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

Ayako Ohno1,*, Nobuo Maita2,*, Takanori Tabata3, Hikaru Nagano4, Kyohei Arita5, Mariko Ariyoshi6, Takayuki Uchida1, Reiko Nakao1, Anayt Ulla1, Kosuke Sugiura1,7, Koji Kishimoto8, Shigetada Teshima-Kondo4, Yuushi Okumura9, Takeshi Nikawa1

Infection of certain influenza viruses is triggered when its HA is cleaved by host cell proteases such as proprotein convertases and type II transmembrane serine proteases (TTSP). HA with a monobasic motif is cleaved by trypsin-like proteases, including TMPRSS2 and HAT, whereas the multibasic motif found in high pathogenic avian influenza HA is cleaved by furin, PCS/6, or MSPL. MSPL belongs to the TMPRSS family and preferentially cleaves [R/K]-K-K-R sequences. Here, we solved the crystal structure of the extracellular region of human MSPL in complex with an irreversible substrate-analog inhibitor. The structure revealed three domains clustered around the C-terminal α-helix of the SPD. The inhibitor structure and its putative model show that the P1-Arg inserts into the S1 pocket, whereas the P2-Lys and P4-Arg interacts with the Asp/Glu-rich 99-loop that is unique to MSPL. Based on the structure of MSPL, we also constructed a homology model of TMPRSS2, which is essential for the activation of the SARS-CoV-2 spike protein and infection. The model may provide the structural insight for the drug development for COVID-19.

DOI 10.26508/lsa.202000849 | Received 13 July 2020 | Revised 12 March 2021 | Accepted 12 March 2021 | Published online 5 April 2021

Introduction

Mosaic serine protease large form (MSPL), and its splice variant TMPRSS13, was originally identified from a human lung cDNA library and is a member of the type II transmembrane serine proteases (TTSPs) (Kim et al, 2001; Kido & Okumura, 2008). TTSPs comprise a transmembrane domain near the N-terminus and a trypsin-like serine protease domain (SPD) at the C-terminus. Human MSPL is expressed in lung, placenta, pancreas, and prostate (Kim et al, 2001). MSPL is reported to cleave the spike protein of porcine epidemic diarrhea virus (Shi et al, 2017), MERS- and SARS-CoV (Zmora et al, 2014), certain influenza virus HAs (Okumura et al, 2010), and pro-hepatocyte growth factor (Hashimoto et al, 2010), but the physiological function of MSPL is poorly understood. TTSPs share a similar overall organization comprising an N-terminal cytoplasmic domain, transmembrane region, and stem/catalytic domains at the C terminus (Szabo & Bugge, 2008). All TTSPs are synthesized as single-chainzymogens 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 (Bugge et al, 2009). The catalytic domain contains a highly conserved “catalytic triad” of three amino acids (His, Asp, and Ser). Like all other trypsin-like serine proteases, MSPL possesses a conserved Asp residue on the bottom of the S1 substrate-binding pocket; therefore, it accepts substrates with Arg or Lys in the P1 position. Based on similarities in the domain structure, the SPD, TTSPs are classified into four subfamilies: hepsin/TMPRSS, matriptase, HAT/DESC, and corin (Szabo & Bugge, 2008, 2011; Antalis et al, 2011; Böttcher-Friebertshäuser, 2018). MSPL belongs to the hepsin/TMPRSS subfamily. In this subfamily, hepsin and matripsin (TMPRSS3) contain a single scavenger receptor cysteine-rich repeat (SRCR) domain in the stem region, whereas MSPL, TMPRSS2, and -3, and -4 contain an additional low-density lipoprotein receptor A (LDLA) domain near the single SRCR domain in the stem region (Antalis et al, 2011; Szabo & Bugge, 2011). Furthermore, enteropeptidase has additional insertions of SEA, LDLA, CUB, MAM, and CUB domains between the transmembrane and the LDLA domain (Kitamoto et al, 1994). The SRCR domain has ~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) (Ojala et al., 2007). The canonical LDLA domain contains ~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 (Daly et al., 1995).

Limited proteolysis of the glycoprotein on the viral surface mediated by a host protease is a key step in facilitating viral infection. The influenza viral HA0 is cleaved by various host proteases and divided into HA1 and HA2 subunits, where HA1 mediates host cell binding as well as the initiation of endocytosis and HA2 controls viral-endosomal membrane fusion (Hamilton et al., 2012). Previous studies show that TMPRSS2, -4, DESC1, HAT, and MSPL activate the influenza virus by cleaving HA0 (reviewed by Böttcher et al. [2006]; Chaipan et al. [2009]; Okumura et al. [2010]; Antalis et al. [2011]; Ohler and Becker-Paul, [2012]; Böttcher-Friebertshäuser et al. [2013]; Zmora et al. [2014]; Böttcher-Friebertshäuser et al. [2018]). A newly synthesized HA is cleaved during its transport to the plasma membrane in the trans-Golgi network by furin or TMPRSS2, whereas HAT cleaves it at the cell surface during viral budding (reviewed in Böttcher-Friebertshäuser et al. [2018]). There are two types of cleavage site sequences; monobasic motifs have single or discrete basic residues such as [Q/E]-[T/X]-R or R-X-X-R (vertical arrow indicates the cleavage groove), and multibasic motifs are composed of Lys/Arg-rich sequences such as R-X-[K/R]-R↓↓. The multibasic motif is found in highly pathogenic avian influenza virus, and is mainly cleaved by furin, PCS↓↓↓ (Stieneke-Gröber et al. [1992]; Horimoto et al. [1994]), and MSPL (Kido et al. [2009]). Some highly pathogenic avian influenza virus variants, such as HSN2 (Lee et al., 2005) and H7N3 (Bulach et al., 2010) have the multibasic motif with Lys at the P4 position (K-K-R↓↓↓), and these HA proteins are not cleaved by furin nor PCS↓↓↓, but MSPL (Thomas, 2002; Remacle et al., 2008; Kido et al., 2009). Therefore, MSPL is a key protease to protect humans from an outbreak of novel avian influenza A virus.

To date, the extracellular region of human hepsin (Somoza et al., 2003; Herter et al., 2005) and serine protease domain (SPD) of enteropeptidase (Lu et al., 1999; Simeonov et al., 2012) are the only structures reported among the hepsin/TMPRSS family. 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 is unknown. To elucidate the spatial arrangement of the three domains and multibasic motif recognition, we determined the crystal structure of the extracellular region of human MSPL in complex with the irreversible peptidic inhibitor decanoyl–RVKR–cmk 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. Although the inhibitor adopts an artificial conformation because of crystal packing, the predicted peptide model explains how MSPL is able to recognize both Lys or Arg as P4 residues. In addition, we constructed a homology model of human TMPRSS2, which is reported to cleave the spike protein of SARS-CoV-2 (Bestle et al., 2020; Hoffmann et al., 2020a, 2020b). The human TMPRSS2 model reveals a wide binding cleft at the S1’ position, suggesting that TMPRSS2 can capture the target peptides of flexible conformations.

Results
Overall structure of the human MSPL extracellular region

The extracellular region of human MSPL is composed of an LDLA domain (residues 203–226), an SRCR domain (residues 227–317) and a SPD (residues 326–561) (Fig 1A). We expressed and purified the extracellular region (residues 187–586) of human MSPL and crystallized the protein with decanoyl–RVKR–cmk, which is known as furin inhibitor I. Diffraction data were collected at the Photon Factory AR-NE3a and the structure was solved to a resolution of 2.6 Å (Fig 1B and Table S1). This is the first published structure of an LDLA-containing hepsin/TMPRSS subgroup protein. The refined model contains the human MSPL with the residue range of 193–563, decanoyl–RVKR–cmk, and a calcium ion (Fig S1B). Residues of 187–192, 324–325, and 564–586 regions were missing because of the disorder. There are four potential N-glycosylation sites, and two N-glycans attached to Asn255 and Asn405 were observed, but no phosphorylated residues were found (Murray et al., 2017).

The extracellular region of human MSPL is composed of the noncatalytic portion of the N-terminal region (LDLA and SRCR domain) and the SPD at the C-terminus (Fig 1B). The three domains are linked to each other by disulfide bonds. The human MSPL is activated by hydrolytic cleavage at Arg325-Ile326 then residues in the 326–586 region are converted to the mature SPD (Okumura et al., 2010). Ile326 (Ile16 in chymotrypsin; hereafter, the residue numbers in parentheses denote the corresponding chymotrypsin residue number, see Fig 3B) is located in a pocket where the N atom interacts with the side chain of Asp510(194) (Fig S1A). Therefore, this structure could represent the active form in which human MSPL is processed by an intrinsic protease during expression in the cell. The LDLA domain of human MSPL 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 (Daly et al., 1995). Therefore, the LDLA of human MSPL lacks half of the canonical N-terminal region. Because the SRCR domain of human MSPL only has two disulfide bonds, it does not belong to either group A or B (Sarrias et al., 2004). Intriguingly, the 3D structures of the SRCR domains of human 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 (Ojala et al., 2007).

To date, 3D structures of SRCR-SPD of hepsin (Protein Data Bank [PDB] entry: 1P57 & 1Z8G) and SPD of enteropeptidase (PDB entry: 1EK8 & 4DGJ) have been reported in the same hepsin/TMPRSS subfamily. Here, we compared the structures of human MSPL and hepsin (Figs 1A and 2A–C). The root-mean-square deviation of the two SPDs (r.m.s.d. of Ca atoms = 0.637 Å), as well as the SRCR domains (r.m.s.d. of Ca atoms = 0.988 Å), are quite small. Although
the SPD and SRCR domains of human MSPL 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 human 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 human MSPL. The LDLA, SRCR, and SP domains of human 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 SP domain was observed in human MSPL, whereas the C-terminal end of hepsin forms an antiparallel β-sheet (Fig 2A).

There are only six residues between the transmembrane domain and the N-terminal Thr193 residue of our structural model. Hence, the extracellular region of human 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 Arg196, Lys198, Lys218, Lys220, and Arg561 (245) (Fig 2C). The extracellular region of hepsin is also thought to lie flat against the plasma membrane (Somoza et al, 2003). Hence, both MSPL and hepsin may bind substrate in close proximity to the transmembrane region. However, the extracellular region of human MSPL is oriented in the opposite way with respect to that of hepsin Supplemental Data 1.

Interaction of the inhibitor decanoyl–RVKR–cmk in the active site of human MSPL

As expected, the SPD of human MSPL displays the conserved architecture of the trypsin- and chymotrypsin-like (S1A family) serine
proteases (Fig 1B). In the activated human MSPL, Ile326(16) at the cleavage site forms a salt bridge with the conserved Asp310(194) residue located immediately before the catalytic Ser511(195) residue (Fig 5A). This interaction is generated by the activating cleavage (Stubbs et al, 1998) as immediately before the catalytic Ser511(195) residue (Fig S1A). This site forms a salt bridge with the conserved Asp510(194) residue located proteases (Fig 1B). In the activated human MSPL, Ile326(16) at the cleavage hairpin loop (189-loop) (Khan & James, 1998). This and P2-Lys side chains are formed (Figs 1C and 3A). Covalent interaction between the decanoyl–RVKR–cmk inhibitor and catalytic residues (His366(57), Ser511(195)) occurs via a 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 (Gly509(193) and Ser511(195)). In addition, the guanidino group of P1-Arg forms a salt bridge with the side chain of Asp505(189), as well as a hydrogen bond with the side chain of Ser506(190). Asp505(189) is located at the bottom of the S1 pocket. These residues are highly conserved among the hepsin/TMPRSS subfamily (Fig 7). The interaction between P1-Arg and human MSPL is characteristic of trypsin-like serine proteases (Halfon & Craik, 1998). The chloromethyl group of the inhibitor irreversibly alkylates His366(57) of the SPD of human MSPL, in addition, a hemiketal if formed to Craik, 1998). The chloromethyl group of the inhibitor irreversibly alkylates His366(57) of the SPD of human MSPL, in addition, a hemiketal if formed to }

**Figure 2. Comparison of human MSPL and hepsin.** (A) Human Hepsin (coloured in marine blue) and human MSPL (coloured in cyan [LDLA], magenta [SRCR], and green [SPD]) were superposed with the SPD. The RMSD value is 0.637 Å calculated with 197 Ca atom positions. A β-sheet interaction of the N-terminus and SPD in MSPL is replaced by the C-terminus in hepsin (red arrow and close-up view in the red box). The hepsin SRCR domain is rotated by about 80° relative to that of human MSPL. (B) Human Hepsin (coloured in blue [SRCR] and pale blue [SPD]) and human MSPL (coloured in magenta [SRCR] and pale green [SPD]) were superposed with the SRCR domain. The RMSD value is 0.988 Å calculated with 59 Ca atom positions. (C) (Left) The electrostatic surface potential of the human MSPL extracellular domain. A characteristic positively-charged patch (magenta oval), composed of Arg196, Lys198, Lys218, Lys220, and Arg561, is thought to act as a contact surface for the cell membrane. The potential map is coloured from red (−5kT/e) to blue (+5kT/e). Middle: A ribbon model of human MSPL is shown with the same orientation. Right: A putative model of the membrane-anchored full-length human MSPL coloured with green (SPD), cyan (LDLA), magenta (SRCR), and orange (transmembrane domain).
The guanidino group of P4-Arg is in close proximity to the negatively charged region around the 99-loop (Glu409(97), Glu410(98), Asp411(99), and Tyr489(175)). Because these residues are unique to MSPL, the structure may explain why this enzyme shows a target preference for the P4-Arg/Lys sequence.

Comparison of the binding mechanisms of decanoyl–RVKR–cmk inhibitor to human MSPL and furin

The crystal structure of the decanoyl–RVKR–cmk inhibitor in complex with mouse furin has been determined (Henrich et al., 2003). 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 (Siezen & Leunissen, 1997). The catalytic domain of furin has a different overall fold from that of human MSPL, which belongs to the trypsin- and chymotrypsin-like (S1A family) serine protease family. Despite the different overall fold of human MSPL and furin, decanoyl–RVKR–cmk inhibits both enzymes. Therefore, we compared the structure of the human MSPL-bound decanoyl–RVKR–cmk inhibitor with that of the furin-bound inhibitor (Fig 5). Except for the P1-Arg and P4-Arg, they are not superimposed. In the human MSPL–decanoyl–RVKR–cmk complex structure, the inhibitor exhibits a bend at the P3-Val. By contrast, in the furin–decanoyl–RVKR–cmk complex structure, the inhibitor fits an extended...
conformation. As a consequence, the P1, P2, and P4 site contacts with furin, whereas the P3 side chain is directed into the solvent. As described earlier, the P3 and P4 site of decanoyl-RVKR-cmk bound to MSPL is most likely an artifact. Compared with the putative model of the MSPL-bound inhibitor, the P1-P3 site is almost identical, whereas the side chain of P4-Arg is in the opposite orientation (Fig 5). As the S4 site of furin is optimized for Arg binding, furin shows reduced affinity for the P4-Lys sequence (Henrich et al, 2003). However, the S4 site of MSPL comprises a wide negatively charged surface that is suitable for multibasic motif binding (Fig 4C).

Discussion

By the end of January 2021, the SARS-CoV-2 pandemic has killed more than 2.2 million people (https://ourworldindata.org/covid-deaths) and resulted in a worldwide recession as people were forced to socially distance. The infection of SARS-CoV-2 requires cleavage at the S1/S2 site by furin, followed by cleavage at the S2’ site by TMPRSS2 (Bestle et al, 2020; Hoffmann et al, 2020a, 2020b). TMPRSS2 displays low sequence preference for the P2-P4 position (Böttcher et al, 2006; Baron et al, 2013). The reason for TMPRSS2 mediated specific cleavage of the S2’ site is therefore unclear.

To date, the experimental structure of TMPRSS2 has not been reported. However, human MSPL shares 45% amino acid identity with TMPRSS2. Consequently, there is sufficient similarity to build a reliable homology model of human TMPRSS2 using MSPL as template (Fig 6 and Supplemental Data 1). Eight out of nine disulfide bonds are conserved (Fig 7), and the relative domain alignment of human TMPRSS2 is similar to that of MSPL. However, the SP domain, specifically at the β12-β13 loop region (60-loop in chymotrypsin), displays significant differences (Fig 6). These structural changes result in a wide substrate-binding cleft (Fig 6), so that human TMPRSS2 may be able to capture the target peptides of not only a stretched conformation but also flexible conformations.

Furthermore, Asp411, one of the important residues for P2-Lys and P4-Arg recognition found in MSPL, is replaced by Lys225 in TMPRSS2 (Figs 4B and 5). This substitution leads to a reduced negatively charged region on the S3–S5 site (Fig 6). Nonetheless, the electrostatic surface potential of the S2 site in TMPRSS2 is still negatively charged (Fig 6) and able to participate in P2-Lys binding.

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 (Wattenhofer et al, 2005; Antalis et al, 2010). One of seven missense mutants associated with the loss of hearing, D103G, was found in the LDLA domain of human TMPRSS3 (Guipponi et al, 2002; Wattenhofer et al, 2005). Because Asp103 in human TMPRSS3 corresponds to Asp222 in human MSPL, the LDLA structure stabilized by calcium-binding may be important for the function of the protein (Fig S1B). Indeed, the mutations in LDLA and SRCR (D103G, R109W, and C194F) as well as the SPD of human TMPRSS3 affect its autoactivation by proteolytic cleavage at the junction site between the SRCR and the SP domains (Guipponi et al, 2002).

In summary, we have elucidated the structure of the extracellular domain of human MSPL and its spatial arrangement of three (LDLA,
SRCR, and SP) domains, as well as the substrate sequence specificity of human MSPL. These findings will be useful in designing novel anti-influenza drugs that prevent HPAI virus uptake into the host cell. Human 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 (Zmora et al, 2014).

The mechanism of infection of SARS-CoV-2 needs to be elucidated as a matter of urgency. The MSPL structure shares the highest sequence homology to TMPRSS2 among the experimentally solved structures. The homology model presented in this article provides novel insight into TMPRSS2 function. However, it is still necessary to solve the structure of TMPRSS2.

Note added in proof

Recent studies have shown that TMPRSS13/MSPL is involved in the cleavage of the spike protein of SARS-CoV-2 as the same extent as TMPRSS2 (Hoffman et al, 2021; Kishimoto et al, 2021).

Materials and Methods

Expression and purification of human MSPL

Soluble recombinant human MSPL was generated using a previously established stable cell line expressing human MSPL (Okumura et al, 2010), which accumulated in serum-free culture medium (SFCM). It should be noted that the human MSPL we used here is a splice variant (GenBank id: BAB39741) of the canonical sequence that includes the single amino acid substitution L586V. Approximately 10 L of SFCM was concentrated by ultrafiltration using a Pellicon XL 50 (Merck-Millipore). The resulting SFCM was applied to an Anti-FLAG M2 agarose gel equilibrated in 50 mM Tris–HCl and 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 human MSPL pooled and dialyzed into PBS.

Complex formation, crystallization, and data collection

The peptide inhibitor (decanoyl–RVKR–cmk) was purchased from Merck-Millipore and reconstituted in DMSO. Human MSPL–inhibitor
complex was formed by incubating purified human MSPL (6.1 mg/ml) with a fourfold 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 precipitate. Crystallization screening was performed by mixing 1 μl of the human MSPL–inhibitor solution with 1 μl of reservoir solution using the hanging-drop vapor diffusion method. The human 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. Before data collection, the single crystal was transferred to the cryoprotectants (20% [vol/vol] glycerol and 80% [vol/vol] of the reservoir) for 5 s, and then flash-frozen in liquid nitrogen. The diffraction dataset for the human MSPL–decanoyl-RVKR-cmk crystal was collected at beamline NE3A in the Photon Factory Advanced Ring. The crystal belonged to space group P212121 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 & Murshudov, 2013). Data collection statistics are summarized in Table S1.

Structure determination and refinement of the human MSPL–inhibitor peptide complex

The structure of the complex was solved by the molecular replacement method using the program MolRep (Vagin & Teplyakov, 2010), with SPD of human plasma kallikrein (PDB entry: 2ANY; Tang et al., 2005), which shows the highest sequence identity score (46.1%), as a search model. The model of SPD was manually fixed with COOT (Emsley & Cowtan, 2004) and refined with Refmac5 (Murshudov et al., 2011). When the SPD of human MSPL was well refined, the interpretable electron density of the unmodeled region was observed, then the model of the LDLA and SRCR domains was manually built. The final model contained human MSPL, decanoyl-RVKR-cmk, 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.

Homology modelling of human TMPRSS2

The sequence alignment of the extracellular region of human MSPL and human TMPRSS2 was obtained using the BLAST web server (https://www.uniprot.org/blast/). The amino acid identity of extracellular regions between human MSPL and human TMPRSS2 was 39.8% with a score of 704, and E-value of $1 \times 10^{-86}$. The homology model of human TMPRSS2 was build using MODELLER (Šali & Blundell, 1993). Electrostatic surface potentials were calculated using the APBS server (http://server.poissonboltzmann.org/).
Data Availability

The coordinates and structure factors of the human MSPL–decanoyl–RVKR–cmk inhibitor complex have been deposited to the RCSB (PDB entry: 6KD5). The homology model of human TMPRSS2 is available from supplementary materials (Supplemental Data 1).

Supplementary Information

Supplementary Information is available at https://doi.org/10.26508/lsa.202000849.

Acknowledgements

We thank the beamline staff at the PF-AR and SPring-8 BL44XU for supporting data collection under the proposal number 2013G075 and 20156373, respectively. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI grant number 15J0096 to A Ohno, 15K13747 & 19K05966 to N Maita, 15K09585 & 18K08453 to Y Okumura, and 18H04981 & 19H04054 to T Nikawa, and by AMED-CREST grant number JP19gm0810009h0104 to T Nikawa from Japan Agency for Medical Research and Development (AMED).

Author Contributions

A Ohno: data curation, formal analysis, funding acquisition, validation, investigation, visualization, methodology, and writing—original draft.

N Maita: data curation, formal analysis, funding acquisition, validation, investigation, visualization, methodology, and writing—original draft.

T Tabata: resources, investigation, and methodology.

H Nagano: resources and investigation.

K Arita: methodology.

M Ariyoshi: validation, methodology, and writing—review and editing.

T Uchida: data curation, investigation, and methodology.

R Nakao: data curation, validation, and investigation.

A Ulla: formal analysis, investigation, and visualization.

K Sugiuira: formal analysis and investigation.

Figure 7. Multiple sequence alignment of the human MSPL extracellular region with members of the hepsin/TMPRSS subfamily. The extracellular region of human MSPL (187–586), human TMPRSS2 (110–492), human TMPRSS3 (70–454), human TMPRSS4 (55–437), and human hepsin (50–417) aligned by Clustal W program (Thompson et al, 1994), followed by colouring with ESPRIET (Gouet et al, 2003). Red asterisk indicates the catalytic triad. The amino acid sequences were obtained from UniProtKB with the id code of human MSPL (UniProt:Q9BYE2), human TMPRSS2 (UniProt:O15393), human TMPRSS3 (UniProt:P57727), human TMPRSS4 (UniProt:Q9NRS4), and human hepsin (UniProt: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. Pink arrowhead indicates the autocleavage site. Pink and cyan circles denote residues that interact with the P1 and P2 site of the decanoyl–RVKR–cmk inhibitor, respectively. Green numbers denote the disulfide pairing of human MSPL.
K Kishimoto: conceptualization, data curation, and validation.
S Teshima-Kondo: conceptualization, data curation, and supervision.
Y Okumura: conceptualization, resources, formal analysis, supervision, funding acquisition, investigation, methodology, writing—original draft, and project administration.
T Nikawa: conceptualization, resources, supervision, project administration, and writing—review and editing.

Conflict of Interest Statement
The authors declare that they have no conflict of interest.

References

Antalis TM, Bugge TH, Wu Q (2011) Membrane-anchored serine proteases in health and disease. Prog Mol Biol Transl Sci 99: 1–50. doi:10.1016/B978-0-12-385504-6.00001-4

Antalis TM, Buzza MS, Hodge KM, Hooper JD, Netzel-Arnett S (2010) The cutting edge: Membrane-anchored serine protease activities in the pericellular microenvironment. Biochem J 428: 325–346. doi:10.1042/BJ20100046

Baron J, Tarnow C, Mayoli-Nüüssle D, Schilling E, Meyer D, Hammami M, Schwalm F, Steinmetzer T, Guan Y, Garten W, et al (2013) Matriptase, HAT, and TMPRSS2 activate the hemagglutinin of H9N2 influenza A viruses. J Virol 87: 1811–1820. doi:10.1128/JVI.02320-12

Batty TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. Pathog Dis 69: 87–203. Cham, Switzerland: Springer. doi:10.1007/978-3-319-75474-1_8

Böttcher E, Matrosovich T, Beyerle M, Klenk HD, Garten W, Matrosovich M (2006) Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci Alliance 3: e202000786. doi:10.26508/lsa.202000786

Böttcher Friefertshauser E (2018) Membrane-anchored serine proteases: Host cell factors in proteolytic activation of viral gycoproteins. Activation of Viruses by Host Proteases, 153–203. Cham, Switzerland: Springer. doi:10.1007/978-3-319-75474-1_8

Böttcher Friefertshauser E, Klenk HD, Garten W (2013) Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. Pathog Dis 69: 87–100. doi:10.1128/MD-0090-2010

Bugge TH, Antalis TM, Wu Q (2009) Type II transmembrane serine proteases. J Biol Chem 284: 23177–23181. doi:10.1074/jbc.R109.021006

Bulach D, Halpin R, Spiro D, Pomery L, Janies D, Boyle DB (2010) Molecular analysis of H7 avian influenza viruses from Australia and New Zealand: Genetic diversity and relationships from 1976 to 2007. J Virol 84: 9957–9966. doi:10.1128/JVI.00930-10

Chapia C, Kobasa D, Bertram S, Glowacka I, Steffen I, Tsegaye TS, Takeda M, Bugge TH, Kim S, Park Y, et al (2009) Proteolytic activation of the 1918 influenza virus hemagglutinin. J Virol 83: 3200–3211. doi:10.1128/JVI.02205-08

Daly NL, Scanlon MJ, Djordjevic JT, Kroon PA, Smith R (1995) Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. Proc Natl Acad Sci U S A 92: 6334–6338. doi:10.1073/pnas.92.14.6334

Debeia M, Goettig P, Magdolen V, Huber R, Schechter NM, Bode W (2007) Structural basis of the zinc inhibition of human tissue kallikrein 5. J Mol Biol 373: 1017–1031. doi:10.1016/j.jmb.2007.08.042

Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132. doi:10.1107/S0907444904019158

Evans PR, Murshudov GN (2013) How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69: 1204–1214. doi:10.1107/S0907444913000061

Gouet P, Robert X, Courcelle E (2003) EPRipt/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucl Acids Res 31: 3320–3323. doi:10.1093/nar/gkg556

Guipponi M, Vuagniaux G, Wattenhofer M, Shibuya K, Vazquez M, Dougherty L, Scamuffa N, Guida E, Okui M, Rossier C, et al (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. doi:10.1093/hmg/11.23.2829

Halfon S, Craik CS (1998) Family S1 of trypsin (clan SA). In Handbook of Proteolytic Enzymes. Barrett AJ, Rawlings ND, Woessner JF (eds.). pp 5–12. San Diego: Academic Press.

Hamilton BS, Whitkatter GR, Daniel S (2012) Influenza virus-mediated membrane fusion: Determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. Viruses 4: 1144–1168. doi:10.3390/v4071144

Hashimoto T, Kato M, Shimomura T, Kitamura N (2010) TMPRSS3, a type II transmembrane serine protease, is inhibited by hepatitis growth factor activator inhibitor type 1 and activators pro-hepatocyte growth factor. FEBS J 277: 4888–4900. doi:10.1111/j.1742-4658.2010.07894.x

Henrich S, Cameron A, Boureunkov GP, Kiefersauer R, Huber R, Lindberg I, Bode W, Than ME (2003) The crystal structure of the proprotein processing protease furin explains its stringent specificity. Not Struct Biol 10: 520–526. doi:10.1083/nsh541

Herter S, Piper DE, Aaron W, Gabrielle T, Cutler G, Cao P, Bhatt AS, Choe Y, Craik CS, Walker N, et al (2005) Hepatocyte growth factor activator inhibitor type 1 and activates pro-hepatocyte growth factor. FEBS J 277: 4888–4900. doi:10.1111/j.1742-4658.2010.07894.x

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schergens TS, Herrler G, Wu NH, Nitsche A, et al (2020a) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181: 271–280. doi:10.1016/j.cell.2020.02.052

Hoffmann M, Kleine-Weber H, Pöhlmann M (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. doi:10.1016/j.molcel.2020.04.022

Hoffmann M, Hofmann-Winkler H, Smith JC, Krüger N, Arora P, Sørensen LK, Søgaard OS, Hasselstrøm JB, Winkler M, Hempel T, et al (2021) Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. EBioMedicine 65: 103255. doi:10.1016/j.ebiom.2021.103255

Horimoto T, Nakayama K, Smekens SP, Kawacca Y (1994) Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses. J Virol 68: 6074–6078. doi:10.1128/JVI.68.12.6074–6078

Khan AR, James MNG (1998) Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. Prot Sci 7: 815–836. doi:10.1002/pro.5560070401

Kido H, Okumura Y (2008) MSPL/TMPRSS13. Front Biosci 13: 754–758. doi:10.2741/2717

Kido H, Okumura Y, Takahashi E, Han HY, Wang S, Chida J, Le TQ, Yano M (2009) Host envelope glycoprotein processing proteases are indispensable for entry into human cells by seasonal and highly pathogenic avian
influenza viruses. J Mol Genet Med 3: 167–175. doi:10.4172/1747-0862.1000029

Kim DR, Sharrmin S, Inoue M, Kido H (2001) Cloning and expression of novel mosquito serine proteases with and without a transmembrane domain from human lung. Biochim Biophys Acta 1518: 204–209. doi:10.1016/S0167-4888(01)00814-1

Kishimoto M, Uemura K, Sanaki T, Sato A, Hall WW, Kariwa H, Orba Y, Sawa H, Sasaki M (2021) TMPRSS11D and TMPRSS13 activate the SARS-CoV-2 spike protein. Viruses 13: 384. doi:10.3390/v13030384

Kitamoto Y, Yuan X, Wu Q, McCourt DW, Sadler JE (1994) Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinct assortment of domains. Proc Natl Acad Sci U S A 91: 7588–7592. doi:10.1073/pnas.91.16.7588

Kyriileis OJ, Huber R, Ong E, Oehler R, Hunter M, Madison EL, Jacob U (2007) Structure of MSPL. EMBO J 26: 2998–3008. doi:10.1038/sj.emboj.7601933

Laskowski RA, Swindells MB (2011) LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51: 2778–2786. doi:10.1021/ci200227u

Lee CW, Swayne DE, Linares JA, Senne DA, Suarez D (2005) H5N2 avian influenza outbreak in Texas in 2004: The first highly pathogenic avian influenza in the United States in 20 years? J Virol 79: 11412–11421. doi:10.1128/JVI.01128-04

Lu D, Fütterer K, Korolev S, Zheng X, Tan K, Waksman G, Sadler JE (2007) Crystal structure of enteroproteidase light chain complexed with an analog of the trypsinogen activation peptide. J Mol Biol 367: 292–361. doi:10.1006/jmbi.2007.11.013

Murray AS, Varela FA, Hyland TE, Schoenbeck AJ, White JM, Tanabe LM, Todi SV, List K (2017) Phosphorylation of the type II transmembrane serine protease, DESC1, in hepatocyte growth factor activator inhibitor-1 inactivation in vivo. J Biol Chem 292: 235–245. doi:10.1074/jbc.M117.775999

Murshudov GN, Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 66: 100–120. doi:10.1107/S090744491001314

Ojala JR, Pikkarainen T, Tuuttila A, Sandalova T, Tryggvason K (2007) Crystal structure of the cysteine-rich domain of scavenger receptor cysteine-rich (SRCR) domain. Structure 11: 1123–1131. doi:10.1016/S0969-2126(03)00148-S

Okumura Y, Takahashi E, Yano M, Ohuchi M, Daidoji T, Nakaya T, Böttcher E, Garten W (1992) Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endopeptidase. EMBO J 11: 2407–2414. doi:10.1093/emboj/11.12.2407

Ojala JR, Pikkarainen T, Tuuttila A, Sandalova T, Tryggvason K (2007) Crystal structure of the cysteine-rich domain of scavenger receptor cysteine-rich (SRCR) domain. Structure 11: 1123–1131. doi:10.1016/S0969-2126(03)00148-S

Sarrias MR, Grønlund J, Padilla O, Madsen J, Holmsovk 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. doi:10.1615/CritRevImmunol.v24.i1.10

Shi W, Fan W, Bai J, Tang Y, Wang L, Jiang Y, Tang L, Liu M, Cui W, Xu Y, et al (2017) TMPRSS2 and MSPL facilitate trypsin-independent porcine epidemic diarrhea virus replication in vero cells. Viruses 9: 114. doi:10.3390/v9050114

Siezen RJ, Leunissen JA (1997) Subtilases: The superfamily of subtilisin-like serine proteases. Protein Sci 6: 501–523. doi:10.1021/pro.650006301

Simeonov P, Zahn M, Sträter N, Zuchner T (2012) Crystal structure of a supercharged variant of the human enteroproteinidase light chain. Proteins 80: 1907–1910. doi:10.1002/prot.24084

Somoza JR, Ho JD, Luong C, Ghate M, Sprengeler PJ, Mortara K, Shadrer S, Sperandio D, Chan H, McGarth ME, et al (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. doi:10.1016/S0969-2126(03)00148-S

Szabo R, Bugge TH (2008) Type II transmembrane serine proteases in development and disease. Int J Biochem Cell Biol 40: 1297–1316. doi:10.1016/j.biocel.2007.11.013

Szabo R, Bugge TH (2011) Membrane-anchored serine proteases in vertebrate cell and developmental biology. Annu Rev Cell Dev Biol 27: 213–235. doi:10.1146/annurev-cellbio-092910-154247

Tang J, Yu CL, Williams SR, Jeffery D, Sprengeler PA, Estevez A, Sampang J, Shadrer W, Spencer J, et al (2005) Expression, crystallization, and three-dimensional structure of the catalytic domain of human plasma kallikrein. J Biol Chem 280: 41077–41089. doi:10.1074/jbc.M506766200

Thomas G (2002) Furin at the cutting edge: From protein traffic to embryogenesis and disease. Nat Rev Mol Cell Biol 3: 753–766. doi:10.1038/nrm934

Thompson JD, Higgins DG, Gibson TJ (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: 4683–4688. doi:10.1093/nar/22.22.4683

Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr 66: 22–25. doi:10.1107/S0907444909042589

Wattenhofer M, Sahin-Calapoglu N, Andreassen D, Kalay E, Ceylan R, Bräillard B, Fowler-Jaeger N, Reynold A, Rossier BC, Karaguzel A, et al (2005) A novel TMPRSS3 missense mutation in a DFNB8/10 family prevents proteolytic activation of the protein. Hum Genet 117: 528–535. doi:10.1007/s00439-005-1322-x

Zmora P, Blazejewska P, Moldenhauer AS, 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. doi:10.1128/JVI.02472-14

License: This article is available under a Creative Commons License Attribution 4.0 International, as described at https://creativecommons.org/licenses/by/4.0/.