann S., Bertram S. (2014). DESC1 and MSPL activate influenza A viruses and emerging coronaviruses for host cell entry. *J. Virol.*, 88, 12087-12097.
5. Okumura Y., Takahashi E., Yano M., Ohuchi M., Daidoji T., Nakaya T., Böttcher E., Garten W., Klenk H. D., Kido H. (2010). Novel Type II Transmembrane Serine Proteases, MSPL and TMPRSS13, Proteolytically Activate Membrane Fusion Activity of the Hemagglutinin of Highly Pathogenic Avian Influenza Viruses and Induce Their Multicycle Replication. *J. Virol.*, 84, 5089-5096.
6. Hashimoto T., Kato M., Shimomura T., Kitamura N. (2010). TMPRSS13, a type II transmembrane serine protease, is inhibited by hepatocyte growth factor activator inhibitor type 1 and activates pro-hepatocyte growth factor. *FEBS j.*, 277, 4888–4900.
7. Szabo R., Bugge T. H. (2008). Type II transmembrane serine proteases in development and disease. *Int. J. Biochem. Cell Biol.*, 40, 1297-12316.
8. Hooper J. D., Clements J. A., Quigley J. P., Antalis T. M. (2001). Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *J. Biol. Chem.*, 276, 857-860.
9. Szabo R, Bugge T. H. (2011). Membrane-anchored serine proteases in vertebrate cell and developmental biology. *Annu. Rev. Cell Dev. Biol.*, 27, 213-235.
10. Ojala J. R., Pikkarainen T., Tuuttila A., Sandalova T., Tryggvason K. (2007). Crystal structure of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic and an acidic cluster that both Contribute to ligand

recognition. *J. Biol. Chem.*, 282, 16654-16666.

11. Daly N. L., Scanlon M. J., Djordjevic J. T., Kroon P. A., Smith R. (1995). Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. *Proc. Natl. Acad. Sci. U. S. A.*, 192, 6334-6338.
12. Böttcher E, Freuer C, Steinmetzer T, Klenk H. D., Garten W. (2009). MDCK cells that express proteases TMPRSS2 and HAT provide a cell system to propagate influenza viruses in the absence of trypsin and to study cleavage of HA and its inhibition. *Vaccine*, 27, 6324-6329.
13. Chaipan, C., Kobasa, D., Bertram, S., Glowacka, I., Steffen, I., Tsegaye, T. S., Takeda, M., Bugge, T. H., Kim, S., Park, Y., Marzi, A., & Pöhlmann, S. (2009). Proteolytic activation of the 1918 influenza virus hemagglutinin. *J. Virol.*, 83, 3200–3211.
14. Ohler A., Becker-Pauly C. (2012). TMPRSS4 is a type II transmembrane serine protease involved in cancer and viral infections. *Biol. Chem.*, 393, 907-914.
15. Hamilton B. S., Whittaker G. R., Daniel S. (2012). Influenza Virus-Mediated Membrane Fusion: Determinants of Hemagglutinin Fusogenic Activity and Experimental Approaches for Assessing Virus Fusion. *Viruses*, 4, 1144-1168.
16. Stieneke-Gröber, A., Vey M., Angliker H., Shaw E., Thomas G., Roberts C., Klenk H. D., Garten. W. (1992). Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J.*, 11, 2407-2414.
17. Somoza J. R., Ho J. D., Luong C., Ghate M., Sprengeler P. J., Mortara K., Shrader W. D., Sperandio D., Chan H., McGrath M. E., Katz B. A. (2003). The Structure of the Extracellular Region of Human Hepsin Reveals a Serine Protease Domain and a Novel Scavenger Receptor Cysteine-Rich (SRCR) Domain. *Structure*, 11, 1123-1131.
18. Murray A. S., Varela F. A., Hyland T. E., Schoenbeck A. J., White J. M., Tanabe L. M., Todi S. V., List K. (2017). Phosphorylation of the type II transmembrane serine protease, TMPRSS13, in hepatocyte growth factor activator inhibitor-1 and -2-mediated cell-surface localization. *J. Biol. Chem.*, 292, 14867–14884.
19. Sarrias M. R., Grønlund J., Padilla O., Madsen J., Holmskov U., & Lozano, F. (2004). 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.

20. Pathak M., Wong S. S., Dreveny I., Emsley J. (2013). Structure of plasma and tissue kallikreins. *Thromb. Haemost.*, 110, 423-433.
21. Henrich S., Cameron A., Bourenkov G. P., Kiefersauer R., Huber R., Lindberg I., Bode W., Than M. E. (2003). The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat. Struct. Biol.*, 10, 520-526.
22. Siezen R. J., Leunissen J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.*, 6, 501-523.
23. Hoffmann M., Kleine-Weber H., Schroeder S., Krüger N., Herrler T., Erichsen S., Schiergens T. S., Herrler G., Wu N. H., Nitsche A., Müller M. A., Drosten C., Pöhlmann S. (2020a). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181, 271–280.
24. Meng T., Cao H., Zhang H., Kang Z., Xu D., Gong H., Wang J., Li Z., Cui X., Xu H., Wei H., Pan X., Zhu R., Xiao J., Zhou W., Cheng L., Liu J. (2020). The insert sequence in SARS-CoV-2 enhances spike protein cleavage by TMPRSS. Biorxiv, February 2020, Doi: 10.1101/2020.02.08.926006.
25. Walls A. C., Park Y. J., Tortorici M. A., Wall A., McGuire A. T., Veesler D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 181, 281-292.
26. Coutard B., Valle C., de Lamballerie X., Canard B., Seidah N. G., Decroly E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.*, 176, 104742.
27. Hoffmann M., Kleine-Weber H., Pöhlmann S. (2020b). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell*, 78, 779–784.
28. Wattenhofer M., Sahin-Calapoglu N., Andreasen D., Kalay E., Caylan R., Braillard B., Fowler-Jaeger N., Reymond A., Rossier B. C., Karaguzel A., Antonarakis S. E. (2005). A novel TMPRSS3 missense mutation in a DFNB8/10 family prevents proteolytic activation of the protein. *Hum. Genet.*, 117, 528-535.
29. Antalis T. M., Buzza M. S., Hodge K. M., Hooper J. D., Netzel-Arnett S. (2010). The cutting edge: membrane-anchored serine protease activities in the pericellular microenvironment. *Biochem. J.*, 428, 325–346.

30. Guipponi M., Vuagniaux G., Wattenhofer M., Shibuya K., Vazquez M., Dougherty L., Scamuffa N., Guida E., Okui M., Rossier C., Hancock M., Buchet K., Reymond A., Hummler E., Marzella P. L., Kudoh J., Shimizu N., Scott H. S., Stylianos E. Antonarakis S. E., Rossier B. C. (2002). The transmembrane serine protease (TMPRSS3) mutated in deafness DFNB8/10 activates the epithelial sodium channel (ENaC) in vitro. *Hum. Mol. Genet.*, 11, 2829-2836.
31. Battye T. G., Kontogiannis L., Johnson O., Powell H. R., Leslie A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, 67, 271-281.
32. Evans P. R., Murshudov G. N. (2013). How good are my data and what is the resolution? *Acta. Crystallogr. D. Biol. Crystallogr.*, 69, 1204-1214.
33. Vagin A., Teplyakov A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 22-25.
34. Emsley P., Cowtan K., (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, 60, 2126-2132.
35. Murshudov G. N., Skubák P., Lebedev A. A., Pannu N. S., Steiner R. A., Nicholls R. A., Winn M. D., Long F., Vagin A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.*, 67, 355-367.
36. Wallace A. C., Laskowski R. A., Thornton J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, 8, 127-134.
37. Šali R. Blundell T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234, 779-815.
38. Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.*, 22, 4673–4680.
39. Gouet P., Robert X., Courcelle E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucl. Acids Res.*, 31, 3320-3323.

40. Lovell S. C., Davis I. W., Arendall W. B., 3rd, de Bakker P. I., Word J. M., Prisant M. G., Richardson J. S., Richardson D. C. (2003). Structure validation by Cα geometry: φ, ψ and Cβ deviation. *Proteins*, 50, 437–450.

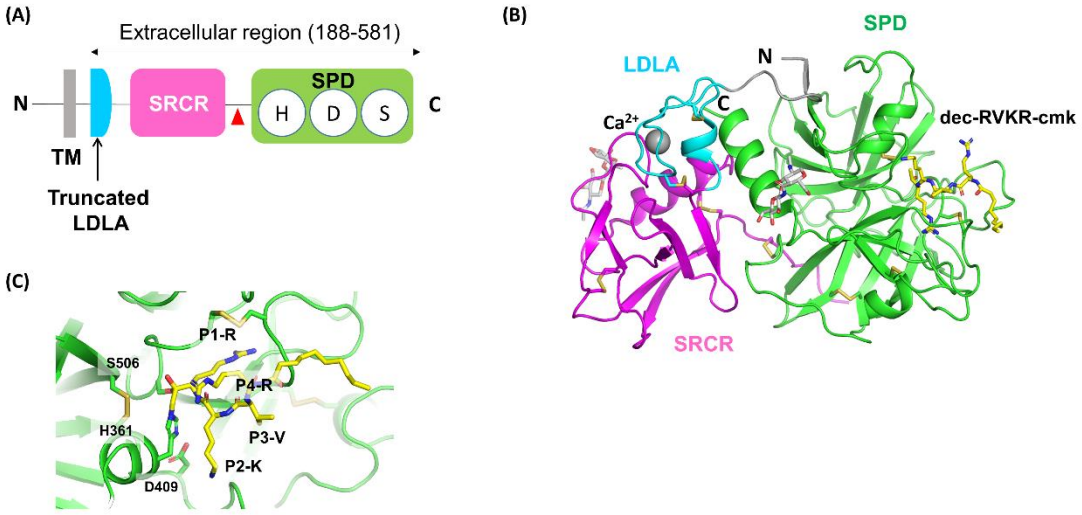

Fig. 1 Overall structure of the MSPL extracellular domain

(A) Schematic illustration of full-length MSPL. MSPL is composed of a cytoplasmic region (1-165), transmembrane helix (166-185), truncated LDL-receptor class A (LDLA) domain (204-220), scavenger receptor cysteine-rich (SRCR) domain (221-314), and serine-protease domain (SPD) (326-558). MSPL is cleaved at Arg321 (red arrowhead) to generate the mature form.

(B) Ribbon representation of the crystal structure of MSPL extracellular region complexed with furin inhibitor (yellow stick model). LDLA domain (cyan), SRCR domain (magenta), and SPD (green) are shown. LDLA domain binds Ca^{2+} in the center of the loop. The N-terminal region (188-191) interacts with SPD by making a β-sheet. Two glycoside chains were observed at Asn250 and Asn400 (white stick model).

(C) A close-up view of bound furin inhibitor and the catalytic triad of residues.

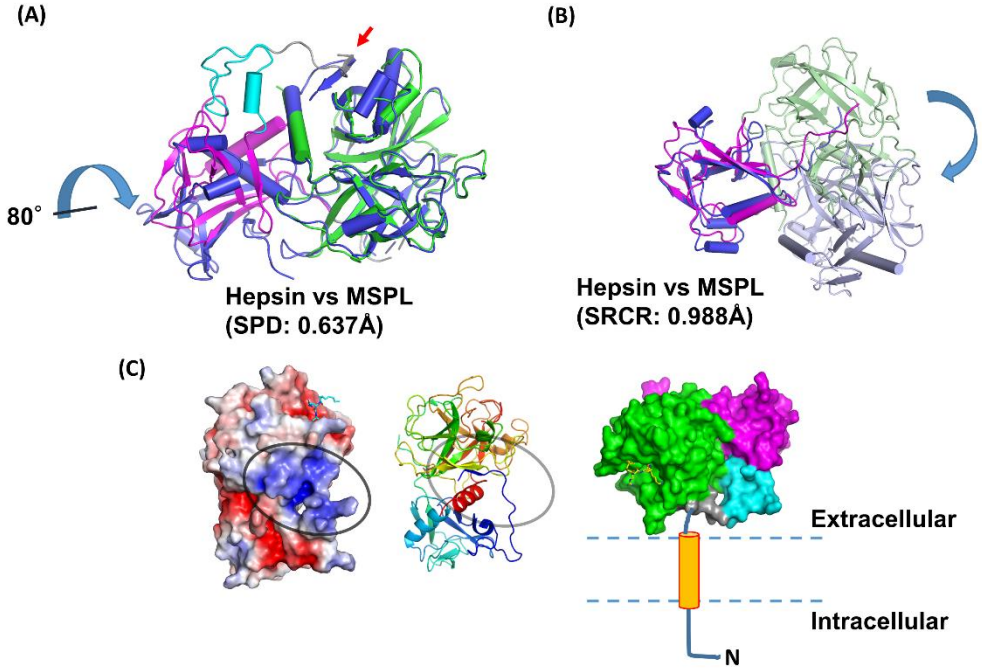

Fig. 2 Comparison of MSPL and hepsin

(A) Hepsin (colored in blue) and MSPL (colored in cyan (LDLA), magenta (SRCR), and green (SPD)) were superposed with the SPD domain. The RMSD value is 0.637 Å calculated with 197 Cα atom positions. A β-sheet interaction of the N-terminus and SPD in MSPL is replaced by the C-terminus in hepsin (red arrow). The hepsin SRCR domain is rotated by about 80° relative to that of MSPL.

(B) Hepsin (colored in blue (SRCR) and pale blue (SPD)) and MSPL (colored in magenta (SRCR) and pale green (SPD)) were superposed with the SRCR domain. The RMSD value is 0.988 Å calculated with 59 Cα atom positions.

(C) (Left) Electrostatic surface potential of the MSPL extracellular domain. A characteristic positively-charged patch (gray oval), composed of Arg191, Lys193, Lys 213, Lys 215, and Arg556, is thought to act as a contact surface for the cell membrane. The potential map is colored from red (-5kT/e) to blue (+5kT/e). (Middle) A ribbon model of MSPL is shown with the same orientation. (Right) A proposed model of the membrane-anchored full-length MSPL.

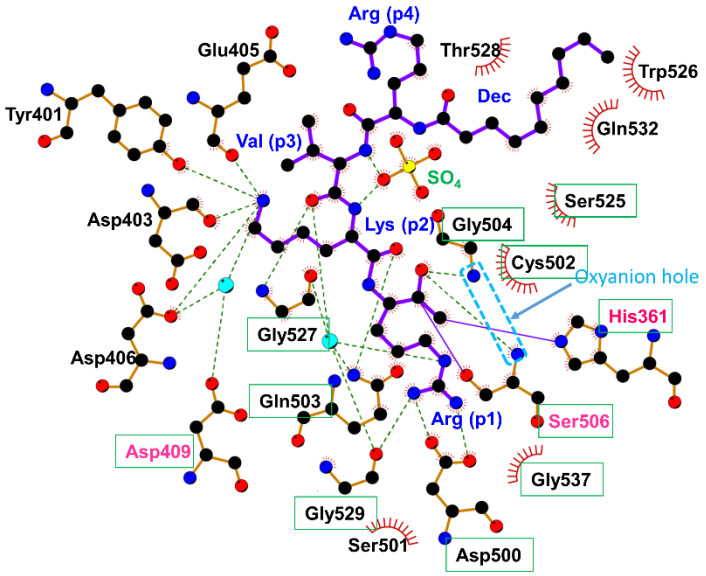

Fig. 3 Interaction of furin inhibitor to MSPL

The SPD of MSPL and furin inhibitor are shown in orange and purple, respectively. Nitrogen atoms, blue; oxygen atoms, red; carbon atoms, black; sulfur atoms, yellow. Dashed lines represent hydrogen bonds. Red semi-circles with radiating spokes denote the residues of the MSPL involved in hydrophobic contacts with furin inhibitor. Cyan spheres denote water molecules. Light-blue dashed rectangle denotes the oxyanion hole. The catalytic triad of three amino acids are highlighted in red. Conserved residues among MSPL, TMPRSS2-4, and hepsin are highlighted in green boxes. The figure was prepared with LIGPLOT (36).

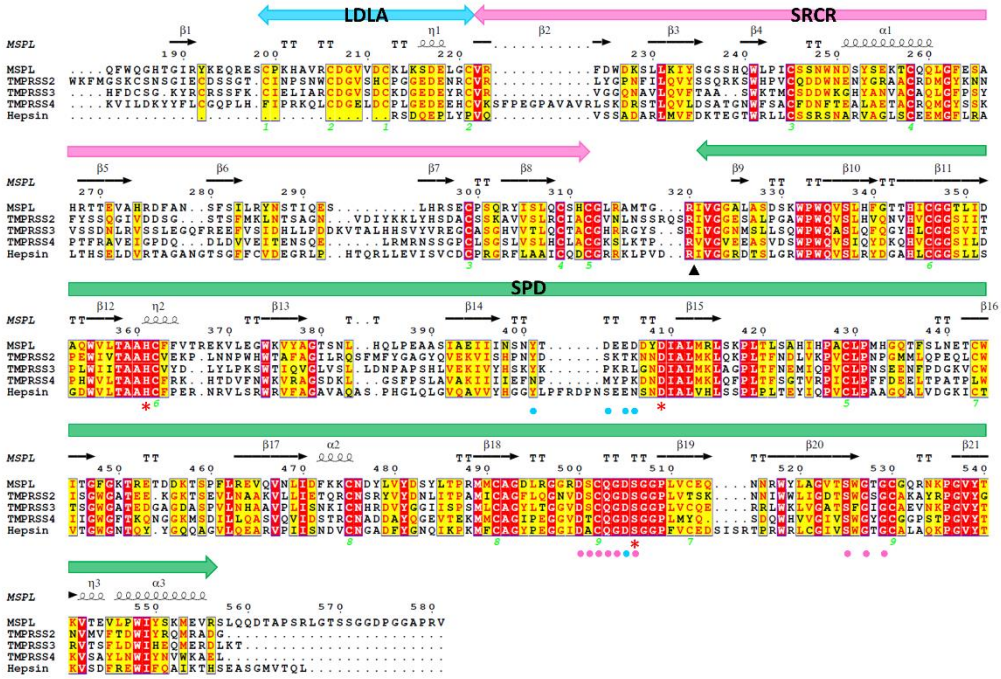

Fig. 4 Multiple sequence alignment of the MSPL extracellular region with members of the hepsin/TMPRSS subfamily

Extracellular region of human MSPL (182-581), human TMPRSS2 (110-492), human TMPRSS3 (70-454), human TMPRSS4 (55-437), and human hepsin (50-417) aligned by Clustal W program (38), followed by coloring with ESPRIPT (39). Red asterisk indicates the catalytic triad. The amino acid sequences were obtained from UniProtKB with the id code of MSPL (Q9BYE2), TMPRSS2 (O15393), TMPRSS3 (P57727), TMPRSS4 (Q9NRS4), and hepsin (P05981). The secondary structure regions identified in MSPL are indicated. Identical residues are shown in white on red, whereas similar residues are shown in red. Black triangle indicates the cleavage site. Pink and cyan circles denote residues that interact with the P1 and P2 site of the furin inhibitor, respectively. Green numbers denote the disulfide pairing of MSPL.

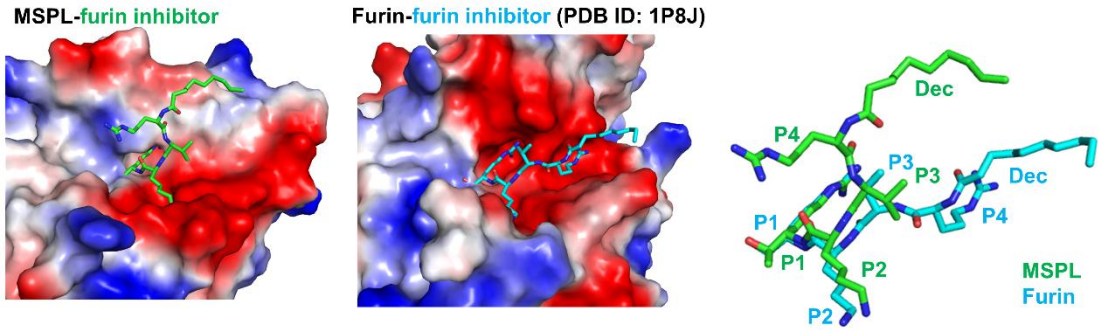

Fig. 5 Conformational differences between the furin inhibitor bound to MSPL and furin
Left panel shows the MSPL:furin inhibitor complex. MSPL and furin inhibitor is shown as an electrostatic surface potential representation and green stick model, respectively. Middle panel shows the furin:furin inhibitor complex (PDB id, 1P8J). Furin and furin inhibitor are shown as an electrostatic surface potential representation and cyan stick model, respectively. Right panel shows the superposition of furin inhibitors bound to MSPL and furin. The potential maps are colored from red (-5kT/e) to blue (+5kT/e).

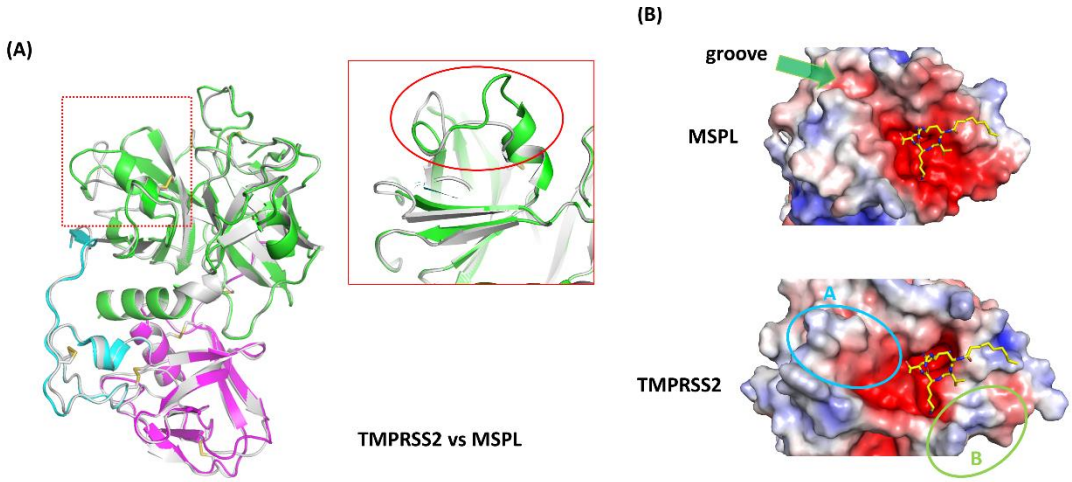

Fig. 6 Homology model analysis of TMPRSS2

(A) A homology model of TMPRSS2 (gray ribbon) was built using MSPL as a template. Superposed analysis revealed large structural differences at the β12-β13 loop region (red rectangle).

(B) Electrostatic surface potential of MSPL and TMPRSS2 SPD. MSPL has a narrow groove that fits with the downstream peptide chain (green arrow). By comparison, in TMPRSS2 the groove is significantly wider, and the peptide binding site is bowl-shaped (cyan oval A). A positively-charged area derived from Lys225 is indicated in green oval B. The potential map is colored from red (-5kT/e) to blue (+5kT/e).

Table S1. Data collection and model refinement statistics [a]

	MSPL (188-558)/decanoyl-RVKR-cmk
Data Collection	
X-ray source	PF-AR NE3A
Space group	$P2_12_12_1$
Unit cell parameters	a = 55.84 Å, b =62.40 Å, c = 171.63 Å, α = 90°, β = 90°, γ = 90°
Wavelength, Å	1.0000
Resolution range, Å	40-2.6 (2.72-2.60)
No. observed reflections	130,814
No. unique reflections	19,086
Multiplicity	6.9 (7.0)
Completeness, %	99.7 (99.6)
$< I >/<\sigma (I) >$	9.5 (2.5)
$CC_{1/2}$	0.995 (0.729)
R_{merge} [b]	0.151 (1.012)
Model Refinement	
Resolution range, Å	40-2.6
No. reflections	17,570
R_{work} / R_{free} [c]	0.180 / 0.232
No. non-H atoms	
Protein	2,911
Oligosaccharide	52
Inhibitor	50
Ion/water	36/81
Average B-factors, Å^2	
Protein	34.3
Oligosaccharide	57.7
Inhibitor	40.8
Ion/water	56.8/26.1
r.m.s deviations	
Bond lengths, Å	0.009
Bond angles, °	1. 241
Ramachandran plot [d], %	
Favored region	95.4
Allowed region	4.6
Outlier region	0.0
PDB code	6KD5

[a] Highest resolution shell is shown in parentheses.

[b] $R_{merge} = \Sigma_{hkl} \, |I_i - \langle I_i \rangle| \, / \, \Sigma_{hkl} \, I_i$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $\langle I_i(hkl) \rangle$ is the average value of I_i(hkl) for all i measurements.

[c] $R_{work} = \Sigma_{hkl} \, ||F_{obs}| - |F_{calc}|| \, / \, \Sigma_{hkl} \, |F_{obs}|$. 8% of the reflections were excluded for R_{free} calculation.

[d] Analyzed with the program *Rampage* (40).

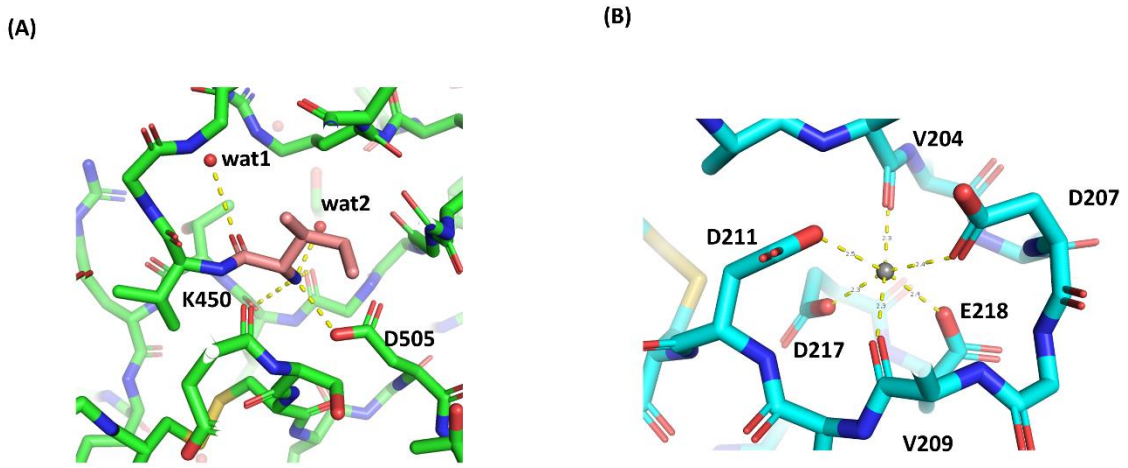

Fig. S1

(A) The interaction of Ile321 (colored in rose red). Ile321 interacts with the side chain of Asp505, backbone of Lys450, and two water molecules.

(B) Calcium ion bound at the loop in the LDLA domain. The calcium ion interacts with octahedral coordinates to Val204, Asp207, Val209, Asp211, Asp217, and Glu218.