Sheep mast-cell proteinases-1 and -3: cDNA cloning, primary structure and molecular modelling of the enzymes and further studies on substrate specificity

Sybil M. McALEESE*, Alan D. PEMBERTON†, Mary E. McGrATH†, John F. HUNTLEY‡ and Hugh R. P. MILLER*

*Department of Veterinary Clinical Studies, University of Edinburgh, Easter Bush Veterinary Centre, Roslin, Midlothian, EH25 9RG, Scotland, U.K., †Axys Pharmaceuticals, Inc., 180 Kimball Way, South San Francisco, California 94080, U.S.A., and ‡Moredun Research Institute, International Research Centre, Pentlands Science Park, Bush Loan, Penicuik, Midlothian EH26 0PZ, Scotland, U.K.

Sheep mast-cell proteinase-1 (sMCP-1) is a serine proteinase expressed predominantly by mucosal mast cells, with specificity for cleavage C-terminal to basic and hydrophobic amino acid residues. A cDNA encoding sMCP-1 has been cloned using reverse transcriptase (RT)-PCR. It appears to be translated as a pre-proenzyme with a 17-amino-acid signal peptide, a basic 226-amino-acid catalytic domain. A second cDNA, encoding a serine proteinase 90% identical with sMCP-1, was also cloned and named sMCP-3. Molecular models were constructed for both enzymes using coordinates for the refined X-ray structures of human chymase G, chymase and rat mast-cell proteinase-2. The model for sMCP-1 suggests that the acidic Asp-226 side chain extends into the substrate-binding pocket, hydrogen-bonding with Ser-190 on the opposite side and bisecting the pocket. The location of an acidic moiety in this position would favour interaction with basic substrate residues and binding of aromatic residues is rationalized by interaction of the positively charged equatorial plane with Asp-226. The balance between chymotryptic and trypsin-like activities of sMCP-1 was found to be sensitive to salt concentration, with increasing univalent cation concentration favouring chymotryptic activity relative to the trypsin. Using a peptide substrate representing residues 36–59 of the human thrombin receptor, increasing salt concentration favoured cleavage at Phe-43 rather than at Arg-41.

INTRODUCTION

Serine proteinases expressed by granular leucocytes are important mediators of inflammation and tissue remodelling [1–4] and determining their precise physiological role is the subject of much current research. One major family of granulocyte proteinases is exemplified by mast-cell chymase [5], encompassing a range of structurally related proteinases with varying activities and cellular localization. Chymase is expressed by mast cells in many mammalian species in which its ability to convert angiotensin I into angiotensin II is conserved [5]. Rodent mast cells express multiple chymases with varying chymotryptic-like activity [6]. The more distantly related granzyme B, cathepsin G and duodenase have yet further differing substrate specificities [7–9]. These specificities are derived primarily from alterations in amino acid residues defining the SI substrate-binding pocket, which binds the amino acid side chain of the substrate residue at which cleavage occurs [10]. (The nomenclature is that of Schechter and Berger [11], who defined seven substrate residues, P4–P3, which flank the scissile bond and the complementary substrate-binding pockets, S4–S3, on the enzyme.) For example, the X-ray crystallographic structure of human chymase shows a relatively unhindered substrate-binding pocket, allowing the insertion of large hydrophobic side chains [12]. However, in granzyme B, molecular modelling and mutational analyses have shown that the protrusion of the basic Arg-226 side chain into the pocket is responsible for its specificity for binding acidic P1 residues [13].

A number of mammalian serine proteinases are now known to possess the ability to cleave substrates for both trypsin and chymotrypsin, namely human chymase G, bovine duodenase and sheep mast-cell proteinase-1 (sMCP-1) [9,14,15]. The X-ray crystal structure of cathepsin G [14] shows the Glu-226 side chain extending across the centre of the primary substrate-binding pocket, presenting the negatively charged residue in a position where it can interact with basic side chains (preferably Lys), as well as aromatic residues, by interaction of negative Glu-226 with the positive equatorial plane of the aromatic moiety. Complete amino acid sequencing of bovine duodenase [16] also revealed a negatively charged residue (Asp) at position 226, suggesting a similar explanation for its dual trypsin–chymotrypsin activity. N-terminal amino acid sequencing and functional studies have shown that sMCP-1 is likely to be homologous with bovine duodenase [15]. This paper presents the complete cDNA sequence of sMCP-1, confirming the similarity with duodenase, and molecular modelling to rationalize the observed dual specificity.

The dual substrate specificity of sMCP-1 makes predicting its behaviour particularly difficult, because it can cleave at more than one type of amino acid residue in a potential substrate. In considering the effect of salt concentration, we have investigated conditions where one type of substrate specificity may be favoured over another.

MATERIALS AND METHODS

RNA extraction

Sheep bone-marrow mast cells were cultured in lymphocyte-conditioned medium (CM) with the addition of interleukin-3 and

Abbreviations used: sMCp, sheep mast-cell proteinase; RT, reverse transcriptase; RACE, rapid amplification of cDNA ends; rMCp-2, rat mast-cell proteinase-2; BPTI, bovine pancreatic trypsin inhibitor; suc-FLF-SBzl, succinyl-Phe-Leu-Phe-thiobenzyl ester; Z-K-SBzl, N-carbobenzoxy-Lys-thiobenzyl ester; BMMC, bone-marrow-derived mast cell.

1 To whom correspondence should be addressed (e-mail: alan.pemberton@ed.ac.uk).

The nucleotide sequences reported in this paper have been submitted to the EMBL nucleotide sequence database: accession nos. sMCp-1 (originally deposited as sMCp-1a) Y14654 and sMCp-3 (originally deposited as sMCp-1b) Y13462. The sequence for sMCp-2 has also been deposited and has the accession no. Y08133.
stem-cell factor, as described by Macaldowie et al. [17]. Total RNA was extracted from the harvested cells using TRI reagent (Sigma Chemical Co.). Abomasal tissue was extracted at slaughter from a sheep infected with the nematode parasite *Teladorsagia circumcincta* and snap-frozen in liquid nitrogen. The frozen tissue was ground up with TRI reagent on dry ice, the mixture allowed to thaw and the total RNA was extracted.

**Reverse transcriptase-PCR (RT-PCR), cloning and sequencing**

mRNA was reverse-transcribed using an oligo(dT) anchor primer (3'-dT-anchor) and Moloney murine leukaemia virus RT (M-MuLV, Boehringer Mannheim U.K. Ltd.). The resulting cDNA was amplified using specific primers (see below and Figure 1). P1 was designed from the N-terminal amino acid sequence of the protein sMCP-1 [15] using the same codons for matching residues as found in sMCP-2 (EMBL: Y08133, unpublished results). It was made degenerate for the third base of the histidine residue. P2 was based on a well-conserved region in other known mast-cell chymases [18–24]. In retrospect, it was found to have three errors in the 24-base oligonucleotide. PCR was done in 10 mM Tris/HCl buffer, pH 8.0, 50 mM KCl, 1.5 mM MgCl₂, with 200 µM each dNTP and 1 µM each primer. After denaturation (95 °C, 10 min), 2 units of Tag DNA polymerase (Boehringer Mannheim U.K. Ltd.) were added at 72 °C and the amplification programme (95 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s) was run for 30 cycles, followed by a final step of 72 °C for 30 min. The reaction mixtures were subjected to electrophoresis on agarose gels and Southern blot analysis. Positive PCR products were cloned into the TA cloning vector pCR2.1 (Invitrogen, Leek, the Netherlands). Positive clones were identified by restriction-enzyme digestion and Southern blot analysis. Nucleotide sequences were determined by the dideoxy chain-termination method using ABI prism dye terminator cycle sequencing (Perkin–Elmer, Norwalk, CT, U.S.A.). Sequence data were analysed using the Wisconsin Package (Genetics Computer Group, Madison, WI, U.S.A.).

**Rapid amplification of cDNA ends (RACE)**

For 3'-RACE, the first strand cDNA was amplified by PCR using P3 and the 3'-dT-anchor primer. The resulting products were reamplified using P4 and the 3'-anchor.

For the 5' end, a 5'/3'-RACE kit (Boehringer Mannheim U.K. Ltd.) was used, following the manufacturer's instructions; total RNA was reverse-transcribed using P5 and avian myeloblastosis virus RT. The cDNA produced was purified and a 3'-end homopolymeric tail was added using dATP and terminase transferase. The dA-tailed cDNA was amplified twice by PCR using the 5'-dT-anchor primer and the gene-specific primers P6 and P7. The thermal cycle profile used was denaturation at 94 °C for 10 min, addition of 2 units of Taq DNA polymerase at 72 °C, then 30 amplification cycles: 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, followed by a final step of 72 °C for 30 min. The RACE products were cloned and sequenced as described above.

**Primers**

The sequences of primers used were as follows: P1 (sense) ATCATCGGGGGCCAYAGGC; P2 (antisense) CAGAAGAGGGCCACTGCTCC; P3 (sense) AATCTTGAAATCCTAATGGAG; P4 (sense) TCAACCAAAGAAGGAATCTTCCTCC; P5 (antisense) GATGACCTGTGGGTCTCCT; P6 (antisense) CTGTCCACGATGTTGTGGGC; P7 (antisense) GGTTAGCCTGTAGTGGGC; P8 (sense) GATCTTGAAGTCCAAAGAGGAGA; P9 (antisense) CAAGAGATGAATGTTTATTAAGTTCCCTATTGCC.

**Southern blot analysis**

PCR products and restriction digestions of cloned products were electrophoresed on agarose gels and blotted on to Hybond N+ (Amersham International, Amersham, U.K.). Positive amplification products from the first PCR reaction were identified by hybridization with a digoxigenin-labelled cDNA probe, corresponding to the complete sequence of sMCP-2 (EMBL: Y08133, unpublished results). This probe hybridized to the P1-P2 PCR product at 50 °C with washing conditions of 0.5 × SSC (where 1 × SSC is 0.15 M NaCl/0.015 M sodium citrate)/0.1% (w/v) SDS, but not at 65 °C with washing conditions of 0.1 × SSC/0.1% (w/v) SDS. 3'- and 5'-RACE products were identified with oligonucleotide probes 3'-end-labelled with digoxigenin (Boehringer Mannheim U.K. Ltd.), prepared from primers P1, P2, P4 and P7. Hybridization was carried out at 40 °C with washing conditions of 1 × SSC/0.1% (w/v) SDS.

**N-terminal and internal amino acid sequencing of native sMCP-1**

N-terminal amino acid sequencing was performed on a gas-phase sequencer (P. Barker, Babraham Institute, University of Cambridge, U.K.). Internal sequence was derived from analysis of the products from autolysis of sMCP-1 in 0.1 M Tris/HCl, pH 8.5, at 45 °C for 3 days. The products (approx. Mᵋ 23000 and 5000) were separated by SDS/PAGE [25], blotted on to Immobilon-P (Millipore, Bedford, MA, U.S.A) and sequenced.

**Molecular modelling studies**

Construction of molecular models of sMCP-1 and sMCP-3 was facilitated by the availability of high-resolution X-ray crystal
structures of three closely related enzymes. When compared with sMCP-1, human cathepsin G [14] is most similar (57 % identity), and human chymase [12] (48 % identity) and rat mast-cell proteinase-2 (rMCP-2) [26] (45 % identity) are also good models. For sMCP-3, identity with these three enzymes is 59 %, 51 % and 48 % respectively.

Alignments of the primary structures of the five proteinases were optimized manually within the Homology module of the Insight suite of programs (MSI, San Diego, CA, U.S.A.). Structurally conserved regions were identified and coordinates transferred from the most related reference protein to the nascent model. Side chains were automatically replaced as necessary during this process. A similar methodology was used to build loop regions. The close relatedness of the sMCPs to cathepsin G meant that the majority of the two models was derived from that protein. However, rMCP-2 and human chymase provided coordinates for several regions where they shared more similarities with the modelled proteins. After assembling the models, refinement with Discover (MSI, San Diego, CA, U.S.A.) was used for energy minimization. Profiles-3D (MSI, San Diego, CA, U.S.A.) was used to quantify the validity of the models.

Investigation of substrate-specificity determinants in the sMCPs was carried out by modelling interactions with substrate-like inhibitors. High-resolution structures of complexes between trypsin and the inhibitor ecos [27] (available from M. E. McGrath, Axys Pharmaceuticals, Inc., San Francisco, CA, U.S.A.), between trypsin and basic pancreatic trypsin inhibitor (BPTI) [28] (pdb: 1brb) and between chymotrypsin and turkey ovomucoid inhibitor third domain [29] (pdb: 1cho) were superimposed on to the sMCP models using coordinates for the catalytic residues. This allowed the detection of putative interactions between the sheep enzymes and substrates. Three different inhibitors were used in order to ascertain the most likely interactions. Inhibitor side chains were replaced to match the observed specificity of sMCP-1 and modelled using the Bio-polymer module of Insight (MSI, San Diego, CA, U.S.A.).

**Effect of ionic strength on hydrolysis of S-ester substrates by native sMCP-1**

Initial rates of hydrolysis of the chymotrypsin substrate succinyl-Leu-Phenylthiobenzoyl (suc-FLF-SBz) and the trypsin substrate N-carbobenzoxy-Lys-thiobenzoyl ester (Z-K-SBz) were determined by adding 10 µl of 5 mM solutions in DMSO to appropriately buffered mixtures (190 µl) containing 10 µl of sMCP-1 [127 nM in H2O; 0.1 % (w/v) Brij35] and 5,5′-dithiobis-(2-nitrobenzoic acid) (10 µl of 10 mM in DMSO). Buffers were prepared to give a final concentration of 50 mM at pH 7.5 of Tris-acetate, Tris/HCl or sodium phosphate, containing between 0 and 0.5 M additional salt (NaCl, NaNO3, Na2SO4, KNO3, CaCl2, MgCl2). Changes in absorbance were measured at 405 nm and were constant over the period of observation (60 s). All incubations were carried out at room temperature (22–24 °C).

**Cleavage of thrombin receptor peptide TR(36–59) by native sMCP-1**

A peptide synthesized on the basis of residues 36–59 of the human thrombin receptor [30], denoted as TR(36–59) (structure: ATLDPRSFLLRNPDKYEPFWEDE), was synthesized by Genosys (Cambridge, U.K.) and checked for purity by using HPLC and MS. Incubations with the synthetic peptide were carried out at ambient temperature (23 °C), comprising TR(36–59) (50 µl; 1 mg/ml in H2O), 75 µl of buffer (see below), 5 µl of internal standard Ala-Trp (1 mg/ml in H2O), 10 µl of H2O plus 10 µl of sMCP-1 (6.4 µM). Buffers used included 2× PBS [20 mM phosphate/300 mM NaCl (pH 7.2)] and 0.1 M Tris-acetate, pH 7.5, in the presence or absence of 1 M NaCl. Aliquots of 15 µl were removed at varying time points and added to 15 µl of 10 % (v/v) acetic acid to stop the reaction. Stopped samples were immediately cooled on ice, then frozen at −20 °C until required for analysis.

**Analysis of TR(36–59)-cleavage products**

Injections of 20 µl of each stopped sample (see above) were loaded on to a C5 reversed-phase HPLC column (Jupiter 10 µm C5 300 Å, 250 × 4.6 mm; Phenomenex U.K. Ltd., Macclesfield, Cheshire, U.K.) and eluted in an 8–80 % acetonitrile gradient in 0.055 % (v/v) trifluoroacetic acid, with UV detection at 214 nm. Millenium 2010 software (Waters U.K. Ltd., Watford, Herts., U.K.) was used to control the HPLC system (Waters 626) and perform peak integration. The appearance and disappearance of peaks were demonstrated by plotting the peak area relative to internal standard for incubations stopped at varying time points. Peptides were identified by collecting peaks from representative incubations and analysing freeze-dried fractions by matrix-assisted laser-desorption MS (I. Davidson, University of Aberdeen, Scotland, U.K.).

**RESULTS**

**Sequencing of sMCP-1 cDNA**

RT-PCR was performed, first using primers 1 and 2. A product of the expected size (approx. 600 bp) was obtained, cloned and sequenced. Gene-specific primers 3–7 were made on the basis of this sequence and used for 3′- and 5′-RACE. The full-length sMCP-1 cDNA sequence was thus obtained from three overlapping fragments. To ensure maximum accuracy, clones were sequenced in both directions. Confirmation of the sequence was obtained from a complete cDNA clone prepared using primers 8 and 9. Clones P1-P2 and P4-3′-dT-anchor were obtained from abomasal tissue RNA and clones 5′-dT-anchor-P7, P4-3′-dT-anchor and the full-length clone P8-P9 were obtained from bone-marrow-derived mast-cell (BMMC) RNA, suggesting that this sequence is mast-cell-derived. A total of 889 bp of sMCP-1 nucleotide sequence was obtained, including 738 bp of open-reading frame. Comparison with other known mast-cell proteinases indicates a pre-proenzyme with a 17-amino-acid signal peptide, a basic 2-amino-acid propeptide and a 226-amino-acid catalytic domain.

**The cDNA sequence for sMCP-3; a closely related sMCP**

A second clone P1-P2 was obtained from BMMC RNA at the same time, and under identical conditions, as that for sMCP-1. The 3′- and 5′-ends were obtained using primers P3b, P4b, P6b and P7b, which were designed to differentiate between the two partial sequences by making maximum use of the differences found (see Figure 2). Clones obtained using P8 and P9 could be either sMCP-1 or sMCP-3, but were readily distinguishable because the cDNA for sMCP-1 has an EcoRI restriction site, which is absent in sMCP-3. The complete sequences were found to be 93 % identical. A total of 920 bp of sMCP-3 nucleotide sequence has been determined. The partial clones P1-P2, P4b-3′-dT-anchor and 5′-anchor-P7 and the complete clone P8-P9 were obtained from BMMC RNA. The partial clone P4-3′-dT-anchor was also obtained from abomasal tissue
RNA. This suggests that sMCP-3 is mast-cell-derived and may be co-expressed with sMCP-1 in abomasal mast cells. Figure 2 compares the deduced amino acid sequences, which show 90% identity. The signal and activation peptides are identical, with only one base change in the nucleotide sequence. However, sMCP-3 has a C-terminal extension of six residues. Whether or not this would be present in mature protein is unknown. Like cathepsin G [31], a portion of the C-terminus may be removed on activation. Important residues of the S1 substrate-binding site, Asn-189 and Asp-226, are the same in sMCP-1 as they are in bovine duodenase. The predicted net charge of the catalytic domain of sMCP-1 is +5 compared with +6 for duodenase. Sheep MCP-3 is much less basic, having a net charge of +1.

**Molecular modelling of sMCP-1 and sMCP-3**

The availability of three high-quality crystal structures for enzymes closely related to sMCP-1 and sMCP-3 allowed assembly of accurate molecular models. The proposed three-dimensional structures of the models were found to be compatible with their primary structures as determined by using Profiles-3D. The sMCP-1 model scored 97.4, where the expected score for a perfectly folded protein of that size is 102.6, which compares well with the rMCP-2 X-ray structure, where the score is 98.8/101.7. The $\phi$, $\psi$ core region occupancy was calculated to be 81.3%. Any inherent error in the modelling process is mitigated by the huge database of structural and enzymic data that exists for the chymotrypsin-like serine proteinases. Thus these constructs can be used in a cautiously predictive manner, especially when the predictions are limited to substrate-binding subsites proximal to the well-studied active site.

The major determinant of specificity for the chymotrypsin-like serine proteinases is the P1–S1 interaction, and so most of the analysis was limited to this region. The S1 pocket of these enzymes is composed of three $\beta$-strands. The Cys-191–Cys-220
Sheep mast-cell proteinase-1 and -3: sequence and structure prediction

Figure 3 Comparison of the deduced amino acid sequence of sMCP-1 with those of other similar enzymes

The sMCP-1 sequence is aligned with bovine duodenase [16], human granzyme B [32], human cathepsin G [31], mouse cathepsin G [33] and chymotrypsinogen [42]. Sequence gaps resulting from the optimization of alignments are indicated by hyphens (—). Residues identical with those of sMCP-1 are indicated by dots. Standard chymotrypsinogen numbering is shown. Symbols used: X, active-site residues; *, residues 189 and 226.

disulphide is conserved throughout much of this proteinase family and serves to fasten the front loop to a side strand of the S1 pocket. This rigidifying feature is missing in the sheep enzymes, as in the other mast-cell proteinases discussed here. Side chains that extend into the base of the pocket determine specificity for cationic amino acids (Asp-189 in trypsin) or large hydrophobic side chains (Ser-190 in chymotrypsin). Amino acids along the walls of the primary specificity pocket may serve to modulate the size of the pocket, as exemplified by elastase, which prefers small alkyl side chains and has valine and threonine at positions 216 and 226 respectively.

The models for both sMCP-1 and sMCP-3 suggest that the side chains of residues 189, 190 and 226 extend into the bottom half of the S1 pocket, forming a plane that bisects the pocket laterally (Figure 4). As a result, the upper part of the pocket is unhindered and the lower part consists of an inner and an outer section, similar to what is observed in the X-ray crystal structure of cathepsin G [14]. In the absence of substrate or inhibitor, there is the potential for a stable interaction between Tyr-228, Ser-190 and Asp-226 and perhaps a water molecule, as seen for the equivalent grouping of residues in the trypsin double mutant Gly-226 → Asp, Asp-189 → Gly [28]. Accommodation of substrate or inhibitor P1 residues was modelled using three proteinase–inhibitor structures. The third domain of turkey ovo-mucoid inhibitor has a P1 leucine and modelling indicates that, consistent with observed substrate preferences for sMCP-1, leucine fits well in S1 (Figure 4). Other preferred P1 residues for this enzyme are lysine, arginine and phenylalanine. Modelling with inhibitors in which P1 has been computationally mutated to any of these residues shows lysine or arginine would be able to form a salt bridge with Asp-226, whereas P1 phenylalanine would probably bind as seen for cathepsin G, where the negative charge in the binding pocket is favourably juxtaposed with the positive equatorial edge of the ring (Figure 4). The shape of the S1 pocket in sMCP-3 would be the same as that predicted for sMCP-1. However, whereas the aspartate at position 189 would seem to dictate specificity for basic residues, the presence of Asn-226 may sterically hinder this interaction.

Effect of salt on hydrolysis of S-ester substrates by native sMCP-1

In the presence of 50 mM Tris-acetate, pH 7.5, the rate of hydrolysis of the chymotrypsin substrate (Suc-FLF-SBzl) was increased with increasing NaCl concentration, whereas hydrolysis of the trypsin substrate (Z-K-SBzl) was relatively unaffected (Figure 5a). The same effect was observed when NaCl was replaced by NaNO₃, Na₂SO₄ or KNO₃, or by using Tris/HCl in place of Tris-acetate (results not shown). However, increasing CaCl₂ concentrations caused both activities to decrease slightly by approximately the same extent (Figure 5b) and MgCl₂ had no significant effect on activity (results not shown). In the absence of added salt, the rate of chymotrypsin substrate hydrolysis was greater in 50 mM sodium phosphate buffer pH 7.5 (Figure 5c), than in Tris-acetate. Using phosphate buffer, increasing [NaCl]
The active-site region for the molecular model of sMCP-1 is shown in stereo, coloured blue. The catalytic triad residues (Asp-102, His-57 and Ser-195) are shown as sticks and labelled, as are the following residues of interest in the S1 substrate-binding pocket: Tyr-228, Ser-190, Asp-226 and Asn-189. The model is shown superimposed on to the structure of human cathepsin G (black) with its bound peptidyl phosphonate inhibitor featuring P1 phenylalanine. Three proteinaceous inhibitors have been modelled at the sMCP-1 active site and their P1 and P2 residues shown in order to visualize the basis for the observed substrate specificity of this proteinase. The naturally occurring P1 leucine of turkey ovomucoid inhibitor domain III is shown in purple, and the naturally occurring P1 lysine of BPTI is shown in green. P1 of ecotin has been mutated to phenylalanine and is coloured red.

Figure 5 Effect of salt concentration on sMCP-1 hydrolytic activity

S-ester substrates were used to monitor the chymotrypsin-like activity (●) (suc-FLF-SBzl) and trypsin-like activity (☐) (Z-K-SBzl) of sMCP-1. In the presence of 50 mM Tris-acetate buffer, pH 7.5, the effect of altering NaCl and CaCl₂ concentration is shown in (a) and (b) respectively. The effect of altering NaCl concentration in 50 mM phosphate buffer, pH 7.5, is shown in (c). (d) shows how the ratio of chymotrypsin-like to trypsin-like activity (C:T ratio) is affected by NaCl concentration in 50 mM Tris-acetate, pH 7.5 (▲) and 50 mM phosphate, pH 7.5 (▲).
Hydrolysis of thrombin receptor peptide TR(36–59) by native sMCP-1

MS analysis was carried out on peptide fragments obtained by HPLC of a representative incubation of TR(36–59) with sMCP-1 in 50 mM Tris/ HCl, pH 7.5. A typical chromatogram is shown in Figure 6(a). Fragment masses indicated that cleavage occurred at Arg-41–Ser-42, Phe-43–Leu-44 and Leu-45–Arg-46, with cleavage at Phe-43 and Arg-41 under differing conditions is shown in (b)–(d). Buffers used were: (b) PBS; (c) [50 mM Tris-acetate (pH 7.5)]; and (d) [50 mM Tris-acetate/0.5 M NaCl (pH 7.5)]. Peak areas are shown relative to the internal standard.

DISCUSSION

It was considered until recently that serum proteinases expressed within mast cells exhibited either trypsin-like (tryptase) or chymotrypsin-like (chymase) activities. However, it is emerging that sMCP-1 in sheep mast cells and cathepsin G in human mast cells (and neutrophils) have substrate specificities that blur this distinction. Like chymase, sMCP-1 is capable of inhibition by plasma serpins and cleaves certain bioactive peptides with chymotrypsin-like specificity [15,34], yet it mimics trypsin in activating fibroblasts, and cleaves fibrinogen very efficiently with trypsin-like specificity [34]. This makes the task of disentangling its role from those of tryptase and chymase all the more complex and fascinating.

The deduced amino acid sequence for sMCP-1 described here verifies that this enzyme belongs to the group of enzymes that includes mast-cell chymases, granzyme B and cathepsin G, which differs from the archetypal serine proteinase, chymotrypsin, in that its members lack the Cys-191–Cys-220 disulphide bridge. (Throughout the discussion, chymotrypsinogen numbering will be used to refer to specific amino acids.) By comparison with sequences of related proteinases, sMCP-1 consists of a 226-amino-acid mature protein, resulting from the processing of a 17-residue signal peptide and cleavage of a basic activation dipeptide. The residues of the serine proteinase catalytic triad (His-57, Asp-102 and Ser-195) are present, as expected. Direct N-terminal and internal amino acid sequencing results obtained using preparations of the purified protein designated sMCP-1 confirm the deduced amino acid sequence obtained from the sMCP-1 cDNA, while elucidating a possible polymorphism at Ile-33. Previous studies addressing the possibility that sMCP-1 preparations might consist of a mixture of enzymes of differing specificity have consistently shown evidence of homogeneity [15].

The cDNA encoding a highly similar enzyme (90% deduced amino acid sequence identity) was also identified and termed sMCP-3. The sMCP-3 protein has not yet been identified and characterized, but from the deduced sequence we are able to predict that it differs from sMCP-1 in that it is much less basic and therefore unlikely to co-purify with sMCP-1 on ion-exchange chromatography. In addition, sMCP-3 may be glycosylated, because it contains a potential N-linked carbohydrate-binding site, whereas sMCP-1 has none. However, it is probable that current immunochemical techniques for detection of sMCP-1 would also detect sMCP-3. Because sMCP-1 staining has not been observed in cells types other than mast cells and because sMCP-3 was cloned from ovine bone-marrow-derived mast cells, the data suggests that both proteinases are most likely mast-cell-derived.

Important differences between the two enzymes are present in the residues defining the S1 substrate-binding pocket, where Asn-189, Ser-190 and Asp-226 in sMCP-1 are transposed to Asp-195 and Ser-226 in sMCP-3. Molecular modelling suggests that the side chains of residues 189, 190 and 226 bisect the lower section of the S1 substrate-binding pocket, whereas sMCP-1 has none. However, it is probable that current immunochemical techniques for detection of sMCP-1 would also detect sMCP-3. Because sMCP-1 staining has not been observed in cells types other than mast cells and because sMCP-3 was cloned from ovine bone-marrow-derived mast cells, the data suggests that both proteinases are most likely mast-cell-derived.
ever, the sMCP-1 model clearly shows that the positively charged end of a lysine side chain could interact with the carboxy group of Asp-226. In a similar manner to that seen for the trypsin mutant, arginine can also be accommodated in sMCP-1. Since greater rearrangement of S1 is probably necessitated by the latter P1 element, cleavage after arginine may be thermodynamically less favourable than after lysine. Indeed, sMCP-1 hydrolyses tosylglycylprolyl-lysine 4-nitroanilide much more efficiently than the corresponding arginine-containing substrate [15].

The binding of aromatic residues in the outer section of the binding pocket may occur by the same mechanism as in cathepsin G, where the carboxy group of Glu-226 is aligned with the equatorial plane of the aromatic ring [14], which has a net positive charge due to delocalization of electrons on either side of the ring. The lack of an acidic residue at 226 in sMCP-3 indicates that it would not be able to bind aromatic residues by this mechanism. However, sMCP-3 might be able to bind basic residues by interaction with Asp-189 at the bottom of the pocket. Alternatively, Asn-226 and Ser-190 may block access to Asp-189, residues by interaction with Asp-189 in the pocket, resulting in an elastase-like specificity for small hydrophobic residues, binding in the upper section of the pocket. Purification and/or expression studies of sMCP-3 will be required to resolve this question.

The sMCPs are probably able to make more extensive contacts with substrate residues C-terminal to the scissile bond (the prime side) than those observed for the paradigm for this family, chymotrypsin. In this region, they are most similar to cathepsin G, but also resemble rMCP-2 and chymase in having a larger loop (30’s loop) than chymotrypsin. In addition, differently from chymotrypsin, but as found for the other mast-cell proteinases, the loop is probably situated closer to the active site and is able to sample substrate residues P2 through P4 via residues 37–41 of that loop. These interactions could be both non-specific main-chain hydrogen bonds, as well as specific side chain interactions, between proteinase and substrate. Their effect would be increased substrate discrimination, when compared with the related digestive enzymes.

We also investigated whether, by altering reaction conditions, one specificity could be favoured over another. Increasing univalent cation concentration (e.g. Na+, K+) in Tris buffers caused an increase in chymase activity, rising to 5-fold at 0.5 M, with no increase in trypatase activity. However, increasing concentrations of bivalent cations (Ca2+, Mg2+) did not increase chymase activity. A similar effect has been noted for human chymase and cathepsin G, where increasing NaCl concentration to 2 M caused an approximate 2 to 3-fold increase in chymotrypsin-like activity [8,35]. In 50 mM sodium phosphate buffer, increasing NaCl concentration did not cause a notable increase in sMCP-1 chymase activity, but a decrease in trypatase activity was observed. Importantly, however, in both Tris and phosphate buffers the overall effect of increasing salt concentration was to favour chymase-like activity relative to trypatase-like. This may be due simply to increasing hydrophobic interactions favouring binding of P1 phenylalanine residues in the S1-binding pocket at increased ionic strength, although the insensitivity of chymotrypsin to increased salt concentration and the absence of this effect with bivalent cations appears to confound this idea.

In a previous study using small peptide fragments representing regions close to the thrombin receptor cleavage site, Molino et al. [36] have shown that cathepsin G can cleave at Arg-41 (which would activate the receptor) as well as Phe-55 (leading to inactivation), suggesting that a peptide encompassing both sites would provide a useful substrate in which to observe changes in relative chymase–trypatase specificity. Hydrolysis of the TR(36–59) peptide by cathepsin G was minimal at Arg-41, compared with cleavage at Phe-55 (results not shown) and, indeed, Parry et al. [37] have already shown that the TR(38–60) peptide was cleaved by cathepsin G in the same manner. However, with sMCP-1, cleavages at Arg-41 and Phe-43 occurred at comparable rates in addition to a lower rate of cleavage at Leu-45. Notably, increasing salt concentration was found to favour cleavage at Phe-43 relative to Arg-41, reinforcing the studies with chromogenic substrates. Therefore it appears likely that if sMCP-1 were to experience a change in univalent cation concentration in vivo, such as in airway surface fluid where salt concentration is significantly lower than in plasma [38], this would affect the relative cleavage specificity of the proteinase in that environment.

Our previous analyses of sMCP-1 substrate specificity and N-terminal sequencing [15] showed close similarities to bovine duodenase [9,39]. The data presented here strongly suggest that sMCP-1 and duodenase are homologous enzymes (85%, amino acid sequence identity), although their cellular localization may be quite different. The sheep enzyme and the homologous enzyme in the goat (goat mast-cell proteinase-1) have been immunolocalized in mucosal mast cells [40,41], whereas Zamolodchikova et al. [39] report that duodenase is present in the granules of secretory epithelial cells of bovine duodenal mucosa, where it has a proposed role in enteropeptidase activation. The expression of homologous enzymes in the same tissue location, but by different cell types, would suggest a cross-over of function in the two species. It would be interesting to discover whether duodenase is expressed by bovine mucosal mast cells during intestinal nematodiasis.

In summary, sequence analysis and molecular modelling have been used to rationalize the dual trypatase–chymase substrate specificity of sMCP-1. The balance of specificity was sensitive to changes in electrolyte composition, suggesting that sMCP-1 activity in vivo may vary in differing compartments.

We thank Dr. R. Tanaka for helpful support and discussion, Ms. A. Mackellar and Mrs. J. Vaagenes for valuable technical assistance and Dr. F. Jackson for supply of ovine gastric mucosa. DNA sequencing was performed by the DNA sequencing service, University of Durham, N-terminal amino acid sequencing by Mr. P. Barker at the Babraham Institute, Cambridge and MS by Mr. I. Davidson, University of Aberdeen.

This work benefited from the use of the SEQNET facility and was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; award no. A01356). J.F.H. was funded by the Scottish Office, Agriculture and Fisheries Department.

REFERENCES

Sheep mast-cell proteinase-1 and -3: sequence and structure prediction


Received 13 February 1998/29 April 1998; accepted 14 May 1998


