Cystal structure of inhibitor-bound human MSPL/TMPRSS13 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

A viral surface envelope glycoprotein, hemagglutinin (HA), is cleaved by host cell proteins of transmembrane protease serine (TMPRSS) family, that triggers a viral infection. 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 the HA, while TMPRSS2 or -4 cleaves monobasic motifs only. To elucidate the HA cleavage of the recognition motif by MSPL, we solved the crystal structure of 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 SPD domain. Furin inhibitor structure shows that the side chain of P1-Arg inserts into highly conserved S1 pocket, whereas side chain of P2-Lys interacts with the Asp/Glu-rich 99's loop that is unique to MSPL. Based on our structure, we designed four inhibitors which showed more specifically to MSPL than the furin inhibitor. We also constructed a homology model of TMPRSS2, that is identified as an initiator of SARS-CoV-2 infection, suggested that TMPRSS2 is more suitable for Ala/Val residues at P2 site than Lys/Arg residues.

Introduction

Mosaic serine protease large form (MSPL) and its splice variant TMPRSS13 identified from human lung cDNA library are members of type II transmembrane serine proteases (TTSPs), characterized by the transmembrane domain near the N-terminus and the catalytic serine protease domain at the C-terminus (Kim, et al., 2001; Kido and Okumura, 2008). TTSPs share the cytoplasmic, transmembrane, stem and catalytic domains in the order of N-terminus to C-terminus (Szabo and Bugge, 2008). 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. Two chains linked by disulfide bridge, so that TTSPs remain membrane bound (Hooper, et al., 2001). 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, revealing the substrate specificity for substrate with Arg or Lys residues in the P1 position. Based on similarities in domain structure, the serine protease domain and the chromosomal location, TTSPs are classified into four subfamilies: Hepsin/TMPRSS, Matriptase, HAT/DESC and Corin (Szabo and Bugge, 2008 & 2011). MSPL and its splice variant TMPRSS13 belong to the Hepsin/TMPRSS subfamily. In this subfamily, Hepsin and Spinesin contain single scavenger receptor cysteine-rich repeat (SRCR) domain in the stem region, while MSPL, TMPRSS2, -3, -4, and -13 contains low-density lipoprotein receptor A (LDLA) domain toward single SRCR domain in the stem region (Szabo and Bugge, 2011). SRCR domain contains approximately 100–110 amino acids and adopts 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, SRCR domain has been divided into three subclasses (A, B and C) (Ojala, et. al., 2007). On the other hands, LDLA domain contains approximately 40 amino acids and contains six conserved cysteine residues that are involved in the formation of disulfide bonds. LDLA domain also carries a calcium ion via highly conserved six residues near the C-terminus. Both disulfide bonds and calcium-binding stabilize the overall structure of LDLA domain. (Daly, et al., 1995). Recently, it was reported that TMPRSS2, -4, -13, and MSPL were involved in the influenza viral spreading by cleaving the glycoprotein hemagglutinin (HA) on the influenza viral surface (Böttcher, et. al., 2009; Chaipan, et al., 2009; Okumura, et. al., 2010; Ohler, et. al., 2012; Zmora, et. Al., 2014). HA is cleaved into HA1 and HA2 subunits by the TMPRSS2, -4, -13, and MSPL. The proteolytic cleavage of HA is essential for the influenza virus infection. HA1 mediates both the host cell binding and the initiation of endocytosis. HA2 controls viral-endosomal fusion (Hamilton, et. al., 2012). Until now, there have been two main HA processing consensus motif in the influenza virus. One is single basic HA processing motif (Q(E)-T/X/-R) in human seasonal influenza viruses. This motif contains a single arginine at the cleavage site. The other is a multiple-basic-residues motif (R-X-X/R-R and K-K/R-K/T-R) in highly pathogenic avian influenza viruses. It contains several basic amino acids at the cleavage site. TMPRSS2 and -4 could recognize single basic HA processing motif, while MSPL and TMPRSS13 could recognize both single basic and multiple basic residues motifs (Okumura, et. al., 2010). It is unclear why only MSPL and its splice variant TMPRSS13 could recognize

multiple-basic-residues 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 (Stieneke-Gröber, et. al., 1992). Previous study showed that the enzyme activity of MSPL was inhibited by the decanoyl-RVKR-cmk that mimics the substrate for the furin (Okumura, et. al., 2010). To date, only one structure of extracellular region of Hepsin have been reported among the Hepsin/TMPRSS family (Somoza, et al., 2003). The crystal structure of Hepsin revealed that SRCR domain was located at the opposite side of the active site of SPD, and these domains are splayed apart. As the Hepsin lacks LDLA domain, the relative orientation of LDLA, SRCR and SPD domains in other members of Hepsin/TMPRSS family, such as MSPL, is still unknown. To elucidate the spatial arrangement of three domains and substrate specificity, we determined the crystal structure of extracellular region of MSPL in complex with the decanoyl-RVKR-cmk peptide at 2.6 Å resolution. Unexpectedly, the overall structure of MSPL provides that the spatial arrangement of SRCR and SPD domains in MSPL is markedly distinct from that in Hepsin. The complex structure provides how the MSPL could recognize the both single- and multiple-basic-residues motif. Based on the structure, we optimized the sequence of peptidyl inhibitors to the MSPL, and obtained one with 220-fold higher inhibitory activity. In addition, we constructed a homology model of TMPRSS2, that is involved in SARS-CoV-2 infection process, and investigated the target sequence preference to S1/S2 site of SARS-CoV-2 spike protein.

Results

Overall structure of 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 have expressed and purified the extracellular region (residues 187-586) of hMSPL and crystallized with decanoyl-RVKR-cmk, known as a furin inhibitor. We collected the diffraction data at Photon Factory AR-NE3a (Tsukuba, Japan) and solved the structure at 2.6 Å (Fig. 1B). To our knowledge, this is the first structure description of the LDLA-containing hepsin/TMPRSS subfamily. The refined model contains the hMSPL with residue range of 188-558, except 319 and 320, decanoly-RVKR-cmk, and a calcium ion. We also observed glycans attached on Asn250 and Asn400.

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 cleaving at Arg320-Ile321 and residues 321-581 region converted to the mature SPD (Okumura, et. al., 2010). We found that Ile321 is located in a pocket and N atom is interacted with Asp505 (Fig. S1A), therefore, this structure could be a mature form and hMSPL was processed by intrinsic protease during the cell expression. The LDLA domain of hMSPL is 24 amino acids in length, and composed of two turns and a short α-helix. A canonical LDLA domain has N-terminal antiparallel β-sheet and three disulfide bonds (Daly, et al., 1995), therefore, LDLA of MSPL lacks the N-terminus half. The SRCR domain of MSPL belongs to the Class C of atypical

cysteine residues pattern like Hepsin. The structures of SRCR domains of MSPL and hepsin are very similar despite low sequence homology between SRCRs (23% sequence identity).

To date, hepsin (PDB entry: 1P57) is the only protein that is available a 3D structure among the same TTSP subfamily, however, hepsin lacks LDLA domain. We compared and superposed hMSPL and hepsin structure (Fig. 2). Those two SPDs showed quite small backbone root mean square deviation (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). Although SPD and SRCR domain of hMSPL and Hepsin are almost identical, the relative arrangement of each domain is quite different (Fig. 2B). When SPD of Hepsin and MSPL are fitted, the SRCR domain of MSPL is rotated ~80 degree against that of Hepsin. The difference may be caused by the presence of LDLA domain in MSPL. By the presence of LDLA domain, SRCR and SPD domains of MSPL have more tightly packed than these domains of Hepsin in which these domains are splayed apart. Accordingly, short parallel β-sheet between the N-terminal segment and SPD domain was observed in MSPL, while in Hepsin, the C-terminal end is located and making an anti-parallel β-sheet (Fig. 2A).

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

As expected, the SPD of MSPL displays the conserved architecture of the trypsin- and chymotrypsin-like (S1 family) serine protease (Fig. 1B). In the activated MSPL, the Ile321 at cleavage site forms a salt bridge with the conserved Asp505 residue located immediately prior to the catalytic Ser506

(Fig. S1A). This interaction might be generated by the activating cleavage. It also helps the formation of the S1 pocket and oxyanion hole by inducing conformational changes in the hairpin loop which formed the S1 pocket and the oxyanion hole (Fig.3). This salt bridge was also observed in other proteases such as plasma Kallikrein and Hepsin (Pathak, et. al., 2013, PDB entry: 1Z8G). 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 dec (decanoyl group) (Fig. 1C, 3). The covalent interaction between the furin inhibitor and catalytic residues His361 and Ser506 is formed by the nucleophilic attack on cmk moiety. P1-Arg inserts into the deep S1 pocket, and its carbonyl oxygen atom directly bind to the backbone amides of the oxyanion hole (Gly504 and Ser506). The guanidino group of P1-Arg forms salt bridges with the side chains of Asp500, as well as 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 hepsin/TMPRSS subfamily (Fig. 4). The interaction between P1-Arg and MSPL was characteristic of trypsin- and chymotrypsin-like serine proteases. On the other hand, P2-Lys interacts with residues at so-called 99-loop (chymotrypsinogen numbering) that containing 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 mediated hydrogen bonds with side chains of Asp406 and catalytic Asp409. Interestingly, with the exception of catalytic Asp409, the residues interacting side chain of P2-Lys are not conserved among Hepsin/TMPRSS subfamily (Fig. 4, cyan dot). This is presumably the

reason why other TMPRSSs and Hepsin do not recognize the di-basic but the single basic motif. Compared to 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. On the other hand, the backbone of P4-Arg forms no hydrogen bond with the MSPL. The side chain of P4-Arg extends into the bulk solvent. N-terminal decanoyl moiety makes van der Waals contacts with Thr528 and Gln532 at 220-loop (chymotrypsinogen numbering). One ordered sulfate ion is located in close proximity to both P3-Val and P4-Arg. It forms hydrogen bonds with the backbone amides of P2 Lys and P3-Val. It also makes 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 (Henrich, et al., 2003). Although furin also has same Ser-His-Asp catalytic triad as MSPL, its catalytic domain belongs to the superfamily of subtilisinlike serine protease (Siezen, et al., 1997). The catalytic domain of Furin has different overall fold from that of MSPL which belongs to trypsin- and chymotrypsinlike (S1 family) serine protease. Despite a different overall fold of MSPL and furin, the inhibitor peptide (decanoly-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. On the other hands, in the Furin:furin inhibitor complex structure, the inhibitor adopts an extended conformation. The P3 site is directed away from the Furin, whereas P1, P2, and P4 site contacts with the Furin. The structural difference between furin-bound inhibitor and MSPL-bound one enables the inhibitor to bind to both the MSPL and the Furin.

Orientation of the extracellular region of MSPL with respect to the plasma membrane

Since only three residues between the transmembrane domain and N-terminal Thr188 from which our structure model starts, the extracellular region of MSPL might be located very close to the plasma membrane. Indeed, the region that was predicted to be close to the plasma membrane enriched in the basic residues, such as Arg191, Lys193, Lys 213, Lys 215, and Arg556 (Fig. 2C). Extracellular region of Hepsin also was suggested to be lying flat against the plasma membrane (Somoza, et al., 2003). MSPL and Hepsin may be in ready for the substrate to bind in close proximity to the transmembrane, however, the extracellular region of MSPL oriented upside down with respect to that of Hepsin.

Inhibitory assay of the new peptidyl inhibitors

In our structure, the P4-Arg of the furin inhibitor are distant from the active site in the SPD domain of MSPL. Therefore, we hypothesized that the replacement of residues at the P3 site by the basic residues might enhance the inhibitory potency towards the human MSPL, because MSPL has the acidic surface that

could interact with basic residues in inhibitors (Fig.5). We designed four inhibitor peptides (Ac-KQRR-cmk, Ac-KKKR-cmk, Ac-KKRR-cmk, and Ac-KRRR-cmk), that derived from the furin inhibitor. To test this hypothesis, we performed enzyme inhibition assays using purified soluble recombinant MSPL. Inhibition potency was determined by measuring the residual enzyme activity using the fluorogenic substrate (Pyr-Arg-Thr-Lys-Arg-MCA) after pre-incubation of recombinant MSPL with varying concentrations of each non-fluorescent inhibitor. As expected, all new inhibitor peptides showed significantly (4-10 folds) more inhibitory potency on recombinant MSPL than decanoyl-RVKR-cmk (Table 1). We also examined their inhibitory potency towards Furin, because the replacement of residues at P3 site by the Arg is expected to reduce the inhibitory potency towards the human Furin. As expected, all new inhibitor peptides showed significantly at least 200-fold high inhibitory potent than decanoyl-RVKR-cmk.

Discussion

In this study, we have elucidated the spatial arrangement of three (LDLA, SRCR, and SPD) domains and substrate specificity of MSPL that could be involved in the multiple infections with the HPAI virus. We also succeeded in the discovery of specific inhibitors for MSPL. The finding might be useful for the design of anti-influenza drug which prevent HAPI virus uptake into a host cell. Indeed, an immunofluorescence assay was performed to investigate the inhibitory activity of one of novel peptides for the mutant HAPAI viral infections (data not shown). It showed that the novel peptide inhibited the mutant HAPAI

viral infection more effectively than the furin inhibitor. Currently, there is no drug for the HAPAI virus with pandemic potential.

MSPL also contributed to cleave and activate severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) spike proteins (Zmora, et al., 2014). Recently, Shi et al. reported that MSPL could cleaved the porcine epidemic diarrhea virus (PEDV) spike protein (Shi, et al., 2017). Shi et al. demonstrated that MSPL facilitated the replication of PEDV and was involved in the viral infection. Although cleavage sites of these spike proteins were still unknown, MSPL-specific inhibitor may be useful for the therapeutic drug development for the treatment of not only HPAI but also SARS-CoV, MERS-CoV and PEDV.

Our structure helps tertiary structure prediction of TMPRSS2, -3, and -4 that contains the LDLA domain toward single SRCR domain in the stem region. In these TMPRSSs, TMPRSS3 and -4 are involved in the epithelial Na^+ channel (ENaC) activation (Antalis, et al., 2010). Moreover, human TMPRSS2 cleaves the spike protein of SARS-CoV-2, which triggers the viral infection (Hoffmann, et al., 2020a; Meng, et al., 2020). To investigate the features of the TMPRSS2, we constructed the homology model (Fig. 6). Eight out of nine disulfide bonds were conserved (Fig. 4), the relative domain alignment of TMPRSS2 is almost similar to MSPL. In the SPD domain, the β12-β13 loop region has changed significantly (Fig. 6). This structural change results in a wide substrate-binding groove, so that TMPRSS2 may become easy to capture the target peptide. Furthermore, the position of Glu404, that is important residue in P2-Lys recognition in MSPL, is replaced by Lys225 in TMPRSS2 (Fig. 4, 6B). As

mentioned earlier, this substitution leads to a preference for 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 (Walls, et al., 2020; Coutard, et al., 2020; Hoffmann, et al., 2020b). In summary, this homology model well reflects the feature of TMPRSS2 target peptide recognition.

TMPRSS4 has been reported to cleave the inhibitory ENaCγ subunit. Moreover, seven missense TMPRSS3 mutants (D103G, R109W, C194F, R216L, W251C, P404L and C407R) that associated the human deafness were unable to activate the ENaC (Antalis, et al., 2010). One of seven missense mutants associated the loss of hearing, D103G, was found in the LDLA domain of TMPRSS3 (Wattenhofer, et. al., 2005; Guipponi, et. al., 2002). Since Asp103 in the TMPRSS3 corresponds with Asp221 in MSPL, LDLA structure stabilizing by calcium-binding may be important for the TMPRSS3's function.

We speculate that the Ca^{2+} binding sites in the LDLA domain may be involved in the internal electrostatic binding to basic residues (Lys/Arg) located near the active sites in SPD domain in the zymogen form of MSPL. In fact, the structure of the LDL receptor (LDLR) which contains seven LDLA domains showed that there are internal interactions between Ca^{2+} binding sites of fourth and fifth LDLA domains and Lys residues of the β-Propeller domain. It is thought that these interactions blocked the ligand such as the LDL access to the Ca^{2+} binding sites the LDLA domain. It is unclear whether the similar internal interaction between the LDLA domain and SPD domain, since no structure of zymogen form of MSPL and other TMPRSSs containing LDLA, SRCR and SPD domains has been solved. If a similar internal interaction between the Ca^{2+} binding sites

of the LDLA domain and the basic residues near the active site in SPD domain exists in the zymogen form of MSPL, it may block its substrate access to the active site. Indeed, the mutations (D103G, R109W and C194F) in LDLA and SRCR as well as SPD domains of TMPRSS3 affected its autoactivation by proteolytic cleavage at the junction site between the SRCR and the SPD domains (Guipponi, et al., 2002). As to the biological function, the study using the TMPRSS13 knockout mice showed that TMPRSS13 deficiency occurred the abnormal skin development. (Madsen, et. al., 2014) However it remains unknown how the TMPRSS13 deficiency causes the epidermal barrier defect. We also found that MSPL as well as Furin could cleave myostatin, a potent negative regulator of myogenesis, into mature form (manuscript in preparation). We also showed that the activation of myostatin by MSPL is related to the unloading caused muscle atrophy. This study is helpful for further studies on biological function of MSPL.

Methods

Cloning, expression, and purification

The recombinant soluble hMSPL were induced into the serum free culture medium (SFCM) from a previously established stable cell line expressing hMSPL (Okumura, et al., 2010). About 10-liter of SFCM were concentrated by ulutrafiltration using Pellicon XL 50 (Merck Millipore) and then the concentrated SFCM were applied to Anti-FLAG M2 agarose gel equilibrated by 50 mM Tris-HCl, 150 mM NaCl, pH 7.4, (TBS). Binding proteins were eluted by 0.1M Glycine-HCl, pH 3.5. The fractions eluted the recombinant hMSPL were collected and were dialyzed by phosphate-buffered saline (PBS).

Complex formation, crystallization, and data collection

The peptide inhibitor (decanoyl-RVKR-cmk) was purchased from Merck Millipore and reconstituted in dimethylsulphoxide (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 precipitation. The crystallization screening was performed by mixing of 1 µL of the MSPL-inhibitor solution with 1 µL of reservoir solution with the hanging-drop vapor-diffusion method. The MSPL-inhibitor complex was crystallized at 15 °C with the reservoir solution comprised of 0.1 M HEPES (pH 7.5), 2.4 M ammonium sulfate. Prior to the data collection, the single crystal was transferred to the cryoprotectants [20% of glycerol and 80% of the reservoir] for the five seconds, and then flash cooled by liquid nitrogen. Diffraction dataset of the MSPL:furin inhibitor complex crystal

were collected at the beamline NE3A at the Photon Factory Advanced Ring (PF-AR). The crystal belongs 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 (Battye, et al. 2011), followed by Aimless (Evans and Murshudov, 2013). 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 (Vagin and Teplyakov, 2010), with SPD of human plasma kallikrein (PDB code: 2ANY), that shows the highest sequence identity score (46.1%), as a search model. The model of SPD was manually fixed with COOT (Emsley and Cowtan, 2004) and refined with Refmac5 (Murshudov, et al., 2011). After the SPD of MSPL was well refined, interpretable electron density of the unmodeled region was appeared, and then the model of the LDLA and SRCR domain were manually built. The final model contained one MSPL, one furin inhibitor, four sugars, 80 ions, and 65 waters, 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 structure in figures were prepared using PyMOL (http://www.pymol.org/). The MSPL/peptide inhibitor interfaces were analyzed using LIGPLOT (Wallace, et. al., 1995). The coordinate and structure factors of MSPL-peptide inhibitor complex have been deposited to

RCSB Protein Data Bank (PDB code: 6KD5).

Homology modelling of TMPRSS2

The sequence alignment of extracellular region of MSPL and TMPRSS2 was obtained by BLAST webserver (https://www.uniprot.org/blast/). The amino acid identity between MSPL and TMPRSS2 was 39.8% with a score as 704, and E-value of 1.1e-86. Homology model of TMPRSS2 was build using MODELLER (Šali and Blundell, 1993). Electrostatic surface potentials were calculated by APBS sever (http://server.poissonboltzmann.org/).

Inhibitor assay

Enzyme inhibition assays were performed using fluorogenic substrate (Pyr-Arg-Thr-Lys-Arg-MCA) by measuring the residual hydrolytic activity after pre-incubation with increasing concentrations of four new designed inhibitors (KQRR-cmk, KKKR-cmk, KKRR-cmk and KRRR-cmk). The enzyme was pre-incubated with an increasing concentration of new designed inhibitors at 37°C for 5 min. After preincubation of enzyme and inhibitor, fluorogenic substrate was added. The residual enzyme activity was measured by reading the fluorescence over time at excitation and emission wavelengths of 360 nm and 465 nm, respectively. IC_{50} was determined from the plots by using different concentrations of new designed inhibitors.

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

Andersen C. B., Moestrup S. K. (2014). How calcium makes endocytic receptors attractive. *Trends Biochem. Sci.*, 39, 82-90.

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.

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.

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.

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.

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.

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.

Emsley P., Cowtan K., (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.*, 60, 2126-2132.

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.

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.

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.

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.

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.

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.

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.

Hohenester E., Sasaki T., Timpl R. (1999). Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. *Nat. Struct. Biol.*, 6, 228-232.

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.

Kido H., Okumura Y., (2008). MSPL/TMPRSS13. *Front. Biosci.*, 13, 754-758.

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.

Madsen D. H., Szabo R., Molinolo A. A., Bugge T. H. (2014). TMPRSS13 deficiency impairs stratum corneum formation and epidermal barrier acquisition. *Biochem J.*, 461, 487-495.

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.

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.

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.

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.

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.

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.

Pathak M., Wong S. S., Dreveny I., Emsley J. (2013). Structure of plasma and tissue kallikreins. *Thromb. Haemost.*, 110, 423-433.

Šali R. Blundell T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, 234, 779-815.

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.

Siezen R. J., Leunissen J. A. (1997). Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.*, 6, 501-523.

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.

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.

Szabo R., Bugge T. H. (2008). Type II transmembrane serine proteases in development and disease. *Int. J. Biochem. Cell Biol.*, 40, 1297-12316.

Szabo R, Bugge T. H. (2011). Membrane-anchored serine proteases in vertebrate cell and developmental biology. *Annu. Rev. Cell Dev. Biol.*, 27, 213-235.

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.

Vagin A., Teplyakov A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.*, 66, 22-25.

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.

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.

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.

Winn M. D., Ballard C. C., Cowtan K. D., Dodson E. J., Emsley P., Evans P. R., Keegan R. M., Krissinel E. B., Leslie A. G., McCoy A., McNicholas S. J., Murshudov G. N., Pannu N. S., Potterton E. A., Powell H. R., Read R. J., Vagin A., Wilson K. S. (2011). Overview of the CCP4 suite and current developments. *Acta. Cryst. D Biol.*

Crystallogr., 67, 235-242.

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.

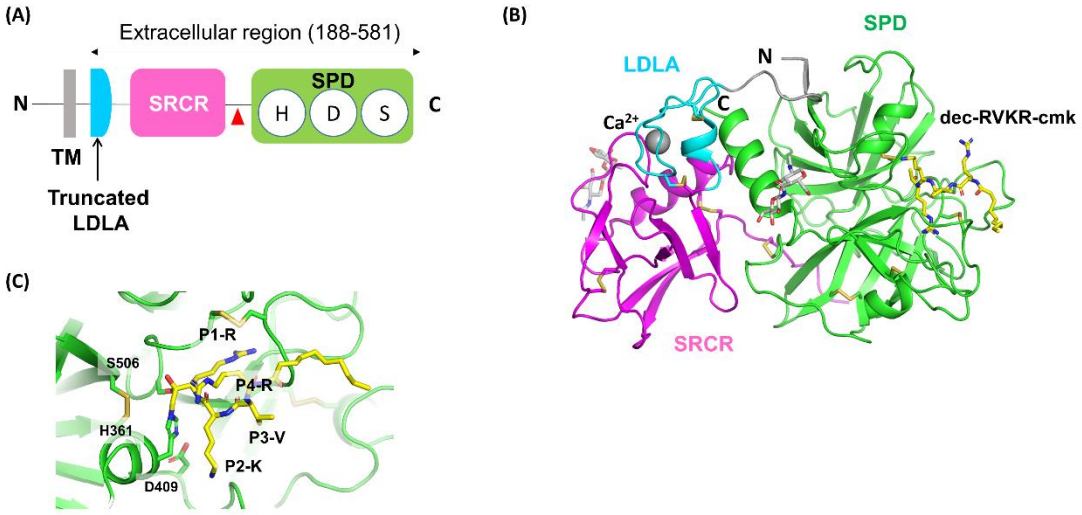

Fig. 1 Overall structure of MSPL extracellular domain.

(A) Schematic presentation of full-length MSPL. MSPL is composed of 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) and converted to a mature form.

(B) Ribbon representation of 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 Ca2+ in the center of the loop. 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 catalytic triad residues.

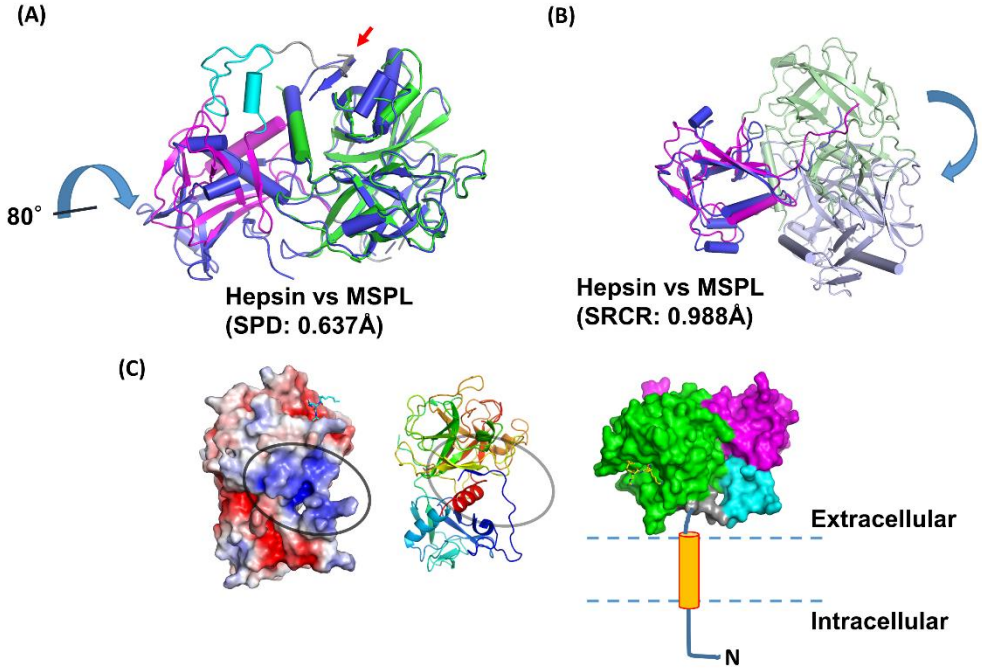

Fig. 2 Comparison of MSPL and Hepsin.

(A) Hepsin (colored in blue) and MSPL (colored in cyan (LDLA), magenta (SRCR), and green (SPD)) were superposed with SPD domain. RMSD value is 0.637 Å calculated with 197 Cα atom position. A β-sheet interaction of N-terminus and SPD in MSPL is replaced by C-terminus in Hepsin (red arrow). The Hepsin SRCR domain is rotated about 80° against 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 SRCR domain. RMSD value is 0.988 Å calculated with 59 Cα atom position.

(C) (Left) Electrostatic surface potential of MSPL extracellular domain. A characteristic positively-charged area (gray oval) composed of Arg191, Lys193, Lys 213, Lys 215, and Arg556, is supposed to be a contact surface of cell membrane. The potential map is colored by 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 membrane-anchored full-length MSPL.

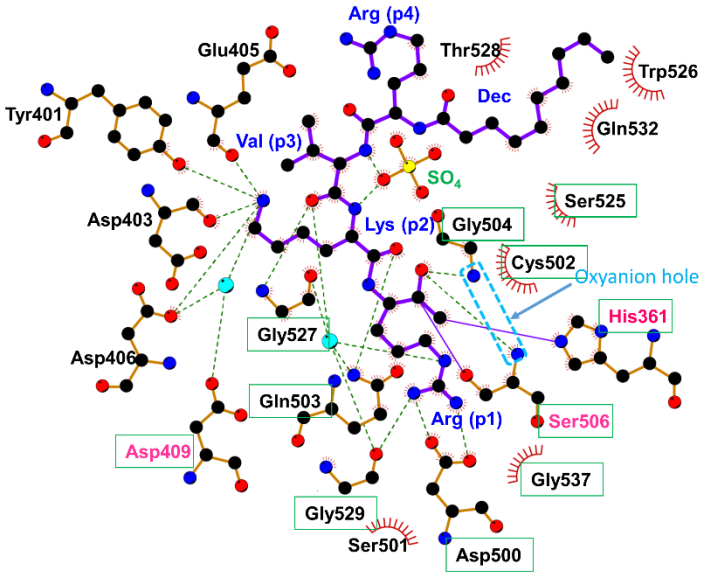

Fig. 3 Interaction of furin inhibitor to MSPL

The SPD of MSPL and furin inhibitor were in orange and purple, respectively. Nitrogen atoms, blue; oxygen atoms, red; carbon atoms, black; sulfur atoms, yellow. Dashed lines indicate 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 square denotes oxyanion hole. The catalytic triad of three amino acids are highlighted in red. The conserved residue among MSPL, TMPRSS2-4, and Hepsin are indicated as green square. The figure was prepared with LIGPLOT (Wallace, et al., 1995)

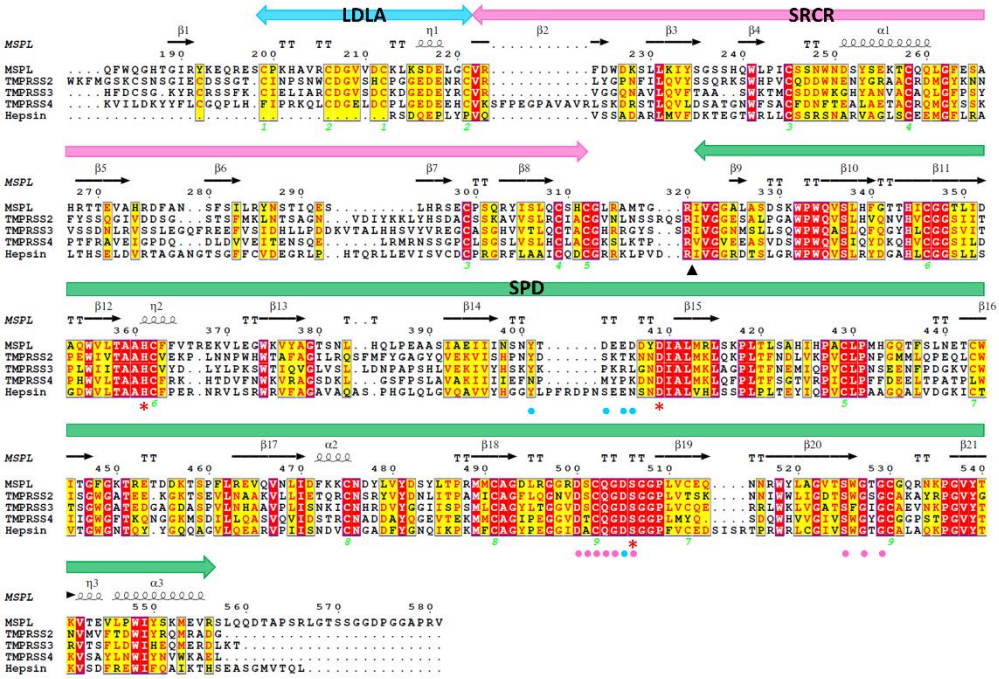

Fig. 4 Multiple sequence alignment of MSPL extracellular region with members of hepsin/TMPRSS subfamily.

Extracellular region of human MSPL (182-581), human TMPRSS2 (110-492), human TMPRSS3 (70-454), human TMPRSS4 (55-437), and human Hepsin (50-417) are aligned by Clustal W program (Thompson, et al., 1994), followed by coloring with ESPRIPT (Gouet, et al., 2003). Red asterisk indicates the catalytic triad. The amino acid sequences were referred 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 the residues that interact with P1 and P2 site of furin inhibitor, respectively. Green numbers denote the disulfide pairing of MSPL.

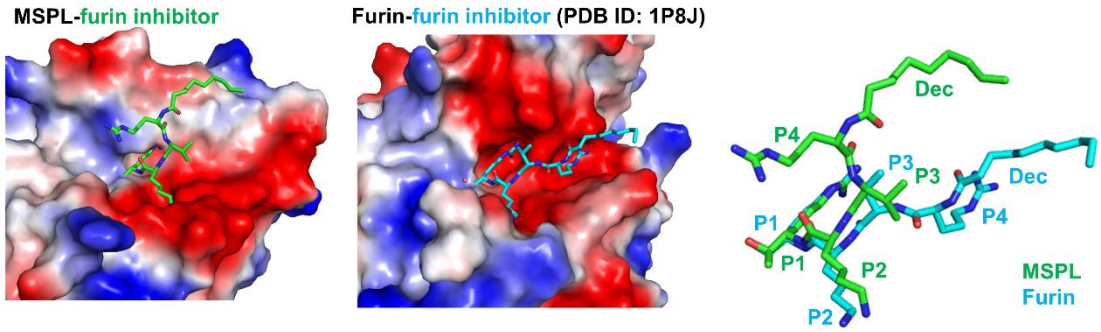

Fig. 5 Conformational difference of the furin inhibitor bound to MSPL and furin.

Left panel shows the MSPL:furin inhibitor complex. MSPL and furin inhibitor is shown in electrostatic surface potential representation and green stick model, respectively. Middle panel shows the Furin:furin inhibitor complex (PDB id, 1P8J). Furin and furin inhibitor is shown in 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 were colored by from red (-5kT/e) to blue (+5kT/e).

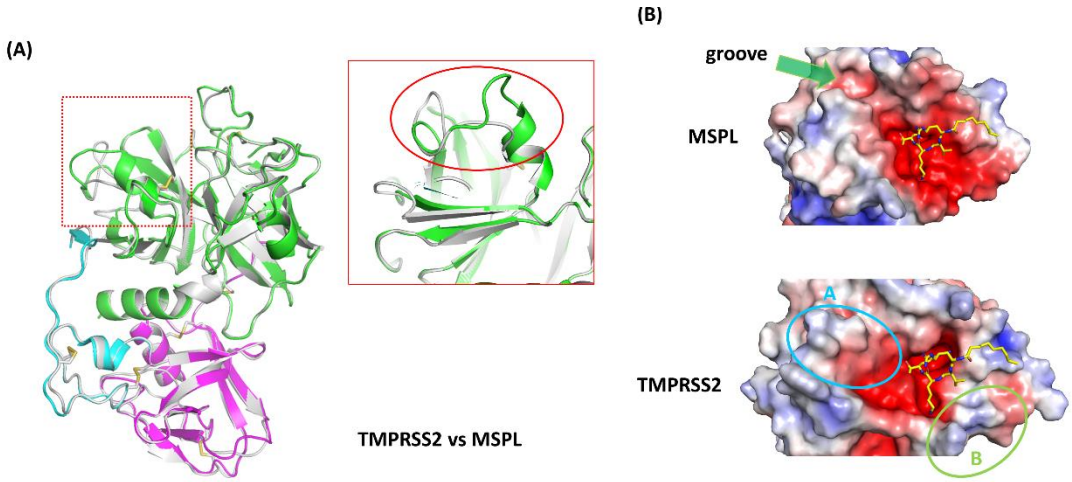

Fig. 6 Homology model analysis of TMPRSS2.

(A) A homology model of TMPRSS2 (gray ribbon) was built using MSPL as a template. Superposed analysis revealed a large structural difference at 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). In TMPRSS2, the groove was widened and the peptide binding site become a bowl-shaped (cyan oval A). A positively-charged area derived from Lys225 is indicated in green oval B. The potential map is colored by from red (-5kT/e) to blue (+5kT/e).

Table 1 Inhibitory effects of designed peptides on the enzyme activity of Furin/MSPL.

Inhibitor	IC_{50} for Furin	IC_{50} for MSPL
dec-RVKR-cmk	2.6 nM	10 nM
Ac-KQRR-cmk	475.8 nM	2.1 nM
Ac-KKKR-cmk	112.4 nM	2.8 nM
Ac-KKRR-cmk	121 nM	1.1 nM
Ac-KRRR-cmk	193 nM	1.7 nM

IC_{50} values of the furin inhibitor and four novel designed inhibitors for Furin and MSPL were determined in the Methods.

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

	MSPL(188-558)/decanoyl-RVKR-cmk
Data Collection	
X-ray source	PF-AR NE3A
Spacegroup	$P2_12_12_1$
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
Mutliplicity	6.9 (7.0)
Completeness, %	99.7 (99.6)
$< I >/<\sigma (I) >$	9.5 (2.5)
R_{merge} [b]	0.169 (0.761)
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 parenthesis.

[b] $R_{merge} = \Sigma_{hkl} |I_i - <I_i>| / \Sigma_{hkl} I_i$, where $I_i(hkl)$ is the intensity of the ith measurement of reflection hkl and $<I_i(hkl)>$ 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*.

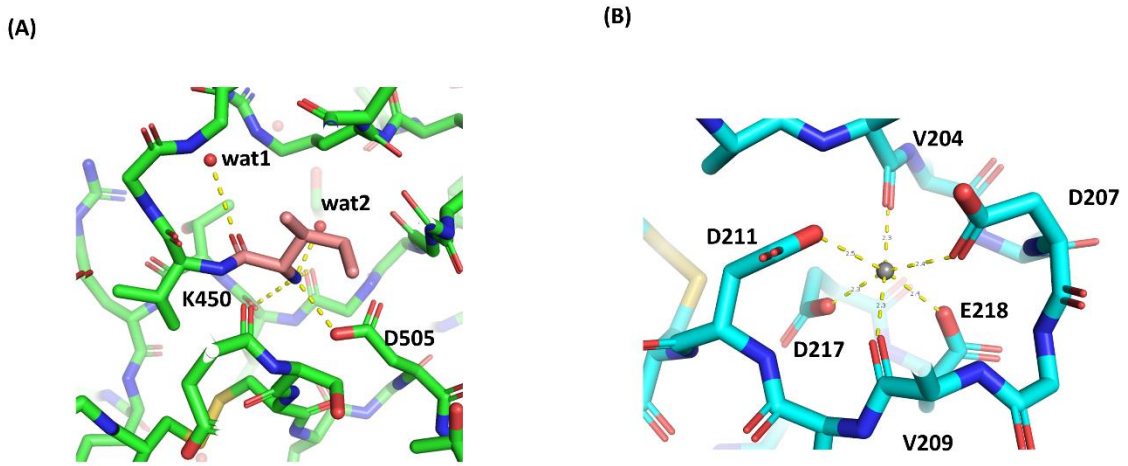

Fig. S1

(A) The interaction of Ile321 (colored in rose red). Ile321 is interacted with side chain of Asp505, backbone of Lys450, and two waters.

(B) A calcium ion bound at the loop in LDLA domain. The calcium ion is interacted an octahedral coordinate with Val204, Asp207, Val209, Asp211, Asp217, and Glu218.