ELSEVIER

Biochimica et Biophysica Acta 1361 (1997) 11–19

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

Short sequence-paper

Molecular cloning of seprase: a serine integral membrane protease from human melanoma

Leslie A. Goldstein, Giulio Ghersi, Mayra L. Piñeiro-Sánchez, Monica Salamone, Yunyun Yeh, Denise Flessate, Wen-Tien Chen *

Lombardi Cancer Center and Department of Cell Biology, Georgetown University Medical Center, 3970 Reservoir Rd. N.W., Washington, DC 20007, USA

Received 22 November 1996; revised 18 March 1997; accepted 1 April 1997

Abstract

Seprase is a homodimeric 170 kDa integral membrane gelatinase whose expression correlates with the invasiveness of the human melanoma cell line LOX. Here, we report the molecular cloning of a cDNA that encodes the 97 kDa subunit of seprase. Its deduced amino acid sequence predicts a type II integral membrane protein with a cytoplasmic tail of 6 amino acids, followed by a transmembrane domain of 20 amino acids and an extracellular domain of 734 amino acids. The carboxyl terminus contains a putative catalytic region (~ 200 amino acids) which is homologous (68% identity) to that of the nonclassical serine protease dipeptidyl peptidase IV (DPPIV). The conserved serine protease motif G-X-S-X-G is present as G-W-S-Y-G. However, sequence analysis of seprase cDNA from LOX and other cell lines strongly suggests that seprase and human fibroblast activation protein α (FAPα) are products of the same gene. We propose that seprase/FAPα and DPPIV represent a new subfamily of serine integral membrane proteases (SIMP). © 1997 Elsevier Science B.V.

Keywords: Seprase; FAPα; DPPIV; Melanoma; cDNA; Gelatinase

At the cell-extracellular matrix (ECM) interface, invasive tumor cells elaborate extracellular ventral membrane protrusions termed 'invadopodia' that exhibit increased proteolytic activities at invasion foci and thus allow metastatic cells to dissolve the surrounding matrix [1]. We reported the identification of a novel 170 kDa integral membrane gelatinase, seprase, from a human malignant melanoma cell line LOX [2]. This cell line had previously been shown to have a high incidence of lung metastasis in athymic mice [3]. And there appears to be a direct correlation between the level of seprase expression and human melanoma invasiveness as measured by an in vitro ECM degradation/invasion assay [2,4]. In a recent study, we isolated the active enzyme from cell membranes and shed vesicles of LOX cells [5]. It had a molecular mass of 170-kDa and consisted of 97-kDa subunits that were *N*-glycosylated. The proteolytic activity of seprase is associated with its dimeric form and is completely blocked by serine-protease inhibitors. Sequence analysis of 3 internal proteolytic fragments of the 97-kDa polypeptide [5] revealed highest homology (67–88%) to the 95-kDa FAPα [6], an inducible cell surface glycoprotein which is reported to be expressed on reactive stromal fibroblasts of epithelial cancers and healing wounds but

* Corresponding author. Fax: +1 202 6873300; E-mail: chenw@gunet.georgetown.edu

0925-4439/97/$17.00 © 1997 Elsevier Science B.V. All rights reserved.
PII S0925-4439(97)00032-X

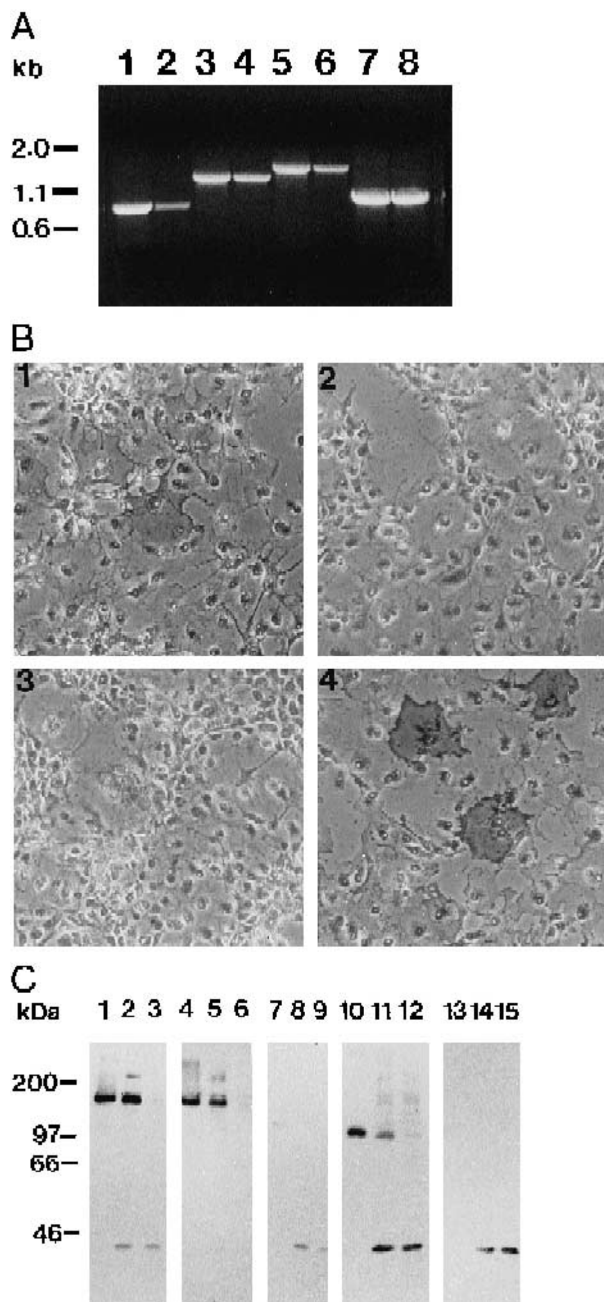
whose function is unknown [7], and to a lower extent (33–70%) with the 110-kDa subunit of DPPIV [8–11] and with (25–50%) DPPX [12,13]. Herein we report the isolation and expression of a cDNA clone that encodes the 97 kDa subunit. Its deduced amino acid sequence confirms that seprase, like DPPIV, is a member of the family of nonclassical serine hydrolases [8–17]. Additionally, based on cDNA sequence analysis of LOX and other cell lines we propose that seprase and FAPα are products of the same gene.

We carried out RT-PCR of LOX and human WI-38 fibroblast total RNA using 3 oligonucleotide primer pairs that correspond to the cDNA sequence of human FAPα (GenBank accession no. U09278; Fig. 1A). The resulting 3 amplicons (825, 1265 and 1468 bp) are overlapping and extend from the 5′ UTR (FAP 1; nt 161) to the 3′ UTR (FAP 6; nt 2523) thus encoding the complete FAPα ORF. No difference was seen in the number or size of the cDNA amplicons generated from LOX and WI-38 RNA. Therefore, to synthesize putative full length seprase cDNA, RT-PCR of LOX RNA utilizing the primer pair FAP 1 (5′ CCACGCTCTGAAGACAGAATT 3′; sense) and FAP 6 (5′TCAGATTCTGATACAGGCT 3′; antisense) was carried out. The resultant ~2.4 kb amplicon was subcloned into the pCR3.1 mammalian expression vector (Invitrogen) [5].

To determine whether this amplicon encoded seprase, we transiently transfected COS-7 cells with the above mentioned recombinant plasmid (pA15), the vector (pA11), and an expression plasmid (pCD26) containing human DPPIV cDNA [10]. We used the antiseprase mAb, D8 [4], and a class matched (IgG2a) antiDPPIV mAb, F4 (negative control), for detection of seprase surface expression. As can be seen in Fig. 1B, the antiseprase mAb specifically stained cells that had been transfected with pA15 (Fig. 1B1) but the D8 mAb did not stain the pCD26 transfected cells (Fig. 1B2). In addition, the F4 mAb did not stain pA15 transfectants (Fig. 1B3) but did stain the pCD26 transfectants (Fig. 1B4). Detergent extracts were prepared from pA15 and pA11 transiently transfected COS-7 cells and Western Blot analysis was performed (Fig. 1C). Western blots (Fig. 1C, Lanes 1–6) of dimeric seprase (170 kDa) utilizing the antiseprase mAbs D8 and D28 [4] show that the pA15 transfected cells (Fig. 1C, Lanes 2 and 5) express a seprase band that is also present in the control LOX cell extract (Fig. 1C, Lanes 1 and 4) but at markedly lower endogenous levels in the pA11 transfected cells (Fig. 1C, Lanes 3 and 6). The specificity of the antiseprase mAbs in these blots was confirmed using a class matched (IgG2a) negative control mAb B5 (Fig. 1C, Lanes 7–9). Likewise, the mAb D8 in the Western blot of denatured seprase (97 kDa) detects comigrating bands for the LOX and pA15 extracts (Fig. 1C, Lanes 10 and 11) and a markedly lower intensity band for the mock transfected lysate (Fig. 1C, Lane 12). The specificity of the D8 mAb for denatured seprase was also confirmed by the negative control mAb B5 (Lanes 13–15). Importantly, we have also detected the expression of seprase proteolytic activity in pA15 transiently transfected COS-7 cell detergent lysates which is reported elsewhere [5]. In these experiments functional dimeric seprase (170 kDa) as determined by gelatin zymography was detected in the pA15 transfectants but not in the mock (pA11) transfected cells.

Sequence analysis of the cDNA insert of pA15 (Accession number U76833) revealed an ORF of 2280 bp which encodes a polypeptide of 760 amino acids with a M_r 87 722 (Fig. 2). The cDNA sequence predicts a type II integral membrane protein with a short cytoplasmic tail (6 amino acids) followed by a hydrophobic transmembrane domain (20 amino acids) and a relatively large extracellular domain composed of 734 amino acids. Three internal peptides generated by Lys C digestion of the 97 kDa subunit [5] yielded 10, 12 and 8 amino acid residues from sequence analysis of their N-termini that correspond to amino acid positions (Fig. 2) Phe^{220} to Thr^{229} (80% identity), Tyr^{461} to Pro^{472} (67% identity) and Lys^{510} to Thr^{517} (88% identity), respectively. Importantly, sequence analysis of 5 additional LOX seprase cDNA clones (all generated by RT-PCR) in the regions corresponding to the peptide sequence data confirm the pA15 sequence. In addition, the cDNA sequence indicates that Phe^{220}, Tyr^{461} and Lys^{510} are all preceded by a Lys residue and that the deduced amino acid sequence identity between seprase and FAPα in the regions corresponding to the 3 peptides are 90, 100 and 100%, respectively. There are 6 potential *N*-glycosylation sites: 5 of the sites are clustered in a membrane proximal region extending from Asn^{49} to Asn^{314} with the sixth site, Asn^{679}, located in the putative catalytic region (see below). There are 12

cysteine residues: 7 of these residues are clustered in a region that extends from Cys^{305} to Cys^{466}. Located at the carboxyl terminus is a putative catalytic region consisting of ~200 amino acids that extends from Tyr^{560} to Asp^{760}. Within this region is the catalytic triad of residues Ser^{624}, Asp^{702} and His^{734} which are in a nonclassical sequence orientation [14–16]. The consensus motif Gly-X-Ser-X-Gly which is characteristic of serine proteases is conserved around Ser^{624}.



A comparison of the nucleotide sequence of the pA15 cDNA to that for FAPα [6] indicated a difference of 8 base pairs in the ORF of their respective cDNAs (Fig. 2). There are 5 base substitutions: 3 give rise to nonconservative amino acid substitutions at residues #207 (Pro to Ala), #229 (Lys to Thr) and #354 (Arg to Thr), respectively. Five additional LOX seprase cDNA clones (produced from 3 RNA preparations) have identical nucleotide sequences to pA15 at the 3 amino acid substitution positions. The fourth and fifth substitutions are silent corresponding to

Fig. 1. (A) RT-PCR of LOX and WI-38 fibroblast RNA. Total RNA (6 μg) from LOX (Lanes 1, 3, 5 and 7) and WI-38 (Lanes 2, 4, 6 and 8) was screened for the presence of seprase/FAPα mRNA (Lanes 1–6) and glyceraldehyde 3-phosphate dehydrogenase (G3PDH) mRNA (Lanes 7, 8). Lanes 1–6 utilized primers that correspond to the FAPα cDNA sequence [6]. Lanes 1 and 2: 5′ CCACGCTCTGAAGACAGAATT 3′(FAP 1; sense) + 5′ GATTCTTAGCTCCAGCCT 3′(FAP 2; antisense) nt 161 to nt 986. Lanes 3 and 4: 5′ TTACATCTATGACCTTAGCA 3′(FAP 11; sense) + 5′ ACAGACCTTACACTCTGAC 3′(FAP 4; antisense) nt 598 to nt 1863. Lanes 5 and 6: 5′ CCAGCAATGATAGCCTCAA 3′(FAP 3; sense) + 5′ TCAGATTCTGATACAGGCT 3′ (FAP 6; antisense) nt 1055 to nt 2523. FAP 1 and FAP 6 correspond to sequences within the 5′ UTR and 3′ UTR, respectively. Lanes 7 and 8 utilized the human G3PDH control amplimer set (Clontech). Samples were electrophoresed on a 1% agarose gel. (B) COS-7 cells transiently transfected with pA15 express seprase. Immunostaining of COS-7 cells that were transiently transfected with recombinant plasmids containing seprase (pA15) and human DPPIV (pCD26) [10] cDNA were stained with the antiseprase mAb D8 and the class matched (IgG2a) antiDPPIV mAb F4. **1**: pA15 transfected cells stained with D8. **2**: pCD26 transfected cells stained with D8. **3**: pA15 transfected cells stained with F4. **4**: pCD26 transfected cells stained with F4. Primary mAbs D8 and F4 were detected by the StrAviGen Multilink Detection System and AEC substrate (BioGenex). (C) Detection of pA15 encoded dimeric (170 kDa) and monomeric (97 kDa) seprase. Western blot analysis was carried out on detergent extracts of pA15 and pA11 (mock) transiently transfected COS-7 cells and on LOX cell extract that was purified by wheat germ agglutinin (WGA) column chromatography [2] using the antiseprase mAbs D8 (Lanes 1–3, 10–12) and D28 (Lanes 4–6) and the class matched (IgG2a) negative control mAb B5 (Lanes 7–9; 13–15). Detergent extract samples were either not heated before undergoing SDS-PAGE (170 kDa) in Lanes 1–9 or heated at ~100°C for 5 min (97 kDa) in Lanes 10–15. Lanes 1, 4, 7, 10, 13: WGA purified LOX extract. Lanes 2, 5, 8, 11, 14: pA15 transfected cell extract. Lanes 3, 6, 9, 12, 15- pA11 transfected cell extract. Detergent extract samples were resolved by 10% SDS-PAGE and transferred to Optitran nitrocellulose membranes (Schleicher and Schuell).

nucleotide positions #252 (A to G) and #2124 (A to G), respectively. Analysis of 6 additional LOX clones all revealed an A nucleotide at position #252. At nucleotide position #2124, 5 of 7 additional LOX clones are in agreement with pA15. The remaining difference of 3 bp is due to the insertion of 3 guanine

```
ATG AAG ACT TGG GTA AAA ATC GTA TTT GGA GTT GCC ACC TCT GCT GTG CTT GCC TTA TTG GTG ATG TGC ATT GTC   75
M   K   T   W   V   K   I   V   F   G   V   A   T   S   A   V   L   A   L   L   V   M   C   I   V     25
TTA CGC CCT TCA AGA GTT CAT AAC TCT GAA GAA AAT ACA ATG AGA GCA CTC ACA CTG AAG GAT ATT TTA AAT GGA  150
L   R   P   S   R   V   H   N   S   E   E   N   T   M   R   A   L   T   L   K   D   I   L   N   G     50
ACA TTT TCT TAT AAA ACA TTT TTT CCA AAC TGG ATT TCA GGA CAA GAA TAT CTT CAT CAA TCT GCA GAT AAC AAT  225
T   F   S   Y   K   T   F   F   P   N   W   I   S   G   Q   E   Y   L   H   Q   S   A   D   N   N     75
ATA GTA CTT TAT AAT ATT GAA ACA GGG CAA TCA TAT ACC ATT TTG AGT AAT AGA ACC ATG AAA AGT GTG AAT GCT  300
I   V   L   Y   N   I   E   T   G   Q   S   Y   T   I   L   S   N   R   T   M   K   S   V   N   A    100
TCA AAT TAC GGC TTA TCA CCT GAT CGG CAA TTT GTA TAT CTA GAA AGT GAT TAT TCA AAG CTT TGG AGA TAC TCT  375
S   N   Y   G   L   S   P   D   R   Q   F   V   Y   L   E   S   D   Y   S   K   L   W   R   Y   S    125
TAC ACA GCA ACA TAT TAC ATC TAT GAC CTT AGC AAT GGA GAA TTT GTA AGA GGA AAT GAG CTT CCT CGT CCA ATT  450
Y   T   A   T   Y   Y   I   Y   D   L   S   N   G   E   F   V   R   G   N   E   L   P   R   P   I    150
CAG TAT TTA TGC TGG TCG CCT GTT GGG AGT AAA TTA GCA TAT GTC TAT CAA AAC AAT ATC TAT TTG AAA CAA AGA  525
Q   Y   L   C   W   S   P   V   G   S   K   L   A   Y   V   Y   Q   N   N   I   Y   L   K   Q   R    175
CCA GGA GAT CCA CCT TTT CAA ATA ACA TTT AAT GGA AGA GAA AAT AAA ATA TTT AAT GGA ATC CCA GAC TGG GTT  600
P   G   D   P   P   F   Q   I   T   F   N   G   R   E   N   K   I   F   N   G   I   P   D   W   V    200
TAT GAA GAG GAA ATG CTT GCT ACA AAA TAT GCT CTC TGG TGG TCT CCT AAT GGA AAA TTT TTG GCA TAT GCG GAA  675
Y   E   E   E   M   L   A   T   K   Y   A   L   W   W   S   P   N   G   K   F   L   A   Y   A   E    225
TTT AAT GAT ACG GAT ATA CCA GTT ATT GCC TAT TCC TAT TAT GGC GAT GAA CAA TAT CCT AGA ACA ATA AAT ATT  750
F   N   D   T   D   I   P   V   I   A   Y   S   Y   Y   G   D   E   Q   Y   P   R   T   I   N   I    250
CCA TAC CCA AAG GCT GGA GCT AAG AAT CCC GTT GTT CGG ATA TTT ATT ATC GAT ACC ACT TAC CCT GCG TAT GTA  825
P   Y   P   K   A   G   A   K   N   P   V   V   R   I   F   I   I   D   T   T   Y   P   A   Y   V    275
GGT CCC CAG GAA GTG CCT GTT CCA GCA ATG ATA GCC TCA AGT GAT TAT TAT TTC AGT TGG CTC ACG TGG GTT ACT  900
G   P   Q   E   V   P   V   P   A   M   I   A   S   S   D   Y   Y   F   S   W   L   T   W   V   T    300
GAT GAA CGA GTA TGT TTG CAG TGG CTA AAA AGA GTC CAG AAT GTT TCG GTC CTG TCT ATA TGT GAC TTC AGG GAA  975
D   E   R   V   C   L   Q   W   L   K   R   V   Q   N   V   S   V   L   S   I   C   D   F   R   E    325
GAC TGG CAG ACA TGG GAT TGT CCA AAG ACC CAG GAG CAT ATA GAA GAA AGC AGA ACT GGA TGG GCT GGT GGA TTC 1050
D   W   Q   T   W   D   C   P   K   T   Q   E   H   I   E   E   S   R   T   G   W   A   G   G   F    350
TTT GTT TCA ACA CCA GTT TTC AGC TAT GAT GCC ATT TCG TAC TAC AAA ATA TTT AGT GAC AAG GAT GGC TAC AAA 1125
F   V   S   T   P   V   F   S   Y   D   A   I   S   Y   Y   K   I   F   S   D   K   D   G   Y   K    375
CAT ATT CAC TAT ATC AAA GAC ACT GTG GAA AAT GCT ATT CAA ATT ACA AGT GGC AAG TGG GAG GCC ATA AAT ATA 1200
H   I   H   Y   I   K   D   T   V   E   N   A   I   Q   I   T   S   G   K   W   E   A   I   N   I    400
TTC AGA GTA ACA CAG GAT TCA CTG TTT TAT TCT AGC AAT GAA TTT GAA GAA TAC CCT GGA AGA AGA AAC ATC TAC 1275
F   R   V   T   Q   D   S   L   F   Y   S   S   N   E   F   E   E   Y   P   G   R   R   N   I   Y    425
AGA ATT AGC ATT GGA AGC TAT CCT CCA AGC AAG AAG TGT GTT ACT TGC CAT CTA AGG AAA GAA AGG TGC CAA TAT 1350
R   I   S   I   G   S   Y   P   P   S   K   K   C   V   T   C   H   L   R   K   E   R   C   Q   Y    450
TAC ACA GCA AGT TTC AGC GAC TAC GCC AAG TAC TAT GCA CTT GTC TGC TAC GGC CCA GGC ATC CCC ATT TCC ACC 1425
Y   T   A   S   F   S   D   Y   A   K   Y   Y   A   L   V   C   Y   G   P   G   I   P   I   S   T    475
CTT CAT GAT GGA CGC ACT GAT CAA GAA ATT AAA ATC CTG GAA GAA AAC AAG GAA TTG GAA AAT GCT TTG AAA AAT 1500
L   H   D   G   R   T   D   Q   E   I   K   I   L   E   E   N   K   E   L   E   N   A   L   K   N    500
ATC CAG CTG CCT AAA GAG GAA ATT AAG AAA CTT GAA GTA GAT GAA ATT ACT TTA TGG TAC AAG ATG ATT CTT CCT 1575
I   Q   L   P   K   E   E   I   K   K   L   E   V   D   E   I   T   L   W   Y   K   M   I   L   P    525
CCT CAA TTT GAC AGA TCA AAG AAG TAT CCC TTG CTA ATT CAA GTG TAT GGT GGT CCC TGC AGT CAG AGT GTA AGG 1650
P   Q   F   D   R   S   K   K   Y   P   L   L   I   Q   V   Y   G   G   P   C   S   Q   S   V   R    550
TCT GTA TTT GCT GTT AAT TGG ATA TCT TAT CTT GCA AGT AAG GAA GGG ATG GTC ATT GCC TTG GTG GAT GGT CGA 1725
S   V   F   A   V   N   W   I   S   Y   L   A   S   K   E   G   M   V   I   A   L   V   D   G   R    575
GGA ACA GCT TTC CAA GGT GAC AAA CTC CTC TAT GCA GTG TAT CGA AAG CTG GGT GTT TAT GAA GTT GAA GAC CAG 1800
G   T   A   F   Q   G   D   K   L   L   Y   A   V   Y   R   K   L   G   V   Y   E   V   E   D   Q    600
ATT ACA GCT GTC AGA AAA TTC ATA GAA ATG GGT TTC ATT GAT GAA AAA AGA ATA GCC ATA TGG GGC TGG TCC TAT 1875
I   T   A   V   R   K   F   I   E   M   G   F   I   D   E   K   R   I   A   I   W   G   W   S   Y    625
GGA GGA TAC GTT TCA TCA CTG GCC CTT GCA TCT GGA ACT GGT CTT TTC AAA TGT GGT ATA GCA GTG GCT CCA GTC 1950
G   G   Y   V   S   S   L   A   L   A   S   G   T   G   L   F   K   C   G   I   A   V   A   P   V    650
TCC AGC TGG GAA TAT TAC GCG TCT GTC TAC ACA GAG AGA TTC ATG GGT CTC CCA ACA AAG GAT GAT AAT CTT GAG 2025
S   S   W   E   Y   Y   A   S   V   Y   T   E   R   F   M   G   L   P   T   K   D   D   N   L   E    675
CAC TAT AAG AAT TCA ACT GTG ATG GCA AGA GCA GAA TAT TTC AGA AAT GTA GAC TAT CTT CTC ATC CAC GGA ACA 2100
H   Y   K   N   S   T   V   M   A   R   A   E   Y   F   R   N   V   D   Y   L   L   I   H   G   T    700
GCA GAT GAT AAT GTG CAC TTT CAG AAC TCA GCA CAG ATT GCT AAA GCT CTG GTT AAT GCA CAA GTG GAT TTC CAG 2175
A   D   D   N   V   H   F   Q   N   S   A   Q   I   A   K   A   L   V   N   A   Q   V   D   F   Q    725
GCA ATG TGG TAC TCT GAC CAG AAC CAC GGC TTA TCC GGC CTG TCC ACG AAC CAC TTA TAC ACC CAC ATG ACC CAC 2250
A   M   W   Y   S   D   Q   N   H   G   L   S   G   L   S   T   N   H   L   Y   T   H   M   T   H    750
TTC CTA AAG CAG TGT TTC TCT TTG TCA GAC TAA                                                         2283
F   L   K   Q   C   F   S   L   S   D                                                               760
```

Fig. 2. Nucleotide and deduced amino acid sequences for the coding region of pA15. Nucleotide and amino acid sequence numbers are shown to the right. The putative transmembrane region is represented in bold characters. Potential *N*-glycosylation sites are shown as bold *italicized* characters. Nucleotide substitutions and insertions relative to the FAPα cDNA sequence [6] are underlined. Amino acids determined by sequence analysis of 3 internal peptides generated by Lys C digestion of the 97 kDa subunit [5] are underlined. The putative catalytic triad of seprase (Ser^{624}, Asp^{702}, His^{734}) and its serine protease consensus motif are represented by bold underlined characters.

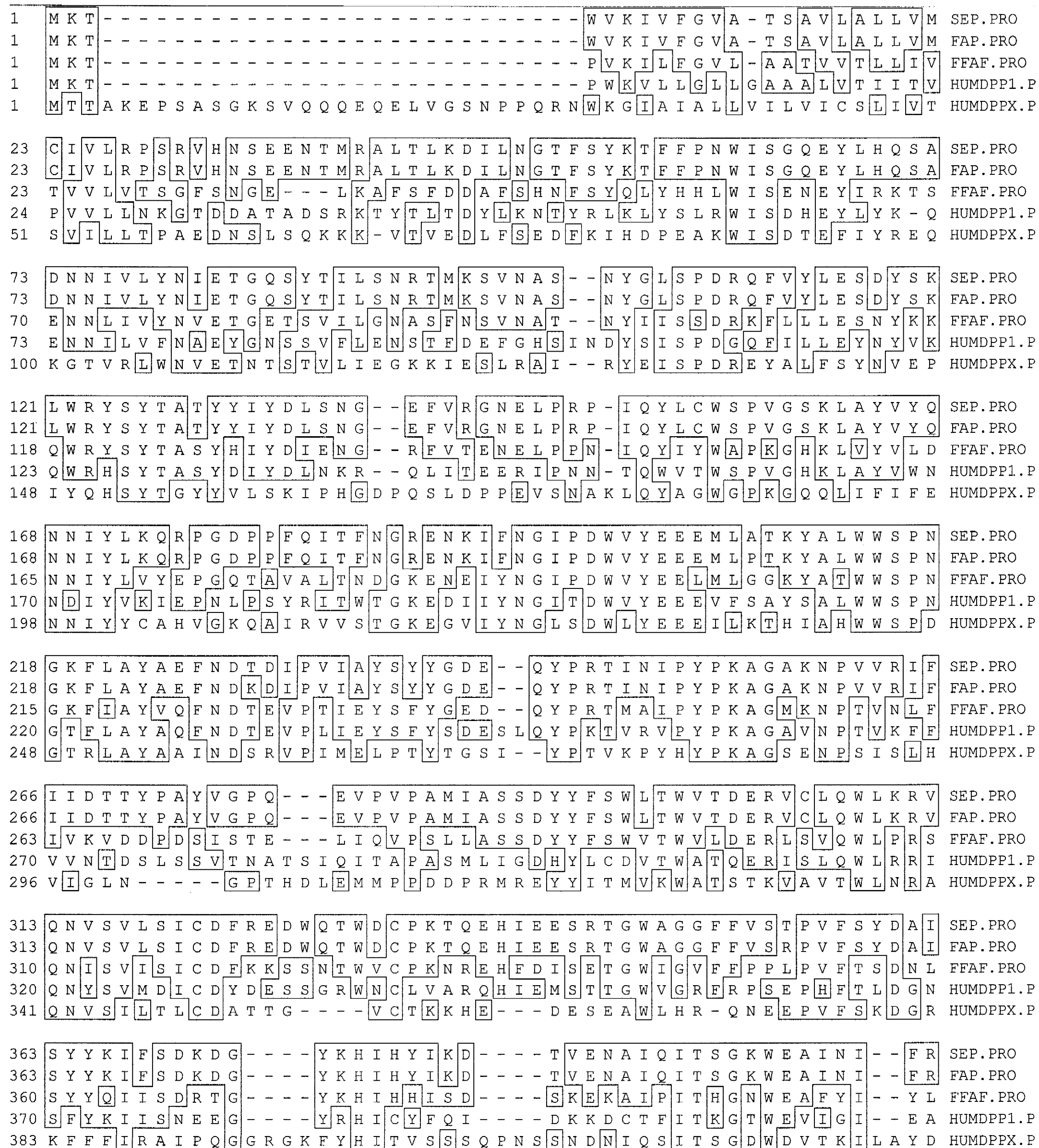


Fig. 3. Alignment of deduced amino acid sequences for seprase (SEP.PRO), human FAPα (FAP.PRO) [6], *Xenopus* fibroblast activation factor alpha (FFAF.PRO) [17], human DPPIV (HUMDPP1.P) [10] and human DPPX-S (HUMDPPX.P) [13] using the Lasergene software Megalign program (DNASTAR). Identical residues are boxed. Sequence gaps are indicated by dashed lines. Amino acid sequence numbers for each protein are shown to the left.

nucleotides at positions #1876 or 1877, 1879 or 1880 and 2010 in the pA15 cDNA. The nucleotide sequence from nucleotide #1876-G to #2010-G encodes a contiguous sequence of 45 amino acids extending from Gly^{626} to Lys^{670} in the putative catalytic region (Fig. 2). We have sequenced the region that encodes the putative catalytic region of seprase in 9 additional LOX seprase cDNA clones (including 2 clones from a lambda gt11 library) and 4 cDNA clones from cell lines other than LOX, including one clone from the WI-38 fibroblast cell line (Gly^{626} to Lys^{670} is identical to the pA15 sequence in this clone), also MDA-MB-436 (a breast carcinoma line), RPMI7951 (a melanotic melanoma line) and a clone

```
403 - V T Q D S L F Y S S N E F E E Y P G R R N I Y R I S I G S Y P P S K K C V T C H L R K E R C Q Y Y SEP.PRO
403 - V T Q D S L F Y S S N E F E E Y P G R R N I Y R I S I G S Y P P S K K C V T C H L R K E R C Q Y Y FAP.PRO
399 L V T D E A I Y Y T S T E Y E G A P G R R H I Y R I G L Q G H L R E K Q C F T C N I R K E R C Q Y Y FFAF.PRO
409 A L T S D Y L Y Y I S N E Y K G M P G G R N L Y K I Q L S D Y T K V T - C L S C E L N P E R C Q Y Y HUMDPP1.P
432 D E K G N K I Y F L S T E - - D L P R R R Q L Y S A N T E G N F - N R Q C L S C D L - V E N C T Y F HUMDPPX.P

452 T A S F S D Y A K Y Y A L V C Y G P G I P I S T L H D G R T D Q E I K I L E E N K E L E N A L K N I SEP.PRO
452 T A S F S D Y A K Y Y A L V C Y G P G I P I S T L H D G R T D Q E I K I L E E N K E L E N A L K N I FAP.PRO
449 T A R F S H K A K Y Y T L Y C Y G P G L P I V T L H E G S N N R E I R I L E D N A D L E A T L H N F FFAF.PRO
458 S V S F S K E A K Y Y Q L R C S G P G L P L Y T L H S S V N D K G L R V L E D N S A L D K M L Q N V HUMDPP1.P
478 S A S F S H S M D F F L L K C E G P G V P M V T V H N T T D K K K M F D L E T N E H V K K A I N D R HUMDPPX.P

502 Q L P K E E I K K L E V D E I T L W Y K M I L P P Q F D R S K K Y P L L I Q V Y G G P C S Q S V R S SEP.PRO
502 Q L P K E E I K K L E V D E I T L W Y K M I L P P Q F D R S K K Y P L L I Q V Y G G P C S Q S V R S FAP.PRO
499 E M P S E K I G K I V Q D D K T F W Y K M T L P P H F N A R K K Y P L L I Y V Y G G P S S Q E V K P FFAF.PRO
508 Q M P S K K L D F I I L N E T K F W Y Q M I L P P H F D K S K K Y P L L L D V Y A G P C S Q K A D T HUMDPP1.P
528 Q M P K V E Y R D I E I D D Y N L P M Q I L K P A T F T D T T H Y P L L L V V D G T P G S Q S V A E HUMDPPX.P

552 V F A V N W I S Y L A S K E G M V I A L V D G R G T A F Q G D K L L Y A V Y R K L G V Y E V E D Q I SEP.PRO
552 V F A V N W I S Y L A S K E G M V I A L V D G R G T A F Q G D K L L Y A V Y R K L G V Y E V E D Q I FAP.PRO
549 S F S V G Y S T Y L A S T E G I I V A S L D G R G T A Y Q G V I V M H A I N N R L G T V A V E D Q I FFAF.PRO
558 V F R L N W A T Y L A S T E N I I V A S F D G R G S G Y Q G D K I M H A I N R R L G T F E V E D Q I HUMDPP1.P
578 K F E V S W E T V M V S S H G A V V V K C D G R G S G F Q G T K L L H E V R R R L G L L E E K D Q M HUMDPPX.P

602 T A V R K F I E M G F I D E K R I A I W G W S Y G G Y V S S L A L A S G T G - - - - L F K C G I A V SEP.PRO
602 T A V R K F I E M G F I D E K R I A I W G W S Y - - - - - E I R F I T G P - - - - - - - - - C - - - - FAP.PRO
599 F A A R K F I A M G F I D E K R I A I W G W S Y G G Y V T S M V L G P R P R - - - - P V K C G M T L FFAF.PRO
608 E A A R Q F S K M G F V D N K R I A I W G W S Y G G Y V T S M V L G S G S G - - - - V F K C G I A V HUMDPP1.P
628 E A V R T M L K E Q Y I D R T R V A V F G K D Y G G Y L S T Y I L P A K G E N Q G Q T F T C G S A L HUMDPPX.P

648 A P V S S W E Y - - - Y A S - - - - - - - - - - V Y T E R - F M G L P T K D D N L E H Y K N S T V M SEP.PRO
635 - - I W N W S F Q M W Y S S G S S L Q L G I L R V C L H R E I H G S P N K D D N L E H Y K N S T V M FAP.PRO
645 P P V S N W E Y - - - Y A S - - - - - - - - - - I Y T E R - Y M G L P T K S D N L E N Y K N S T V M FFAF.PRO
654 A P V S R W E Y - - - Y D S - - - - - - - - - - V Y T E R - Y M G L P T P E D N L D H Y R N S T V M HUMDPP1.P
678 S P I T - - D F K L - Y A S A - - - - - - - - - - F S E R - Y L G L H G L D N R - - A Y E M T K V A HUMDPPX.P

684 A R A E Y F R N V D Y L L I H G T A D D N V H F Q N S A Q I A K A L V N A Q V D F Q A M W Y S D Q N SEP.PRO
683 A R A E Y F R N V D Y L L I H G T A D D N V H F Q N S A Q I A K A L V N A Q V D F Q A M W Y S D Q N FAP.PRO
681 T R A E Q F R K V D Y L L V H G T A D D N V H F Q Q A A Q I S K A L V E A Q V D F Q A M W Y T D K D FFAF.PRO
690 S R A E N F K Q V E Y L L I H G T A D D N V H F Q Q S A Q I S K A L V D V G V D F Q A M W Y T D E D HUMDPP1.P
712 H R V S A L E E Q Q F L I I H P T A D E K I H F Q H T A E L I T Q L I R G K A N Y S L Q I Y P D E S HUMDPPX.P

734 H G L S G L S T N - H L Y T H M T H F L K Q C F S L S D SEP.PRO
733 H G L S G L S T N - H L Y T H M T H F L K Q C F S L S D FAP.PRO
731 H S I S G S A K K - H L F T H M T H F L K N W F N L FFAF.PRO
740 H G I A S S T A H Q H I Y T H M S H F I K Q C F S L P HUMDPP1.P
762 H Y F T S S S L K Q H L Y R S I I N F F V E C F R I Q D K L P T V T A K E D E E E D HUMDPPX.P
```

Fig. 3. (continued).

from human umbilical vein smooth muscle. All exhibit the insertion of 3 guanine nucleotides that correspond with the pA15 sequence. Thus, the difference of 8 bp between the seprase and FAPα cDNAs gives rise to a total difference of 48 amino acids in their deduced amino acid sequences or an identity of ~ 94%. The significance of the 3 nonconservative amino acid substitutions remains to be determined. The substitution at position #229 (Lys to Thr) generates a potential *N*-glycosylation site in seprase (see below). However, the 3 guanine nucleotide insertions in the pA15 cDNA result in a significant divergence of the amino acid sequence (45 contiguous residues) in the putative catalytic region (Fig. 3; see below). Importantly, the consensus sequence motif Gly-X-Ser-X-Gly which is identical in seprase and the related mammalian protease DPPIV (Gly-Trp-Ser-Tyr-Gly) is reported to be Gly-Trp-Ser-Tyr-Glu in FAPα [6]. The substitution of Gly^{626} with Glu in the reported FAPα cDNA sequence is significant since it has been shown that substitution of this Gly residue in rat DPPIV results in complete loss of proteolytic activity [18] and that this mutant is retained and degraded in the endoplasmic reticulum [19,20].

Seprase is also homologous to the recently reported *Xenopus laevis* homologue of human FAPα [17]. It exhibits a 58% amino acid identity with the *Xenopus* homologue which increases to 67% in the putative catalytic region (Fig. 3). Five of the 6 *N*-glycosylation sites present in seprase are conserved in this FAPα homologue including the site at Asn^{227} that is reported not to be present in human FAPα. Seven Cys residues are conserved between seprase and the *Xenopus* homologue including 6 of 7 Cys residues present in the seprase Cys rich region. Importantly, the *Xenopus* fibroblast factor is highly homologous (70% nucleotide identity; 62% amino acid identity) to seprase in the Gly^{626} to Lys^{670} region (Fig. 3; see above). Conversely, the amino acid identity in this region between the reported sequence for human FAPα and its *Xenopus* homologue is 18% although the nucleotide identity is 69%. Both seprase and the *Xenopus* FAPα homologue have the same serine protease consensus motif (Gly-Trp-Ser-Tyr-Gly).

In addition to FAPα, there is also a striking sequence identity with the multifunctional exopeptidase DPPIV [8–11]. Like seprase, this serine protease is homodimeric consisting of 2 subunits with a M_r 110 000 to 130 000 that must be associated for activity [21,22]. However, they have different substrate specificities [2,5,9]. Both are type II integral membrane proteins that have a short cytoplasmic tail consisting of 6 amino acids followed by a transmembrane domain of 20 amino acids for seprase and 22 for DPPIV. The extracellular domain of both proteases, 734 and 738 amino acids for seprase and DPPIV, respectively, can be divided into an *N*-glycosylated membrane proximal region; a central region that is rich in Cys residues; and a carboxyl terminal region which contains the catalytic triad in the nonclassical orientation [14–16]. The sequence identity between seprase and human DPPIV [10] is 52%. However, in the putative catalytic region the identity increases to 68% (Fig. 3). The *N*-glycosylation and Cys rich regions are 49% and 47%, respectively. Four of six potential *N*-glycosylation sites in seprase are conserved in DPPIV; however, DPPIV has five additional *N*-glycosylation sites. Both have 12 Cys with 6 residues being conserved in the Cys rich region and 3 in the catalytic region. Within the 'stalk' segment that consists of the 20 amino acids that immediately follow the transmembrane domain there is only a 15% identity. This region is reported to show the greatest divergence in sequence among DPPIV-like proteins [14]. In addition, only 4 residues are conserved in the transmembrane domain (20%) but 4 of 6 amino acids in the cytoplasmic tail of both proteins are conserved.

Seprase also exhibits a 32% amino acid identity with human DPPX (DPPX short form-DPPX-S), a DPPIV related integral membrane protein of unknown function which is predominantly expressed in mammalian brain [6,12,13]. Dipeptidyl peptidase X lacks the catalytic Ser residue present in the other DPPIV related proteins (Fig. 3). Interestingly, the 6 Cys residues that are conserved between seprase and DPPIV in the Cys rich region are also conserved in DPPX and 2 of 3 Cys residues located in the catalytic region of seprase and DPPIV are conserved in DPPX. Two potential *N*-glycosylation sites present in seprase at Asn^{227} and Asn^{314} are also present in DPPIV and DPPX as well as Asp and Gly residues that correspond to Asp^{599} and Gly^{626} in seprase. These residues have been reported to be required for the transport of mouse [23] and rat DPPIV[19,20] to the cell surface, respectively.

The results from immunostaining, Western blot and gelatin zymography [5] analyses of transiently transfected COS-7 cells confirm that pA15 encodes the 170 kDa integral membrane protease originally described by Aoyama and Chen [2]. Also, our work confirms the peptide sequence data obtained from the 97 kDa subunit which indicated that seprase exhibited highest homology with human FAPα [5]. Sequence data from pA15 and several additional cDNA clones (13 total) from LOX and other cell lines including the fibroblast line WI-38 strongly suggests that seprase and FAPα are the same protein and thus products of the same gene. This conclusion is supported by the RT-PCR data obtained with FAPα primers for LOX and WI-38 RNA (Fig. 1A). The possibility that seprase and FAPα represent alternatively spliced variants that diverge in their putative catalytic regions seems unlikely based on the genomic organization of the human [24] and mouse [25] DPPIV genes. The exon-intron structure of the DPPIV genes predicts that the genomic sequence which encodes the region of divergence in seprase and FAPα (Gly^{626} to Lys^{670}) is contained within 2 con-

tiguous exons (exon 22 and 23) but the divergence in sequence does not align with the exon-intron junctions for these exons. Additionally, the high degree of sequence identity (80%) and the lack of any insertions or deletions of amino acids in the cDNA sequences of seprase and human DPPIV in the region encoded by exons 22 and 23 suggests that no intron-exon junction sliding [26,27] has occurred between the DPPIV and seprase genes in this region. Also, except for the 3 missing guanine nucleotides, the nucleotide sequence that encodes the Gly^{626} to Lys^{670} region (nt 1876-G to nt 2010-G; Fig. 2) is identical in both seprase and FAPα cDNAs. Thus, we propose that seprase/FAPα and DPPIV represent novel proteases within the nonclassical serine hydrolase family. Namely, serine integral membrane proteases (SIMP) with similar structural organizations and membrane topologies that require subunit dimerization for activation of proteolytic activity [5,21,22]. Multiple nonproteolytic functions have been ascribed to DPPIV [9,28–31]. It is interesting to speculate whether this DPPIV related protein, seprase/FAPα, may function in roles other than as an integral membrane protease. The isolation of the seprase cDNA should help elucidate what the precise role of this protein is in cancer metastasis.

We are grateful to Linda Howard for critical review of this manuscript and to the Lombardi Cancer Center Macromolecular Synthesis and Sequencing Core Facility for the synthesis of oligonucleotide primers and DNA sequencing. We also thank Dr. C. Morimoto for the pCD26 expression plasmid and Dr. S. Okada for human umbilical vein smooth muscle RNA. This work was supported by USPHS grant number R01 CA-39077 and HL-33711, and U.S. Army grant DAMD17-94-J-4033 to W.-T.C., by U.S. Army postdoctoral fellowship grant number DAMD17-94-J-4170 to L.A.G, and in part by USPHS grant number 2P30-CA-51008 to the Lombardi Cancer Center.

References

[1] W.-T. Chen, Enzyme Protein 49 (1996) 59–71.
[2] A. Aoyama, W.-T. Chen, Proc. Natl. Acad. Sci. U.S.A. 87 (1990) 8296–8300.
[3] O. Fodstad, I. Kjonniksen, S. Aamdal, J.M. Nesland, M.R. Boyd, A. Pihl, Cancer Res. 48 (1988) 4382–4388.
[4] W.L. Monsky, C.-Y. Lin, A. Aoyama, T. Kelly, S.C. Mueller, S.K. Akiyama, W.-T. Chen, Cancer Res. 54 (1994) 5702–5710.
[5] M.L. Piñeiro-Sánchez, L.A. Goldstein, J. Dodt, L. Howard, Y. Yehn, W.-T. Chen, J. Biol. Chem. 272 (1997) 7595–7601.
[6] M.J. Scanlan, B.K. Mohan Raj, B. Calvo, P. Garin-Chesa, M.P. Sanz-Moncasi, J.H. Healey, L.J. Old, W.J. Rettig, Proc. Natl. Acad. Sci. USA 91 (1994) 5657–5661.
[7] P. Garin-Chesa, L.J. Old, W.J. Rettig, Proc. Natl. Acad. Sci. USA 87 (1990) 7235–7239.
[8] Y. Misumi, Y. Hayashi, F. Arakawa, Y. Ikehara, Biochim. Biophys. Acta 1131 (1992) 333–336.
[9] B. Fleischer, Immunol. Today 15 (1994) 180–184.
[10] T. Tanaka, D. Camerini, B. Seed, Y. Torimoto, N.H. Dang, J. Kameoka, H.N. Dahlberg, S.F. Schlossman, C. Morimoto, J. Immunol. 149 (1992) 481–486.
[11] D. Darmoul, M. Lacasa, L. Baricault, D. Marguet, C. Sapin, P. Trotot, A. Barbat, G. Trugnan, J. Biol. Chem. 267 (1992) 4824–4833.
[12] K. Wada, N. Yokotani, C. Hunter, K. Doi, R.J. Wenthold, S. Shimasaki, Proc. Natl. Acad. Sci. USA 89 (1992) 197–201.
[13] N. Yokotani, K. Doi, R.J. Wenthold, K. Wada, Hum. Mol. Genet. 2 (1993) 1037–1039.
[14] D. Marguet, A.M. Bernard, I. Vivier, D. Darmoul, P. Naquet, M. Pierres, J. Biol. Chem. 267 (1992) 2200–2208.
[15] F. David, A.M. Bernard, M. Pierres, D. Marguet, J. Biol Chem. 268 (1993) 17247–17252.
[16] L. Polgar, FEBS Lett. 311 (1992) 281–284.
[17] D.D. Brown, Z. Wang, J.D. Furlow, A. Kanamori, R.A. Schwartzman, B.F. Remo, A. Pinder, Proc. Natl. Acad. Sci. USA 93 (1996) 1924–1929.
[18] S. Ogata, Y. Misumi, E. Tsuji, N. Takami, K. Oda, Y. Ikehara, Biochemistry 31 (1992) 2582–2587.
[19] T. Fujiwara, E. Tsuji, Y. Misumi, N. Takami, Y. Ikehara, Biochem. Biophys. Res. Commun. 185 (1992) 776–784.
[20] E. Tsuji, Y. Misumi, T. Fujiwara, N. Takami, S. Ogata, Y. Ikehara, Biochemistry 31 (1992) 11921–11927.
[21] G. Püschel, R. Mentlein, E. Heymann, Eur. J. Biochem. 126 (1982) 359–365.
[22] E.F.J. Walborg, S. Tsuchida, D.S. Weeden, M.W. Thomas, A. Barrick, K.D. McEntire, J.P. Allison, D.C. Hixson, Exp. Cell Res. 158 (1985) 509–518.
[23] F. David, L. Baricault, C. Sapin, X. Gallet, D. Marguet, A. Thomas-Soumarmon, G. Trugnan, Biochem. Biophys. Res. Commun. 222 (1996) 833–838.
[24] C.A. Abbott, E. Baker, G.R. Sutherland, G.W. McCaughan, Immunogenetics 40 (1994) 331–338.
[25] A.-M. Bernard, M. Marie-Genevieve, M. Pierres, D. Marguet, Biochemistry 33 (1994) 15204–15214.
[26] C.S. Craik, W.J. Rutter, R. Fletterick, Science 220 (1983) 1125–1129.
[27] C.S. Craik, Q.-L. Choo, G.H. Swift, C. Quinto, R.J. Mac-

Donald, W.J. Rutter, J. Biol. Chem. 259 (1984) 14255–14264.
[28] G.A. Piazza, H.M. Callanan, J. Mowery, D.C. Hixson, Biochem. J. 262 (1989) 327–334.
[29] N.H. Dang, Y. Torimoto, K. Sugita, J.F. Daley, P. Schow, C. Prado, S.F. Schlossman, C. Morimoto, J. Immunol. 145 (1990) 3963–3971.
[30] R.C. Johnson, D. Zhu, H.G. Augustin-Voss, B.U. Pauli, J. Cell Biol 121 (1993) 1423–1432.
[31] J. Kameoka, T. Tanaka, Y. Nojima, S.F. Schlossman, C. Morimoto, Science 261 (1993) 466–469.