Catalytic Domain Structures of MT-SP1/Matriptase, a Matrix-degrading Transmembrane Serine Proteinase*

Rainer Friedrich‡, Pablo Fuentes-Prior‡, Edgar Ong§, Gary Coombs§, Michael Hunter§, Ryan Oehler§, Diane Pierson§, Richard Gonzalez§, Robert Huber‡, Wolfram Bode‡, and Edwin L. Madison§

From the §Max-Planck-Institut für Biochemie, Abteilung Strukturforschung, Am Klopferspitz 18a, 82152 Martinsried, Germany and §Corvas International, San Diego, California 92121

The type II transmembrane multidomain serine proteinase MT-SP1/matriptase is highly expressed in many human cancer-derived cell lines and has been implicated in extracellular matrix re-modeling, tumor growth, and metastasis. We have expressed the catalytic domain of MT-SP1 and solved the crystal structures of complexes with benzamidine at 1.3 Å and bovine pancreatic trypsin inhibitor at 2.9 Å. MT-SP1 exhibits a trypsin-like serine proteinase fold, featuring a unique nine-residue 60-insertion loop that influences interactions with protein substrates. The structure discloses a trypsin-like S1 pocket, a small hydrophobic S2 subsite, and an open negatively charged S4 cavity that favors the binding of basic P3/P4 residues. A complementary charge pattern on the surface opposite the active site cleft suggests a distinct docking of the preceding low density lipoprotein receptor class A domain. The benzamidine crystals possess a freely accessible active site and are hence well suited for soaking small molecules, facilitating the improvement of inhibitors. The crystal structure of the MT-SP1 complex with bovine pancreatic trypsin inhibitor serves as a model for hepatocyte growth factor activator inhibitor 1, the physiological inhibitor of MT-SP1, and suggests determinants for the substrate specificity.

The activity of proteolytic enzymes is required at multiple stages during the growth, invasion, and progression of human tumors (for a review, see Ref. 1). For example, these complex processes entail extensive re-modeling of the extracellular matrix as well as the activation of latent growth factors and pro-angiogenic proteins. Consequently, the high level expression of particular proteinases often correlates with poor patient survival for several different cancers (see, for instance, Ref. 2). Among these cancer-associated enzymes are serine proteinases such as urokinase-type plasminogen activator (uPA), elastase, plasmin, and cathepsin G; matrix metalloproteinases including gelatinases, interstitial collagenases, stromelysins, matrilysin, and membrane-type metalloproteinases; and lysosomal cysteine proteinases such as cathepsin B. Recently, several members of an important, emerging subfamily of serine proteinases, the type II transmembrane serine proteinases (TTSPs; reviewed in Ref. 3), have also been implicated in tumor growth and progression.

MT-SP1 (matriptase/TADG-15/i suppressor of tumorigenicity 14; EC 3.4.21) was first isolated by Shi et al. (4) as a novel proteinase that was expressed by human breast cancer cells. The enzyme was initially assigned as a gelatinase, because of its gelatinolytic properties and gelatinase-like molecular weight. However, isolation and sequencing of the cDNA revealed a 683-residue multidomain proteinase with a C-terminal serine proteinase domain. The enzyme was then named matriptase to emphasize its matrix degrading properties and trypsin-like specifity (5). Independently, Takeuchi and co-workers (6) cloned and characterized a type-II membrane-bound trypsin-like serine proteinase from a human prostate cancer cell line, which they called membrane-type serine proteinase 1, MT-SP1. This 855-residue proteinase contained two tandem repeats of the complement component C1r/s domain (CUB, derived from complement factor/1r-urchin embryonic growth factor/bone morphogenetic protein) and four tandem repeats of the low density lipoprotein receptor (LDLR) class A domain between the N-terminal transmembrane signal anchor and the C-terminal catalytic domain (5, 6). Because the matriptase sequence reported by Lin turned out to be part of the translated MT-SP1 cDNA sequence, matriptase is likely to be a form of MT-SP1 produced by ectodomain shedding (7). Alternatively, the two cDNAs may result from alternative splicing. MT-SP1 is highly expressed in prostate, breast, and colorectal cancers in vitro and in vivo (8), and inhibition of this enzyme suppresses both primary tumor growth and metastasis in a rat model of prostate cancer (5, 6). A mouse homologue was cloned by another group and called epithin (9).

The substrate specificity of MT-SP1 has been mapped using a positional scanning synthetic combinatorial library and substrate phase display (10). The preferred cleavage sequences contained Arg/Lys at P4 and basic residues or Glu at P3, small residues at P2, Arg or Lys at P1, and Ala at P1. This specificity profile corresponds well to the cleavage sequences of recognized surface localized protein substrates of MT-SP1 such as the proteinase-activated receptor-2 (PAR2), single-chain uPA

Received for publication, October 11, 2001
Published, JBC Papers in Press, November 5, 2001, DOI 10.1074/jbc.M109830200

© 2002 by The American Society for Biochemistry and Molecular Biology, Inc.
(sc-uPA), the proform of MT-SP1, and the hepatocyte growth (scattering) factor, which have been shown in vitro and/or in vivo to be efficiently activated by MT-SP1 (10, 11).

Although human breast cancer cells produce MT-SP1 primarily as the free enzyme, in human milk and normal tissues the enzyme is found in complex with an inhibitor called hepatocyte growth factor activator inhibitor 1 (HAI-1; Ref. 12). This membrane-bound inhibitor was originally isolated from human stomach carcinoma cells (13) as a 478 residue glycoprotein containing two Kunitz-type domains separated by an LDLR domain and followed by a transmembrane segment, but has been subsequently detected in several tissues. Soluble, presumably proteolytically cleaved forms of HAI-1, lacking the C-terminal hydrophobic domain, have also been reported (14). In addition to HAI-1, a smaller inhibitor (HAI-2) has been identified and characterized, which lacks the LDLR domain separating the two Kunitz-type modules in HAI-1. Site-directed mutagenesis studies suggested that the first Kunitz domain of HAI-2 is responsible for the inhibitory activity toward hepatocyte growth factor activator (15).

We have expressed and purified the catalytic domain of human MT-SP1 (MT-SP1(cd)), and have crystallized and solved the high resolution x-ray crystal structure of this enzyme in the presence of benzamidine (Bz). Because recombinant HAI-1 was not available, we also determined the structure of the MT-SP1 complex with the Kunitz-type bovine pancreatic trypsin inhibitor (BPTI), which shares a 36% sequence identity with the first Kunitz domain of HAI-1 and is a nanomolar range inhibitor of MT-SP1. These crystal structures provide important new insights into the molecular determinants of the unique specificity of MT-SP1 not obtainable from modeling (16), and give hints about the interaction of this protease with the physiological inhibitor HAI-1. This information is expected to facilitate the design of potent, selective small molecule inhibitors of MT-SP1 that may yield lead compounds for the development of novel anti-cancer agents. Because of the high accessibility of the substrate binding site, the MT-SP1(cd) crystals are well suited for soaking of small molecule inhibitors facilitating the further elaboration of initial lead compounds.

**Experimental Procedures**

**Cloning and Purification**—The human prostate adenocarcinoma cell line, PC-3, was purchased from ATCC (CRL-1435). PC-3 cells were lysed in Trizol reagent (Invitrogen, Carlsbad, CA) and total RNA was isolated according to the manufacturer’s protocol. Poly(A)+ RNAs were purified using oligo(dT) beads (Oligotex; Qiagen, Valencia, CA) and subsequently converted to single-stranded cDNAs by reverse transcription using a gene-specific primer, 5′-CACCCCTTCTTTACGACTC-3′, and purified into pCR2.1-TOPO (Invitrogen). Single-stranded cDNAs from PC-3 cell RNA were subjected to PCR with sense and antisense degenerate oligonucleotide primers (sense primer, 5′-TGGTGTIV1VTUVSV1GC1DIR1RIAYTG3′; antisense primer, 5′-1G1G1G1C1C1C1S11S11S11W1T1R1C1A1G1H1T1C1-3′, where R = A or G; V = A, G, or C; W = A or T; S = G or C; Y = C or T; and H, A, or T). The primer sequences corresponded to two highly conserved regions in chymotrypsin-like serine proteinases. PCR products were purified using a gel extraction kit (Qiagen gel extraction kit; Qiagen), ligated into pCR2.1-TOPO (Invitrogen), and transformed into Escherichia coli TOP10 cells (Invitrogen). To obtain additional MT-SP1 cDNA sequences, both rapid amplification of cDNA ends and gene-specific amplification reactions were performed. A human prostate Marathon-Ready cDNA (CLONTECH, Palo Alto, CA) was used to isolate part of the cDNA encoding MT-SP1. The 3′ region of MT-SP1 cDNA was successfully obtained by PCR amplification of cDNA ends reaction using a gene-specific primer, 5′-5CICACCCCTTCTTTACGACTC-3′ for the sense primer and 5′-5GAGCTTCTTGCGAGCTGTCGTTG-3′ for the anti-sense primer. These fragments were subcloned into pCR2.1-TOPO.

After transformation into *E. coli* cells, the insert DNAs were characterized by Southern blot analysis (using the internal cDNA fragment as probe) and by DNA sequence analysis. To obtain a cDNA encoding the entire protease domain of MT-SP1, an end-to-end PCR amplification reaction was performed using primers 5′-TCTCCTGGGACAGAAAT-3′ and 5′-GGGACAGCGTGGGACAGTA-G3′ for the 5′ end and 5′-ATTCCGGCCGCGCTATACCCCGAGTTCTTCTTGTGCAA3′ for the 3′ end. An 800-bp DNA fragment was amplified, purified, digested with XhoI and NotI, and subcloned into the *Pichia pastoris* expression vector, pPIC9KX. Transformation was performed in a Bio-Rad GenePulser II (voltage = 1500 V, capacity = 50 μF, and resistance = 200 ohms). The screening of transformed *Pichia* clones by expression was performed by testing clones with Spectrozyme t-PA (CH₂SO₄-D-HHT-Gly-Ar-pNA.HCl; American Diagnostica).

The production of multimi/migenuous amounts of MT-SP1 was carried out by fermentation in a BioFlo 3000 fermentor (New Brunswick Scientific, NJ) using a SMD1168/pPIC9K-MT-SP1 Sac SC1 clone. The medium was inoculated with an overnight culture of the *P. pastoris* transformant. Cells and cell debris were removed by centrifugation, the supernatant was concentrated, and the buffer was exchanged into 50 mM Tris-HCl, 50 mM NaCl, 0.05% Tween 80, pH 8.0 (buffer A). The concentrated MT-SP1-containing solution was applied onto a 150-μl benzamidine column equilibrated with buffer A, and the column was washed with 50 mM Tris-HCl, 1.0 mM NaCl, 0.05% Tween 80, pH 8.0 (buffer B) and eluted with 50 mM Tris-HCl, 1.0 mM NaCl, 0.05% Tween 80, pH 8.0 (buffer C). Fractions containing MT-SP1 activity were pooled and concentrated to 50 mM NaHPO₄, 125 mM NaCl, pH 5.5 (buffer D), and the partially purified MT-SP1 was passed through a Q-Sepharose Fast Flow Hitrap column (Amersham Biosciences, Inc.) pre-equilibrated with buffer D. The flow-through was collected, and the protein concentration was determined by measurement of *A*₅₀₀ using an extinction coefficient of 2.012 mg/ml *A*₅₀₀. Purified MT-SP1 was deglycosylated with endoglycosidase H (Prozyme, 5 units/ml) and further purified on an Akta Explorer system using a 7-ml Source 15Q anion exchange column (Amersham Biosciences, Inc.). The protein was eluted in a buffer containing 50 mM HEPES, pH 6.5, with a 0–0.33 M NaCl gradient. Fractions containing protein were pooled, and benzamidine was added to a final concentration of 10 mM. The protein purity was examined by SDS-PAGE, and the protein concentration was determined by measurement of *A*₂₈₀.

**Crystallographic Refinement**—Plate-like crystals of the Bz-MT-SP1 complex were grown from 0.1 M Tris-HCl, pH 8.0, 1.5 M ammonium sulfate, 3% ethanol at 18 °C using the hanging drop vapor diffusion technique. These crystals belong to the orthorhombic space group C222₁ with two MT-SP1 molecules in the asymmetric unit. The best crystals used for data collection were obtained from a 1.3 Å resolution, and have one molecule in the asymmetric unit. The Bz-MT-SP1 crystals were transferred to 0.1 M Tris-HCl, pH 8.0, 1.5 M ammonium sulfate, 23% glycerol. A complete native data set to 1.3-Å resolution was collected from a single crystal under a nitrogen stream at 100 K using synchrotron radiation and a CCD system (MAR Research, Hamburg, Germany) at DESY, Hamburg, Germany. The crystals were evaluated with the MOSFLM package (43) and loaded and scaled using SCALA from the CCP4 program suite (17). For the determination of the orientation and position of the MT-SP1 molecules in the crystals, rotational and translational searches were performed with AMoRe (18) using data from 20- to 3.5 Å resolution and a modified enteropeptidase search model (19) with all nonidentical residues reduced to Ala. A unique solution was found with a correlation factor of 39.4% and an *R*-factor of 46.2%; the corresponding values of the next best solution were 12.8 and 56.3%, respectively. Crystallographic refinement was done in several cycles consisting of model building performed with MAIN (20) and conjugate gradient minimization and simulated annealing using CNS (21). The target parameters of Engh and Huber (22) were used. This procedure converged rapidly, yielding a model with excellent parameters (see Table I). In the final model building/refinement cycles, water molecules were inserted at stereochemically reasonable sites, and individual restrained atomic *B*-values were refined. 5.1% of all reflections were omitted from the refinement to calculate the *R* *free*; the final *R* and *R* *free* were 18.4 and 19.3%, respectively, for all data to 1.3 Å. The whole main chain of the MT-SP1 catalytic domain is in appropriate electron density. Only a few side chains were found to be partially occupied in the electron density; the occupancy of all undefined atoms was set to zero. Modeling was performed interactively using MAIN; these models were energy refined with CNS.

Crystals of the complex were grown from 0.1 M HEPES, pH 6.5, 20% polyethylene glycol 4000, 2% CaCl₂ at 18 °C using the hanging drop vapor diffusion technique. These crystals belong to the triclinic space
TABLE I
Crystallographic data

| Inhibitor  | Bz | BPTI |
|-----------|----|------|
| Data collection |      |      |
| Space group | C222 | P1 |
| Molecules in asymmetric unit | 1 | 2 |
| Cell constants |      |      |
| a | 66.92 Å | 47.10 Å |
| b | 141.60 Å | 54.23 Å |
| c | 51.94 Å | 67.82 Å |
| α | 90.00° | 107.62° |
| β | 90.00° | 96.86° |
| γ | 90.00° | 103.36° |
| Limiting resolution | 1.39 Å | 2.93 Å |
| Rmerge, overall | 8.6% | 12.0% |
| Rmerge, outermost shell | 29.4% | 34.3% |
| Unique reflections | 58,805 | 12,200 |
| Completeness, overall | 96.2% | 94.4% |
| Completeness, outermost shell | 93.5% | 93.1% |
| Non-hydrogen protein atoms | 1864 | 4618 |
| Heterogen atoms | 14 | 0 |
| Solvent atoms | 381 | 80 |
| Refinement statistics |      |      |
| Test set size | 5.1% | 8.1% |
| Resolution range | 18.0–1.30 Å | 12.0–2.93 Å |
| Completeness, overall | 96.2% | 94.4% |
| Completeness, outermost shell | 91.7% | 93.1% |
| R | 18.4% | 19.9% |
| R, outermost shell | 23.7% | 30.0% |
| Rmerge | 19.3% | 27.9% |
| Root mean square deviation | 25.0% | 40.8% |
| Bond length | 0.010 Å | 0.007 Å |
| Bond angles | 1.6° | 1.4° |
| Average B-value/S.D. | 13.7 Å² | 14.4 Å² |
| Ramachandran plot |      |      |
| Most favored region | 89.3% | 74.3% |
| Allowed region | 10.7% | 23.5% |
| Generously allowed region | 2.3% |      |

RESULTS

Overall Structure of the MT-SP1 Catalytic Domain—Our cloning, expression, and purification procedure described under “Experimental Procedures” yielded millimilligram quantities of highly purified MT-SP1(cd), which in the presence of Bz formed crystals diffracting to beyond 1.3-Å resolution. MT-SP1(cd) resembles an oblate ellipsoid with diameters of 35 and 50 Å. Similar to other trypsin-like serine proteinases, the chain starts with Val16 (corresponding to the catalytic domain of another membrane-type serine proteinase, enteropeptidase/enterokinase (Fig. 1B). 222 Ca atoms of topologically equivalent residues are found within a 2.0-Å distance, corresponding to an root mean square deviation of 0.70 Å, with 109 of these topologically equivalent residues being identical. The next best fit, with a 0.73-Å root mean square deviation for 212 Ca atoms, is observed with bovine trypsin (25), followed by bovine chymotrypsin (26) and human thrombin (27). The topological equivalence with chymotrypsinogen (ogen) formed the basis for the sequence alignment and the chymotrypsinogen numbering of the MT-SP1 catalytic domain used in this paper (Fig. 2). This alignment requires a nine-residue insertion between residues Ile60 and Pro61, single-residue insertions behind residues Gly184, Glu186, Ala204, and Ala221, and single-residue deletions at positions 149 and 218. Inserted residues are marked by suffixes following the residue number of the preceding common residue.

Loops Surrounding the Active Site Cleft—The narrow substrate specificity of MT-SP1 arises from unique structural determinants of its active site cleft that is shaped in part by the surrounding surface loops. In the following brief description, these loops (defined according to the residue number of their central residue) will be addressed in an anticlockwise manner with respect to the standard orientation displayed in Fig. 1A.

To the east of the catalytic triad, the rigid 37 loop projects out of the molecular surface of MT-SP1. This loop contains two of the "zymogen triad" residues, Ser42 and His46, which, together with Asp194, would stabilize the inactive zymogen-like conformation of the proenzyme (28, 29). Around Gin58, this loop deviates markedly from the path followed in most other chymotrypsin-like proteinases (Fig. 1B). This conformation seems to be mainly stabilized by the Gin58 side chain, which is held in an exposed position via hydrogen bonds made through its terminal carboxamide group with Tyr60(G) Oγ and with the π-electron system of the Phe60(E) phenyl ring.

The most striking feature of MT-SP1 is the unusually large 60 insertion loop, which is of the same length (nine additional residues) and exhibits a similar β-hairpin conformation as the corresponding loop in thrombin (27), but is oriented differently (Fig. 1B). The eight residues of MT-SP1 between Tyr59 and...
Ser^{60}(H) form a protruding irregular β-hairpin loop, which is stabilized through internal main chain-side chain hydrogen bonds made by the Asp^{60}(A) and Asp^{60}(B) carboxylate groups. In thrombin, the rigid 60 loop partially occludes the active site and hence contributes to the narrow substrate specificity (27). In MT-SP1, the loop is rotated away from the active site creating a more open cavity than in thrombin. Unlike the basic side chains of Lys^{60}(F) in thrombin or Arg^{60}(G) in enteropeptidase (19), the Arg^{60}(F) side chain of MT-SP1 is directed away from the active site cleft and located between the side chains of Asp^{60}(A) and Asp^{60}(B).

To the north of the active site, the 99 loop of MT-SP1 protrudes from the molecular surface and forms a "roof" on top of the active site cleft. Noteworthy is the clustering of aromatic side chains around this loop. The benzyl side chain of Phe^{99}, which protects the hydrogen bonding interaction between the

---

**Fig. 1.** Ribbon stereo plots of MT-SP1(cd) shown in standard orientation. Figure was prepared with SETOR (38). A, the fold, secondary structure elements, and loops of MT-SP1. Helices and β-strands are given as red helices and blue strands. The active site residues, Asp^{189}, and the disulfide bridges (yellow) are shown as stick models. B, MT-SP1 (red) superimposed with the enteropeptidase catalytic domain (green; Ref. 19), bovine chymotrypsin A (yellow; Ref. 39), and the heavy chain of human α-thrombin (blue; Ref. 27). The active site residues of MT-SP1, Asp^{189}, and the bound benzamidine are shown as stick models.

**Fig. 2.** Structure-based amino acid sequence alignment of the human MT-SP1 catalytic domain with human enteropeptidase (40), human plasmin, bovine chymotrypsinogen A, bovine trypsin, and human thrombin. The numbers given correspond to the chymotrypsinogen numbering. β-Sheets and α-helices in MT-SP1 are indicated by arrows and cylinders, respectively. Figure was made with ALSCRIPT (41).
catalytic residues Asp<sup>102</sup> and His<sup>57</sup> from the solvent, delimits the S2 subsite. Unique to MT-SP1 is the well defined benzyl side chain of Phe<sup>97</sup>, which extends away from the apex of the 99 loop partially shielding the circularly arranged carbonyl groups of Asp<sup>96</sup>, Phe<sup>97</sup>, and Thr<sup>98</sup> (see Fig. 3).

The “autolysis” or 145 loop of MT-SP1 is one residue shorter and less exposed than the corresponding segment of chymotrypsin. This loop encircles the extended side chain of His<sup>143</sup> and forms the southern boundary or “floor” of the active site cleft (Fig. 4). The presence of Gly<sup>151</sup> in this loop is noteworthy, because this would allow for the accommodation of bulkier P2<sup>103</sup> side chains of peptide substrates.

In the pancreatic serine proteinases, the 70–80 loop forms the calcium binding site, with the carboxylates of Glu<sup>70</sup> and Glu<sup>80</sup> coordinating the calcium ion. In MT-SP1, both positions are occupied by aliphatic residues (Leu<sup>70</sup> and Val<sup>80</sup>), hence rendering the catalytic domain calcium-independent. Together with the proximal side chain parts of Arg<sup>76</sup> and Ala<sup>77(A)</sup>, the hydrophobic side chains at positions 70 and 80 clamp both loop ends together in a manner similar to that for the ion metal in the calcium-containing serine proteases.

**Active Site and Substrate Binding Sites of MT-SP1**—The residues of the active site triad, Ser<sup>195</sup>, His<sup>57</sup>, Asp<sup>102</sup>, and other catalytic elements such as the oxyanion hole created by the main chain nitrogens of Gly<sup>193</sup> and Ser<sup>195</sup> are arranged in the active site cleft exactly as in trypsin and chymotrypsin. The specificity pocket S1, which opens to the west of Ser<sup>195</sup>, is bordered by segments Val<sup>213</sup>–Gly<sup>220</sup>, Ser<sup>190</sup>–Ser<sup>195</sup>, Pro<sup>225</sup>–Tyr<sup>228</sup>, and the Cys<sup>191</sup>–Cys<sup>220</sup> disulfide bridge. The side chains of all MT-SP1 residues lining the interior of this pocket with their side chains are virtually superimposable with the corresponding residues in trypsin. The specificity for basic residues, but also to Ser<sup>190</sup>, making Lys<sub>P1</sub> residues equally acceptable as Arg, and residues forming the inner wall (Gly<sup>226</sup>, Tyr<sup>228</sup>, and Val<sup>214</sup>). The phenyl group of the bound benzamidine molecule is sandwiched between the parallel peptide groups Trp<sup>215</sup>–Gly<sup>216</sup> and Cys<sup>191</sup>–Gln<sup>192</sup>, whereas its amido group opposes the Asp<sup>189</sup> carboxylate at the bottom of the pocket forming a two-O/two-N salt bridge, with the distal ni-
Asp217 and His40 bridge, and bordered by the isobutyl side chain of Ile41 and the as stick models. The large rotation of the Phe99 side chain and the adaptation of the 60 loop upon binding of BPTI are clearly visible. The view is -Ala16(I) and the P3 /H11032

MT-SP1, with its guanidyl group stacking between the Phe97 side chain. On the primed side of the reactive center loop, the Phe99 benzyl group away from this site, in this way enlarging the S2 cavity and making it accessible for the disulfide bridge (Fig. 5). The Phe99 side chain, if positioned as observed in Bz-MT-SP1, would clash with the Cys14(I)-Cys38(I) disulfide bridge. In this respect it strongly differs from the also exposed but much more rigid thrombin 60 loop, which delimits the S2 cavity and normally prevents access of bulkier substrates and protein inhibitors (30). The 60 loop of MT-SP1, in contrast, does not impair BPTI binding, but instead strengthens it by forming a number of additional favorable interactions (Figs. 5 and 6).

Interaction with Kunitz-type Inhibitors—Because the prototypic Kunitz-type inhibitor BPTI is closely related to the first Kunitz domain of HAI-1 at the sequence level, we crystallized the complex of BPTI with MT-SP1 and solved its structure at 2.9-Å resolution (Figs. 5 and 6). BPTI docks into the concave substrate binding surface of MT-SP1 via the reactive site loop (Thr11(I) to Ile18(I), with the Lys15(I) –Ala16(I) scissile bond) and the secondary binding segment (Gly36(I) to Arg39(I)), similar to the prototypical trypsin complex (23). The reactive site loop of BPTI runs anti-parallel to MT-SP1 segments Ser214-Pro13(I), which is typical for Kunitz-type inhibitors (30). The 60 loop of MT-SP1, in contrast, does not allow for instance the formation of a salt bridge between Asp60(B) (MT-SP1) and Arg20(I), and of charged hydrogen bonds between Arg60(C) and the carbonyl groups of Lys34(I) and Asn44(I). The well defined 60 loop of MT-SP1 component showing negative (red) and positive (blue) surface potential (as above). Some surface residues and the thrombin-like anion binding exosite I are indicated by labels. The complex is rotated about around a vertical axis for 90° compared with the standard orientation in Fig. 1. Figure was made with WEBLABVIEWER (www.msi.com).

Fig. 5. Section of the complex formed between BPTI (blue color) and MT-SP1 (orange) around the contact interface, superimposed with parts of the Bz-MT-SP1 structure (red). The active site residues and Asp189 (green), the benzyl groups of Phe97 and Phe99 and Asp60(B) and Arg60(C) from the 60 loop of both MT-SP1 structures, as well as Pro35(I), Lys15(I), Ala16(I), Arg17(I), Ile18(I), and Arg20(I) of BPTI are shown as stick models. The large rotation of the Phe99 side chain and the adaptation of the 60 loop upon binding of BPTI are clearly visible. The view is slightly tilted in comparison with Fig. 1. Figure was made with SETOR (38).

Fig. 6. Docking of BPTI (yellow) and of the (covalently linked) fourth LDLR domain (green) to opposite sites of the MT-SP1 catalytic domain. Solid surface representations, with the colors of the MT-SP1 component showing negative (red) and positive (blue) surface potential (as above). Some surface residues and the thrombin-like anion binding exosite I are indicated by labels. The complex is rotated about around a vertical axis for 90° compared with the standard orientation in Fig. 1. Figure was made with WEBLABVIEWER (www.msi.com).

DISCUSSION

The catalytic domain of human MT-SP1 reveals the overall fold of a (chymo)trypsin-like serine proteinase, but displays unique properties such as the hydrophobic/acidic S2/S4 subsites and an exposed 60 loop, which considerably affect its substrate recognition and binding properties. The MT-SP1 polypeptide fold deviates notably from that of the serine proteinase trypstat, however, which exhibits six novel surface

The characteristic protruding 60 loop of MT-SP1 provides a unique properties such as the hydrophobic/acidic S2/S4 subsites and an exposed 60 loop, which considerably affect its substrate recognition and binding properties. The MT-SP1 polypeptide fold deviates notably from that of the serine proteinase trypstat, however, which exhibits six novel surface
loops through which the tryptase monomers form the intermolecular interactions that stabilize the unusual tetramer structure (31). Thus, MT-SP1 is not closely related to trypsin, and the name “matritase” is therefore potentially confusing. However, because enterokinase and hepsin are members of the “MT-SP” family that were discovered and described before matritase, assignment of the name MT-SP1 to matritase may also produce controversy. One potential solution to these nomenclature issues would be to adopt either the “MT-SP” or the TTSP (3) terminology that was suggested previously by others, and to assign numbers to individual family members based on their date of discovery.

MT-SP1 can cleave selected synthetic substrates as effectively as trypsin, but exhibits a significantly more restricted specificity than trypsin (10). This may reflect, on the one hand, the near identity of the S1 specificity pockets of these two enzymes, presumably allowing both tight binding of substrates and optimal presentation of their scissile peptide bonds to the enzyme active site, and, on the other hand, clear structural distinctions between the extended binding sites of the two proteinases. With respect to shape and chemical composition, the S1 pocket of MT-SP1 (and trypsin) provides good complementarity for Lys as well as arginine P1 residues. The efficient accommodation of P1-Lys residues, which has been demonstrated experimentally (5, 10), is facilitated by Ser219, whose side chain provides an additional hydrogen bond acceptor to stabilize the buried α-ammonium group. P1-Arg residues are accommodated equally well, because of the overall better space filling by the guanidinium group (30).

The hydrophobic S2 groove of MT-SP1 is shaped to accommodate small to medium-sized hydrophobic side chains of P2-α-amino acids, with a wide hydrophobic exit toward the bulk solvent that would expose longer and more polar side chains. In addition, the rotation of the Phe99 benzyl group upon BPTI binding suggests that the S2 subsite of MT-SP1 is not rigid. This observation is consistent with experimental findings from positional scanning (10), which indicated that MT-SP1 accepts a broad range of amino acids in the P2 position.

Craik and co-workers (10) have reported previously that MT-SP1 exhibits a strong preference for peptide substrates that contain Arg or Lys at P4 position, and a certain preference for either Gln or basic residues at P3. This interesting specificity may be mediated by electrostatic interactions with the acidic side chains of Asp-217 and/or Asp-96 (see Fig. 4), which could favorably pre-orient specific basic peptide substrates as they approach the enzyme active site cleft. Because canonically binding substrates align antiparallel to the Ser214–Asp217 segment in an extended conformation, the side chains of P4-Arg or Lys residues presumably extend toward the 99 loop with their guanidyl or ammonium groups not reaching the Asp96 carboxylate directly but forming a charged hydrogen bond with Phe97 O. In addition, they can favorably interact with the overall negative potential created by the 96-, 97-, and 98-carbonyls, in full agreement with the preference of MT-SP1 for basic P4 residues. This MT-SP1 S4 subsite is reminiscent of the corresponding region of coagulation factor Xa, which has also been shown to function as a cation binding site (32). The Arg29(1) side chain of BPTI (see Fig. 5), although provided by the secondary binding segment of the inhibitor, is a nice example of such interactions; in the BPTI-MT-SP1 complex, the hydrogen bond between the Arg29(1) guanidyl and the Phe97 O is shielded from solvent by the unique Phe97 benzyl side chain (see Figs. 4 and 5), which would significantly strengthen these interactions.

In canonically bound peptide substrates, the side chain of a P3 residue projects out of the active site cleft, where it can hydrogen-bond the carboxamide group of Gln192. It is also possible, however, that a bound substrate could adopt a “kinked” conformation at the P3 position, as seen for Kunitz-type inhibitors (Fig. 5). In this alternative conformation, the P3 side chain extends into the S4 subsite, i.e. toward the 99 loop. To allow the guanidyl or the ammonium group of a P3-Arg or Lys residue, bound to MT-SP1 in the alternative conformation, to form direct hydrogen bonds with Phe97 O, the substrate main chain around P3 would have to rotate and thereby weaken considerably the inter-main chain hydrogen bonds to Gly216. In either conformation, however, a basic P3 side chain would interact favorably with the negative potential of the MT-SP1 S4 pocket (Fig. 4). Long range electrostatic interactions between side chain charges and complementary surface potentials are often found in protein-protein complexes. For instance, in the thrombin complexes with hirudin and other protein inhibitors, the removal of the charges of acidic residues not involved in direct salt bridges has been shown to affect the electrostatic interactions strongly (33, 34). The relatively low probability of the simultaneous occurrence of Arg/Lys residues at P3 and P4 in good MT-SP1 substrates (10) would then be explained by mutual charge compensation and exclusion from the same (S4) site.

The specificity of MT-SP1 differs substantially from that of trypsin, because MT-SP1 does not indiscriminately cleave peptide substrates at accessible Lys or Arg residues, but instead requires recognition of additional residues surrounding the scissile peptide bond. This requirement for recognition of an extended primary sequence for efficient catalysis of substrates suggests that MT-SP1 is a relatively specific proteinase that may play a regulatory role (5, 6). Recognition of an extended primary sequence appears to be also required for efficient cleavage of macromolecular substrates by MT-SP1. Efficient auto-activation of MT-SP1 entails recognition and cleavage of an Arg-Gln-Ala-Arg P4-P1 target sequence. MT-SP1 can also efficiently activate the proteinase-activated receptor-2 (PAR2), sc-uPA (10), and the hepatocyte growth factor/scatter factor (11). These extracellular surface-localized proteins display the P4 to P1 target sequences Ser-Lys-Gly-Arg, Pro-Arg-Phe-Lys, and Lys-Gln-Gly-Arg, respectively, which match closely the MT-SP1 cleavage specificity requirements observed for small peptidic substrates (10). Another indication of the substrate specificity of MT-SP1 is that the enzyme does not activate proteins closely related to these substrates, such as PAR-1, PAR-3, PAR-4, and plasminogen, that do not display target sequences matching the extended MT-SP1 specificity near the scissile bond.

Because MT-SP1 has been found co-localized with sc-uPA in several cell types, it has been suggested that MT-SP1 may be a physiologically relevant activator of this proteinase zymogen (10). uPA plays an important role in angiogenesis and/or tumor progression and can also activate both plasminogen and matrix metalloproteinases (35). The latter, in turn, play important roles in matrix degradation and re-modeling, events that are required both during angiogenesis and tumor invasion and metastasis. In addition, MT-SP1 can directly activate hepatocyte growth factor, a protein that promotes cell growth as well as angiogenesis; therefore, it may play both direct and indirect roles in cell growth and migration and, when improperly regulated, may contribute to tumor angiogenesis, growth, and progression.

In normal tissues, the proteolytic activity of proteinases is carefully controlled localizing their action in time and space. As mentioned above, HA1-1, a type-II transmembrane protein containing two Kunitz-type domains, appears to be the primary physiological inhibitor of MT-SP1 (12). Based on the structure
of our BPTI-MT-SP1 complex, we have modeled the complex between MT-SP1 and the first Kunitz domain of HAI-1, which is 36% identical to BPTI and appears more likely to exhibit high affinity for the MT-SP1 active site than the second HAI-1 Kunitz domain. This model suggests that a large number of favorable interactions could form between the first HAI-1 Kunitz domain and MT-SP1, both between the Gly\(^{22}(I)\)-Arg\(^{13}(I)\)-Cys\(^{14}(I)\)-Arg\(^{15}(I)\) \& Gly\(^{64}(I)\)-Ser\(^{17}(I)\)-Phe\(^{18}(I)\) reactive site loop (using the BPTI nomenclature, with Arg\(^{15}(I)\) \& Gly\(^{64}(I)\) representing the scissile bond) and the active site cleft, and also at secondary interaction sites such as made by the 60 loop. In the primed side, the first HAI-1 Kunitz domain possesses an uncommon P3'-Phe\(^{4}(I)\), which could, because of the small P1'- Gly residue, be nicely packed in the large hydrophobic S1'/S3' pocket of MT-SP1 (Fig. 4). The side chain of the P3-equivalent Arg\(^{13}(I)\) is expected to extend into the S4 subsite, where its guanidyl group would make favorable electrostatic interactions with the 99 loop carbonyls, similarly to those observed for Arg\(^{59}(I)\) in the BPTI complex (see Fig. 5).

The putative Gly\(^{12}(I)\)-Leu\(^{15}(I)\)-Cys\(^{14}(I)\)-Lys\(^{15}(I)\) \& Glu\(^{16}(I)\)-Ser\(^{17}(I)\)-Phe\(^{18}(I)\) reactive site loop of the second Kunitz domain of HAI-1, in contrast, does not match the reported substrate specificity of MT-SP1. The second HAI-1 domain does, however, possess a number of negatively charged residues in its C-terminal \(\alpha\)-helix that could form favorable electrostatic interactions with basic surface residues of MT-SP1 that map in or near the region corresponding to the anion binding exosite I of thrombin (such as Arg\(^{55},\) Arg\(^{58},\) Arg\(^{65},\) Lys\(^{110}\); see Fig. 6). Such an additional exosite binding of the second HAI-1 domain accompanying the interaction of the first domain with the MT-SP1 active site would considerably increase the affinity and specificity of HAI-1 for MT-SP1. A similar cooperative action with consequent increase in affinity and specificity has been previously observed in the complex between thrombin and ornithodorin, a two-domain Kunitz-type inhibitor derived from the blood-sucking tick Ornithodoros moubata (36), where the first and the second Kunitz domains interact (noncanonically, however) with the active site and electrostatically with exosite I, respectively, of thrombin.

The role of the four LDLR domains that precede the catalytic domain remains unclear, but they have been implicated in mediating interactions with other membrane or membrane-associated proteins (5, 6). The LDLR (4) domain of MT-SP1 was modeled (Fig. 6) based on the structure of the fifth low density lipoprotein class A binding domain of LDLR (37), which shares a conserved disulfide bonding pattern and a 45% amino acid identity with LDLR (4). In full-length MT-SP1, the last cysteine residue (Cys\(^{604}(g)\)) of LDLR (4) is separated by only one amino acid (Asp\(^{605}(g)\)) from the first residue (Cys\(^{608}(g)\) = Cys\(^{2}\)) of the catalytic domain, which, in turn, forms an intradomain disulfide bridge with Cys\(^{122}\). This implicates close proximity of both modules. A careful inspection of the electrostatic potentials suggests a distinct rotational orientation of LDLR (4) and the catalytic domain relative to each other, which would create both favorable interactions between electrostatic potentials of the two domains and a good steric fit of the two complementary surfaces. Our docking experiment predicts four interdomain salt bridges (Glu\(^{593}(g)\)-Arg\(^{206}\), Lys\(^{592}(g)\)-Asp\(^{125}\), Asp\(^{605}(g)\)-Arg\(^{119}\), and Asp\(^{605}(g)\)-Lys\(^{606}(g)\)) and charged hydrogen bonds from Ly\(^{239}(g)\) and Arg\(^{238}\) to the 588(g)-590(g) backbone. The LDLR (4) domain of MT-SP1 carries all four acidic side chains (of Asp\(^{598}(g)\), Asp\(^{504}(g)\), Asp\(^{600}(g)\), and Glu\(^{601}(g)\)) engaged in calcium coordination in the fifth low density lipoprotein-A domain of LDLR (37). The suggested association of LDLR (4) with the catalytic domain does not directly involve this putative calcium binding site, which would be located close to the inter-

---

**REFERENCES**

1. Mignatti, P., and Rifkin, D. B. (1993) *Physiol. Rev.* 73, 161–195
2. Kato, T. E., Grundahl-Hansen, J., Shaer, S. M., Long, R., Russell, G., and Brunner, R. (1998) *Breast Cancer Res. Treat.* 47, 9–16
3. Hooper, J. D., Clements, J. A., Quigley, J. P., and Antalis, T. M. (2001) *J. Biol. Chem.* 276, 857–860
4. Shi, Y. E., Torri, J. Y., Weisel, A., Wellstein, A., Lippman, M. E., and Dickson, R. B. (1993) *Cancer Res.* 53, 1409–1415
5. Lin, C. Y., Anders, J., Johnson, M., Sang, Q. A., and Dickson, R. B. (1999) *J. Biol. Chem.* 274, 18231–18236
6. Takeuchi, T., Shuman, M. A., and Craik, C. S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96, 11054–11061
7. Benda, C., Dickson, R. B., and Lin, C. Y. (2001) *Bull. J. Biochem.* 268, 1439–1447
8. Oberst, M., Anders, J., Xie, B., Singh, B., Ossandon, M., Johnson, M., Dickson, R. B., and Lin, C. Y. (2001) *Am. J. Pathol.* 158, 1301–1311
9. Kim, M. G., Chen, C., Lyu, M. S., Cho, E. G., Park, D., Konak, C., and Schwartz, R. E., and Vlasuk, G. (2001) *Ann. N. Y. Acad. Sci.* 928, 1–15

---
R. H. (1999) Immunogenetics 49, 420–428
10. Takeuchi, T., Harris, J. L., Huang, W., Yan, K. W., Coughlin, S. R., and Craik, C. S. (2000) J. Biol. Chem. 275, 26333–26342
11. Lee, S. L., Dickson, R. B., and Lin, C. Y. (2000) J. Biol. Chem. 275, 26333–26342
12. Lin, C. Y., Anders, J., Johnson, M., and Dickson, R. B. (1999) J. Biol. Chem. 274, 18337–1842
13. Shimomura, T., Denda, K., Kitamura, A., Kawaguchi, T., Kito, M., Kondo, J., Kagaya, S., Qin, L., Takata, H., Miyazawa, K., and Kitamura, N. (1997) J. Biol. Chem. 272, 6370–6376
14. Shimomura, T., Denda, K., Kawaguchi, T., Matsumoto, K., Miyazawa, K., and Kitamura, N. (1999) J. Biol. Chem. 274, 18237–18242
15. Qin, L., Denda, K., Shimomura, T., Kawaguchi, T., and Kitamura, N. (1998) FEBS Lett. 436, 111–114
16. Enyedy, I. J., Lee, S. L., Kuo, A. H., Dickson, R. B., Lin, C. Y., and Wang, S. (2001) J. Med. Chem. 44, 1349–1355
17. Collaborative Computational Project No. 4. (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
18. Navaza, J. (1994) Acta Crystallogr. Sect. A 50, 157–163
19. Lu, D., Futterer, K., Korolev, S., Zheng, X., Tan, K., Waksman, G., and Sadler, J. E. (1999) J. Mol. Biol. 292, 361–373
20. Turk, D. (1992) Further Development of a Program for Molecular Graphics and for the Manipulation of Electron Densities and Its Application to Various Protein Structure Determinations. Ph.D. thesis, Technische Universität, München
21. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D Biol. Crystallogr. 54, 905–921
22. Eng, H., and Huber, R. (1991) Acta Crystallogr. Sect. A 47, 392–400
23. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Detsienhofer, J., and Steigmann, W. (1974) J. Mol. Biol. 89, 73–101
24. Nienaber, V., Wang, J. Y., Davidson, D., and Henkin, J. (2000) J. Biol. Chem. 275, 7239–7248
25. Bode, W., and Schwager, P. (1975) J. Mol. Biol. 98, 693–717
26. Birktoft, J. J., and Blow, D. M. (1972) J. Mol. Biol. 68, 187–240
27. Bode, W., Turk, D., and Karshikov, A. (1992) Protein Sci. 1, 426–471
28. Bode, W., Schwager, P., and Huber, R. (1978) J. Mol. Biol. 118, 99–112
29. Madsen, E. L., Kobe, A., Gething, M. J., Sambrook, J. C., and Goldsmith, E. J. (1993) Science 262, 419–421
30. Bode, W., and Huber, R. (1992) Eur. J. Biochem. 204, 433–451
31. Pereira, P. J., Bergner, A., Macedo-Silveira, S., Huber, R., Matschiner, G., Fritz, H., Sommerhoff, C. P., and Bode, W. (1998) Nature 392, 306–311
32. Stubbs, M. T., Huber, R., and Bode, W. (1995) FEBS Lett. 375, 103–107
33. Karshikov, A., Bode, W., Tulinsky, A., and Stone, S. R. (1992) Protein Sci. 1, 727–735
34. Stubbs, M. T., and Bode, W. (1995) Trends Biochem. Sci. 20, 131
35. Andrews, P. A., Egelund, R., and Petersen, H. H. (2000) Cell. Mol. Life Sci. 57, 25–40
36. van de Locht, A., Bode, W., Huber, R., Le Bonniec, B. F., Stone, S. R., Esmon, C. T., and Stubbs, M. T. (1997) EMBO J. 16, 2977–2984
37. Fass, D., Blacklow, S., Kim, P. S., and Berger, J. M. (1997) Nature 388, 691–693
38. Evans, S. V. (1990) J. Mol. Graph. 11, 134–138
39. Blevins, R. A., and Tulinsky, A. (1985) J. Biol. Chem. 260, 4264–4275
40. Kitamoto, Y., Veile, R. A., Donis-Keller, H., and Sadler, J. E. (1995) Biochemistry 34, 4562–4569
41. Barton, G. J. (1993) Protein Eng. 6, 37–40
42. Nicholls, A., Bharadwaj, R., and Honig, B. (1993) Biophys. J. 64, A166
43. Leslie, A. (1991) in Crystal Computing V (Moras, D., Podjarny, A. D., and Thierry, J. C., eds) pp. 27–38, Oxford University Press, Oxford
Catalytic Domain Structures of MT-SP1/Matriptase, a Matrix-degrading Transmembrane Serine Proteinase

Rainer Friedrich, Pablo Fuentes-Prior, Edgar Ong, Gary Coombs, Michael Hunter, Ryan Oehler, Diane Pierson, Richard Gonzalez, Robert Huber, Wolfram Bode and Edwin L. Madison

J. Biol. Chem. 2002, 277:2160-2168.
doi: 10.1074/jbc.M109830200 originally published online November 5, 2001

Access the most updated version of this article at doi: 10.1074/jbc.M109830200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 41 references, 12 of which can be accessed free at http://www.jbc.org/content/277/3/2160.full.html#ref-list-1