Eur. J. Biochem. *254*, 650−654 (1998)

Mouse fibroblast-activation protein
Conserved *Fap* gene organization and biochemical function as a serine protease

Joachim NIEDERMEYER, Barbara ENENKEL, John E. PARK, Martin LENTER, Wolfgang J. RETTIG, Klaus DAMM and Andreas SCHNAPP
Department of Oncology Research, Boehringer Ingelheim Pharma KG, Biberach an der Riss, Germany

(Received 6 February 1998) − EJB 98 0182/2

The human fibroblast-activation protein (FAP), a member of the serine protease family, was discovered as an inducible type-II cell-surface glycoprotein selectively expressed by reactive stromal fibroblasts of epithelial cancers and healing wounds. Antibodies directed against human FAP have a clinical use for antibody-based tumor imaging. As part of an effort to generate animal models of FAP expression in epithelial tumorigenesis and wound healing, we previously cloned the cDNA encoding the mouse FAP homolog. In this study, we used PCR/restriction-fragment length polymorphism, identified in interspecific back-crosses between *Mus musculus* and *Mus spretus*, to map the *Fap* gene locus to a region of mouse chromosome 2, known to be syntenic to the previously identified *FAP* gene locus on human chromosome 2q23. The *Fap* gene spans approximately 60 kb and contains 26 exons ranging in size from 46 bp to 195 bp. This genomic organization is very similar to that of the human FAP locus. Similar to the gene encoding dipeptidyl peptidase IV (DPP IV), the nucleotides encoding the serine protease consensus motif, WGWSYGG, are split between two exons, a feature distinct from classical serine proteases. Consistent with the similarity to DPP IV, a chimeric FAP fusion protein expressed in a baculovirus system has dipeptidyl peptidase activity.

Keywords: type-II integral membrane protein; dipeptidyl peptidase; serine protease; gene organization; gene localization.

The invasive growth of epithelial neoplasms is associated with characteristic changes in the supporting stroma. For example, epithelial cancers induce the formation of tumor blood vessels, the recruitment of reactive tumor stromal fibroblasts, lymphoid and phagocytic infiltrates, the release of peptide mediators and proteolytic enzymes, and the production of an altered extracellular matrix [1−3]. Reactive tumor stromal fibroblasts are not transformed but show characteristic patterns of gene expression not found in resting fibrocytes of normal adult tissues [2, 4−7]. A highly consistent trait of tumor stromal fibroblasts is the induction of fibroblast-activation protein (FAP), a cell-surface-bound member of the serine protease family. A potential role for the FAP protease in extracellular-matrix degradation or growth-factor activation has been suggested [8]. In humans, FAP expression is detected in the stroma of over 90% of malignant breast, colorectal, lung, skin and pancreatic tumors, and in fibroblasts of healing wounds and in some fetal mesenchymal cells [8, 9]. In contrast, most normal adult tissues and benign epithelial tumors show little or no detectable human FAP expression.

Human FAP is a 95-kDa type-II integral membrane protein able to form homomeric high-molecular-mass complexes on cells [10]. Human FAP shows 48% amino acid identity to the T-cell activation antigen CD26, also known as dipeptidyl peptidase (DPP) IV [8]. In contrast to the restricted expression pattern observed with human FAP, DPP IV is widely expressed [11]. The three catalytic residues (His, Asp and Ser) common to DPP IV and other serine proteases are conserved in human FAP, suggesting a similar function as a proteolytic enzyme [8].

To investigate the function of FAP during embryogenesis and epithelial tumorigenesis, the mouse was selected as a model system, since it is readily amenable to genetic manipulations and *in vivo* experimentation. We recently cloned the cDNA encoding mouse FAP, which shares 89% amino acid sequence identity with the human protein, including a conserved catalytic triad [12]. Mouse FAP transcripts can be detected in cultured mouse embryo fibroblasts and embryonic tissues. In addition, the host-derived fibroblast-rich stroma of human epithelial-cancer xenografts grown in immunodeficient mice expresses mouse FAP [12]. To obtain more information about possible regulatory mechanisms involved in the restriction of FAP expression to certain stages of embryogenesis and tumorigenesis, we isolated and characterized the genes encoding human and mouse FAP. Additionally, we mapped the chromosomal localization of the mouse *Fap* gene. To confirm the enzymatic function of FAP, a chimeric FAP fusion protein was expressed in a baculovirus expression system and tested for DPP activity.

Correspondence to A. Schnapp, Boehringer Ingelheim Pharma KG, Birkendorfer Strasse 65, D-88397 Biberach an der Riss, Germany
Fax: +49 7351 54 5146.
E-mail: andreas.schnapp@bid.de

Abbreviations. FAP, fibroblast-activation protein; *Fap*, mouse FAP gene locus; *FAP*, human FAP gene locus; DPP, dipeptidyl peptidase; NH-F_3-Mec, 7-amido-4-trifluoromethyl-coumarin; UTR, untranslated region; RFLP, restriction-fragment-length polymorphism.

Note. The nucleotide sequence data for human FAP and mouse Fap are available from the Genbank/EMBL database under accession numbers U09278 and Y10007, respectively.

EXPERIMENTAL PROCEDURES

Screening of genomic libraries. A mouse λ genomic library (129/Sv mouse strain; Stratagene) was screened with either a full-length mouse FAP cDNA or a 0.4-kb *Bam*HI−*Xba*I frag-

ment of the 5′ end of the mouse FAP cDNA [8], radioactively labeled with [α-^{32}P]dCTP, as a probe. To isolate the human *FAP* gene, a human placental genomic library from Clontech was screened using the human FAP full-length cDNA as a probe. Positive clones were plaque purified to homogeneity and the DNA was analyzed by restriction enzyme digestion using *Sac*I, *Xba*I, *Bam*HI and *Eco*RI. Southern blots were hybridized to mouse and human FAP cDNA probes. Positive restriction fragments were subcloned into pBluescript II KS (+) (Stratagene) and sequenced by the dideoxynucleotide chain-termination method [13] to determine the exon-intron boundaries. In some instances the subcloned fragments were used as probes to confirm the overlap between adjoining clones.

PCR analysis. PCR amplification of selected genomic regions was performed as described [14] using Taq polymerase (Boehringer Mannheim; Perkin Elmer), or the Expand PCR system (Boehringer Mannheim).

Genomic location of Fap by restriction-fragment-length polymorphism (RFLP) analysis. 50−100 ng genomic DNA isolated from C57BL/6J mice or from *Mus spretus*, obtained from the Jackson Laboratory, were used as templates in PCR reactions with two primers from the 5′ untranslated region of the mouse FAP cDNA (forward primer hsFAP17, 5′-CTAATTTTACAGAAATCTTTTGAAACTTGGC-3′; reverse primer mFAP67, 5′-CTTCATTTTTCCAGATGTTTTTGC-3′). From the genomic DNA of both strains a 200-bp fragment was PCR amplified and sequenced. Due to a single nucleotide exchange between the two amplicons, an *Mbo*II restriction site is only present in the C57BL/6J genomic DNA. Using this polymorphism, the two Jackson laboratory interspecific back-cross panels BSS [(C57BL/6JEi * SPRET/Ei)$_{F1}$ * SPRET/Ei] and BSB [(C57BL/6J * *M. spretus*) * C57BL/6J] were screened and the localization of the *Fap* gene locus was determined as described [15, 16]. The raw data sets can be obtained via the Internet at http://www.jax.org/resources/documents/cmdata.

Baculovirus expression. DNAs encoding the mouse CD8 protein [17] or fusion proteins between the extracellular domain of mouse CD8 (amino acids 1−189) and the extracellular portion of mouse FAP (amino acids 27−761) or human FAP (amino acids 27−760) were cloned into the vector pVL1393 (Invitrogen), resulting in the plasmids pmCD8, pmCD8-mFAP and pmCD8-hsFAP, respectively. These constructs were transfected with viral DNA into Sf9 cells (Invitrogen) using the BaculoGold system according to the manufacturer's protocol (Pharmingen). After 4 days the Sf9 cell supernatants were collected and used to infect HighFive cells (Invitrogen) grown in serum-free medium in a first amplification step. After five rounds of amplification the HighFive cell supernatants were collected and analysed for protein expression and activity.

PAGE. SDS/PAGE was performed according to Laemmli [18] in the absence of reducing agents. Apparent molecular masses were determined by coelectrophoresis of the marker proteins myosin (200 kDa), β-galactosidase (116 kDa), phosphorylase *b* (97.5 kDa), BSA (66 kDa) and ovalbumin (45 kDa) (Bio-Rad).

Immunoblotting. For immunoblotting studies, proteins separated by SDS/PAGE were transferred to nitrocellulose membranes (Amersham), blocked for 1 h with 4% milk powder in 0.21 g/l KH_2PO_4, 9 g/l NaCl, 0.726 g/l $Na_2HPO_4 \cdot 7\ H_2O$ (NaCl/P_i), and incubated with the rat anti-CD8 Ig 5H10-1 (Pharmingen; 10 µg/ml), at 4°C in NaCl/P_i containing 4% milk powder. After washing with NaCl/P_i, immunoreactive proteins were visualized with peroxidase-conjugated goat anti-rat Ig (Dianova; 1:5000) using the enhanced chemiluminescence system according to the manufacturer's protocol (Amersham Buchler).

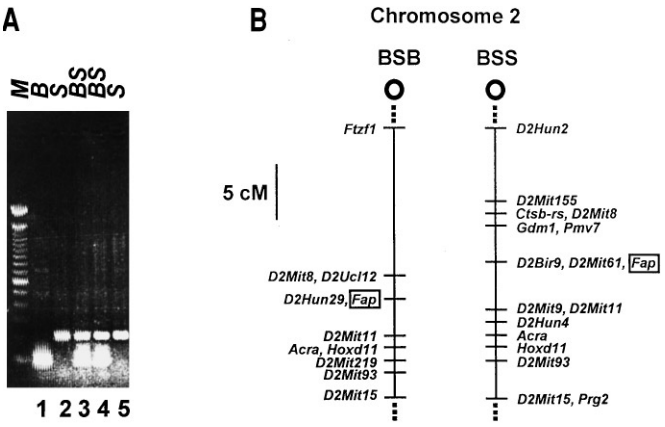

Fig. 1. Genomic mapping of mouse *Fap*. (A) Genomic DNA isolated from *M. musculus* strain C57BL/6J (B, lane 1), *M. spretus* (S, lane 2), or from interspecific backcross animals (lanes 3 to 5) was PCR amplified and digested by *Mbo*II as described in Materials and Methods. Due to a single nucleotide exchange between the two amplicons, an *Mbo*II restriction site is present in the C57BL/6J genomic DNA, whereas the *M. spretus* DNA fragment is not digested by this enzyme. (B) Assignment of mouse *Fap* locus to mouse chromosome 2 by backcross analysis. Using the backcross panels BSB and BSS in combination with a PCR-based genetic marker assay (A), the *Fap* locus was assigned to the mid-portion of mouse chromosome 2. The maps are depicted with the centromere toward the top. The scale bar represents 5 cM.

DPP assay. For immunopurification of CD8, CD8-mouse-FAP, or CD8-human-FAP, 96-well plates (Canberra Packard) were coated with the rat anti-CD8 mAb 5H10-1 (500 ng/well; Pharmingen) for 2 h at 37°C. After washing twice with 1 ml NaCl/P_i, the plates were blocked with NaCl/P_i containing 4% fetal calf serum for 1 h at room temperature. Supernatants obtained from SF9 cells infected with baculoviruses encoding CD8, CD8-(mouse FAP) or CD8-(human FAP) were added and incubated with the precoated 5H10-1 antibody for 2 h at room temperature. The amount of supernatants corresponded to 5, 14 and 28 ng of each recombinant protein, respectively. After four washing steps in NaCl/P_i, the DPP assay was performed in the presence of 0.5 µM of the substrate alanylprolyl-7-amido-4-trifluoromethyl-coumarin (Ala-Pro-NH-F_3-Mec; Bachem) for 1 h at 37°C in 100 mM Tris/HCl, pH 7.8, 100 mM NaCl, followed by fluorometric measurement (SLT Fluostar; excitation wavelength 390 nm, emission wavelength 538 nm).

Computer analysis. Sequence analysis was performed by means of the Genetic Data Environment version 2.2. For synteny analysis, the NCBI GenBank data (http://www3.ncbi.nlm.nih.gov./Synteny/human2.html and http://www3.ncbi.nlm.nih.gov/Homology/mouse2.html) were used.

RESULTS

Chromosomal location of the mouse *Fap* gene. To obtain the chromosomal localization of the mouse *Fap* gene, the two Jackson laboratory interspecific back-cross panels BSS and BSB [19] were screened by a *Mbo*II-RFLP analysis. Statistical analysis of the data obtained from the two panels was performed by determining the degree of matching among the typings for all animals and all loci and by arranging the loci in linkage groups, ordered to minimize apparent double cross-overs with known loci. Thus an overall representation of the entire chromosome complement of each N2 animal could be examined. In both back-cross panels the *Fap* gene mapped to the mid-portion of mouse chromosome 2, approximately 5 cM proximal to the loci for acetylcholine

Exon	Species	Exon size	Acceptor site	Donor site	Intron size
E1	mFAP			ATG AAG**gt**gtat	0.4 kb
	hsFAP			ATG AAG**gt**tagt	0.3 kb
E2	mFAP	85 bp	atgt**ag**ACA TGG CTG A	CCC TCA AGA G**gt**aaca	6.8 kb
	hsFAP		atgt**ag**ACT TGG GTA A	CCT TCA AGA G**gt**aaga	15 kb
E3	mFAP	99 bp	ccag**ag**TT TAC AAA CC	TGG ATT TCA G**gt**aaat	1.4 kb
	hsFAP		ccaa**ag**TT CAT AAC TC	TGG ATT TCA G**gt**aagt	1 kb
E4	mFAP	95 bp	ttaa**ag**AA CAA GAA TA	T AGC ACC ATG**gt**atgt	1.0 kb
	hsFAP		ttaa**ag**GA CAA GAA TA	T AGA ACC ATG**gt**atgt	0.9 kb
E5	mFAP	75 bp	tttc**ag**AAA AGT GTG A	T TAT TCA AAG**gt**atca	1.5 kb
	hsFAP		tttc**ag**AAA AGT GTG A	T TAT TCA AAG**gt**atca	0.8 kb
E6	mFAP	53 bp	taac**ag**CTC TGG CGA T	TT CAG AAT GG**gt**aaaa	4 kb
	hsFAP		tgac**ag**CTT TGG AGA T	TT AGC AAT GG**gt**aaat	7 kb
E7	mFAP	73 bp	ttct**ag**G GAA TTT GTA	G AGT AAA TTA**gt**aagt	0.8 kb
	hsFAP		ttct**ag**A GAA TTT GTA	G AGT AAA TTA**gt**aagt	0.7 kb
E8	mFAP	121 bp	tttc**ag**GCA TAT GTA T	GTT TAT GAA G**gt**aagc	1.3 kb
	hsFAP		ttttag**GCA TAT GTC T	GTT TAT GAA G**gt**aagc	0.9 kb
E9	mFAP	155 bp	tttc**ag**AG GAA ATG CT	G TAT CCA AAG**gt**atgt	1.7 kb
	hsFAP		tttc**ag**AG GAA ATG CT	A TAC CCA AAG**gt**atgt	2.2 kb
E10	mFAP	104 bp	tggc**ag**GCT GGG GCT A	TA GCC TCA AG**gt**caqt	2.1 kb
	hsFAP		tggc**ag**GCT GGA GCT A	TA GCC TCA AG**gt**agtt	2.1 kb
E11	mFAP	136 bp	tggc**ag**T GAC TAT TAT	A TGT CCA AAG**gt**aaaa	5.5 kb
	hsFAP		ttgc**ag**T GAT TAT TAT	T TGT CCA AAG**gt**aggg	4.5 kb
E12	mFAP	45 bp	ttgt**ag**AAC CAG GAG C	G GCT GGT GGA**gt**atga	3.5 kb
	hsFAP		ttttag**ACC CAG GAG C	G GCT GGT GGA**gt**atga	6.5 kb
E13	mFAP	105 bp	ttca**ag**TTC TTT GTT T	A GAC ACT GTG**gt**acgt	0.2 kb
	hsFAP		tttc**ag**TTC TTT GTT T	A GAC ACT GTG**gt**acgt	0.1 kb
E14	mFAP	68 bp	aataagGAA AAT GCT A	AG GAT TCA CT**gt**aagt	2.2 kb
	hsFAP		ttat**ag**GAA AAT GCT A	AG GAT TCA CT**gt**aagt	2.3 kb
E15	mFAP	54 bp	tttc**ag**G TTT TAT TCT	AC ATC TAC AG**gt**aatt	2 kb
	hsFAP		tttc**ag**G TTT TAT TCT	AC ATC TAC AG**gt**aatt	1.9 kb
E16	mFAP	125 bp	ctct**ag**A ATT AGC ATT	GTC TGC TAT G**gt**aggt	4 kb
	hsFAP		ctct**ag**A ATT AGC ATT	GTC TGC TAC G**gt**tggt	2.8 kb
E17	mFAP	48 bp	acac**ag**GC CCT GGC CT	ACA GAC CAA G**gt**actc	4.9 kb
	hsFAP		ccac**ag**GC CCA GGC AT	ACT GAT CAA G**gt**actc	4.9 kb
E18	mFAP	99 bp	ttaa**ag**AA ATA CAA GT	GGG GGA CTG A**gt**aaga	0.5 kb
	hsFAP		tttc**ag**AA ATT AAA AT	GAT GAA ATT A**gt**aaga	0.5 kb
E19	mFAP	70 bp	tcaa**ag**CT TTC TGG TA	TA ATT CAA GT**gt**atat	1 kb
	hsFAP		tcaa**ag**CT TTA TGG TA	TA ATT CAA GT**gt**atgt	0.7 kb
E20	mFAP	195 bp	ctttag**G TAT GGT GGT	CA GCT GTC AG**gt**gagc	0.7 kb
	hsFAP		ctttag**G TAT GGT GGT	CA GCT GTC AG**gt**gagc	5 kb
E21	mFAP	55 bp	cttc**ag**A AAA TTC ATA	A TGG GGC TGG**gt**gagt	0.3 kb
	hsFAP		ctttag**A AAA TTC ATA	A TGG GGC TGG**gt**gagt	8 kb
E22	mFAP	100 bp	ctac**ag**TCC TAC GGA G	GAA TAT TAC G**gt**atgg	1.3 kb
	hsFAP		ctac**ag**TCC TAT GGA G	GAA TAT TAC G**gt**atgg	1 kb
E23	mFAP	65 bp	ttac**ag**CA TCT ATC TA	A CAC TAT AAA**gt**aagc	0.6 kb
	hsFAP		ctat**ag**CG TCT GTC TA	G CAC TAT AAG**gt**aagt	0.5 kb
E24	mFAP	73 bp	ttcc**ag**AAT TCA ACT G	ACA GCA GAT G**gt**gagg	0.4 kb
	hsFAP		ttcc**ag**AAT TCA ACT G	ACA GCA GAT G**gt**gagg	0.3 kb
E25	mFAP	74 bp	ttgc**ag**AT AAT GTG CA	C CAG GCG ATG**gt**acat	1.4 kb
	hsFAP		ttgc**ag**AT AAT GTG CA	C CAG GCA ATG**gt**acat	1.8 kb
E26	mFAP		ttgc**ag**TGG TAC TCT G	TCA GAC TGAaccaat	
	hsFAP		ttcc**ag**TGG TAC TCT G	TCA GAC TAAaaacga	

Fig. 2. Intron-exon organization and splice-junction sequences of the human and mouse *FAP* genes. The nucleotide sequences of exon-intron junctions were determined from genomic subclones by means of oligonucleotides corresponding to cDNA sequences as primers. Exon and intron sequences at the splice acceptor and donor sites are shown in uppercase and lower-case letters, respectively. The sizes of the individual exons and introns are indicated. mFAP, mouse *Fap*; hsFAB, human *FAP*.

receptor alpha (*Acra*) and *Hoxd11* and roughly 5 cM distal to the *Gdm1* (FAD-linked glycerol-3-phosphate dehydrogenase) locus (Fig. 1).

Structural analysis of the mouse and human FAP genes. The full-length mouse FAP cDNA was used as a probe to isolate 20 λ phage clones from a 129/Sv mouse genomic library. A preliminarily clustering of these clones was achieved by typing with oligonucleotide probes covering distinct regions of the mouse FAP cDNA. Seven independent phage clones spanning the mouse FAP coding region were used for further analysis.

To isolate the human *FAP* gene, a human placental genomic library was screened with the full-length human FAP cDNA and seven phage clones were isolated, purified and clustered by means of probes covering distinct regions of the human FAP cDNA. Further analysis focused on five independent phages containing various coding segments from the known cDNA sequence.

Restriction fragments positive for mouse and human FAP coding sequences were subcloned, and partially sequenced to localize exons. Genomic regions that could not be detected in the phage clones were amplified from genomic DNA using

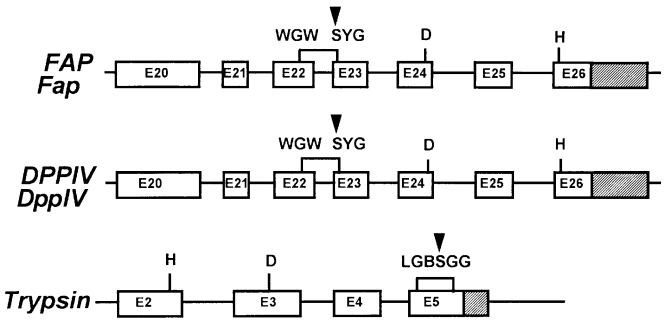

Fig. 3. Comparison of the mouse and human FAP gene structures with those of other serine hydrolases. The gene segments analyzed, corresponding to the catalytic domains of human and mouse FAP, are compared with those of human and mouse *DPP IV* [21, 22] and rat *trypsin* [27]. Exons are represented as open boxes, the non-coding regions as hatched boxes, and the introns (not to scale) as lines. The locations of amino acids forming the catalytic triad of mouse FAP, Ser624 (S), Asp702 (D), and His734 (H), are indicated.

exon-specific primers and long-range PCR techniques. All exon-intron boundaries were sequenced bidirectionally and intron sizes were determined by restriction mapping analysis and by PCR amplification with primers from adjacent exons.

The results from these analyses showed that the *Fap* gene spans approximately 60 kb from the 5′ untranslated region (UTR) to the 3′ UTR. The *FAP* gene is approximately 70 kb. The number and length of the exons and the positions of the exon-intron boundaries are conserved between the human and the mouse genes (Fig. 2). Both genes are composed of 26 exons that range in size from 46 bp to 195 bp. The exons are separated by introns that vary in length from 96 bp to more than 14 kb. All splice junctions obey the GT-AG rule [20].

The 5′ untranslated region and the first two amino acids of FAP are encoded by exon 1. Exon 2 comprises the hydrophobic amino acids forming the transmembrane domain. Exons 3−26 encode the large extracellular region of FAP. The domains of human and mouse FAP thought to be directly involved in serine protease activity are encoded by exons 21−26. Similar to other serine proteases, such as trypsin, the catalytic-triad residues of mouse FAP are encoded by separate exons (Ser624 by exon 22, Asp702 by exon 24, and His734 by exon 26; Fig. 3). A further comparison with the gene structure of other serine hydrolases revealed a close similarity between the genes encoding mouse and human FAP and those encoding mouse and human DPP IV/CD26 [21, 22]. The number and size of the exons and the intron sizes are remarkably conserved between the *FAP* and *DPP IV* genes in both species. In particular, in all four genes the serine consensus motif, XGXSXG, is split between two exons. This feature is not shared by other classical serine proteases, such as trypsin (Fig. 3).

Recombinant FAP displays DPP activity. Although the close similarity between DPP IV and human and mouse FAP suggests a similar biochemical activity, experimental evidence for a functional DDP activity of FAP was missing. To address this question, mouse and human FAP were expressed as CD8-FAP fusion proteins in insect cells. The CD8-(human FAP) and CD8-(mouse FAP) cDNAs encode chimeric 110-kDa proteins comprising the 85-kDa extracellular domain of either mouse FAP (amino acids 27−761) or human FAP (amino acids 27−760), fused C-terminally to the 25-kDa extracellular fragment of mouse CD8 (amino acids 1−189). The CD8 part contains a leader peptide for efficient protein secretion, a dimerization motif, and provides a bait for immunopurification.

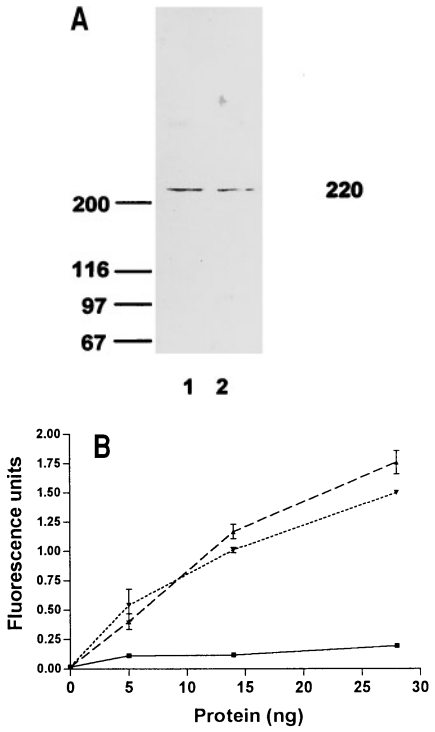

Fig. 4. FAP fusion proteins have DPP activity. (A) Expression of CD8-FAP fusion proteins in the baculovirus system. Supernatants obtained from HighFive cells infected with baculoviruses encoding CD8-(mouse FAP) or CD8-(human FAP) were separated by SDS/PAGE under non-reducing conditions. The presence of the fusion proteins was detected by western blot analysis using the anti-CD8 mAb, 5H10-1. Lane 1, 30 µl supernatant obtained from cells infected with the CD8-(mouse FAP) baculovirus; lane 2, 30 µl supernatant isolated from cells infected with the CD8-(human FAP) baculovirus. The positions of molecular-size markers are shown on the right (in kDa). (B) Immunocapture assay for DPP activity. 5, 14 or 28 ng CD8-(mouse FAP) (▲), CD8-(human FAP) (▼), or CD8 (■) were immunopurified with the anti-CD8 mAb 5H10-1 as described in Materials and Methods. The immunocomplexes were incubated with the synthetic peptide Ala-Pro-NH-F_3Mec. Release of NH_2-F_3Mec was detected at 538 nm after 1 h incubation.

Western blot analysis of supernatants obtained from HighFive cells, infected either with baculoviruses encoding CD8-(human FAP) or CD8-(mouse FAP), was performed under non-reducing conditions. In both cases, a 220-kDa protein band was recognized by the anti-CD8 mAb, 5H10-1 (Fig. 4A), whereas no protein band was detectable in supernatants from uninfected cells (data not shown). The 5H10-1-immunoreactive bands were shifted under reducing conditions to an apparent molecular mass of 110 kDa, indicating that the CD8-FAP fusion proteins are secreted as homodimers (data not shown).

The putative protease activity of the fusion proteins was analyzed by incubating different amounts of immunopurified CD8, CD8-(human FAP), and CD8-(mouse FAP) with the synthetic dipeptide Ala-Pro-NH-F_3-Mec. Release of the leaving group, NH_2-F_3-Mec, due to enzymatic cleavage was detected fluorometrically. Immunopurified CD8-(human FAP) and CD8-(mouse FAP) cleaved the Ala-Pro-NH-Mec in a dose-dependent manner, whereas mCD8 exhibited no activity.

DISCUSSION

Serine proteases form a large family of diverse peptide hydrolases that can be grouped on the basis of functional and structural properties. FAP was identified recently as a cell-surface-bound member of this family. Amino acid and cDNA sequence comparisons revealed significant similarity to the T-cell activation antigen CD26, also known as DPP IV [8]. This similarity is substantiated by our analysis of the genomic organization, localization and function of the human and mouse *FAP* genes.

The number and size of exons and introns, and thus the overall size of the FAP gene, is highly conserved between the human and mouse genomes. Human and mouse FAP genes both contain 26 exons with identically located exon-intron boundaries. In a previous study, we identified splice variants of mouse FAP expressed during embryonic development [12]. The comparison of the cDNA and genomic sequences suggests that alternative use of splice donor sites is responsible for generating the different transcripts. In the case of one alternative transcript, mouse $FAP^{\Delta 33}$, all of exon 3 is missing, leading to an internally deleted protein lacking amino acids 31−63. Another alternative transcript, mouse $FAP^{\Delta 5}$, which is lacking amino acids 31−35, is apparently generated by the use of a cryptic splice donor site within exon 3. No alternative processing has been observed for human FAP.

Using an *Mbo*II polymorphism in *M. spretus* and *M. musculus*, the *Fap* gene has been mapped to the mid-portion of mouse chromosome 2, a region known to be syntenic to human 2q23, the locus of human FAP [23].

The sequence similarity and similar architecture of the genes encoding FAP and CD26/DPP IV suggest that both arose from the duplication of an ancestral gene. This notion is supported by the finding that the CD26 and FAP genes map close to each other in two species, namely to human chromosome 2p23-qter [23, 24] and to the mid portion of mouse chromosome 2 [22]. In contrast, DPP X, another structural relative of FAP and CD26, is encoded by a gene on human chromosome 7 [25].

Similar to other serine proteases, FAP and CD26 have a catalytic serine residue arranged within the consensus sequence Gly-Xaa-Ser-Xaa-Gly. Unique to these two genes is the split of this motif into two exons. Gly-Trp is located at the very end of exon 21 and Ser-Tyr-Gly at the beginning of exon 22. This arrangement differs from the typical serine proteases where the complete serine consensus site is encoded within one exon.

Previous biochemical studies revealed that DPP IV cleaves N-terminal dipeptides from polypeptides with L-proline or L-alanine at the penultimate position [26]. To extend the similarity of FAP with serine proteases to a functional level, recombinant FAP was assayed for the presence of a DPP activity. For this purpose, mouse and human FAP were expressed in the baculovirus system as chimeric CD8-FAP fusion proteins. The ability of immunopurified chimeric CD8-FAP to cleave a Ala-Pro-NH-F_3 Mec substrate clearly reveals a dipeptidyl peptidase activity for the FAP enzyme. In contrast to the widely expressed *CD26/DPP IV* gene, expression of *FAP* is tightly regulated, with evidence of a induction on the surface of reactive stromal fibroblasts of epithelial tumors and healing wounds. These differences may be due to distinct promoter and enhancer elements located in the 5′ flanking region of these genes. Based on the present study, a detailed functional characterization of the FAP-promoter region is warranted.

We thank Mr B. Guilliard, Ms I. Lauritsch and Ms S. Zeiler for expert technical assistance, Drs U. Bamberger and P. Garin-Chesa for valuable contributions, and Profs H. Kleinig and R. Werner for their support and encouragement.

REFERENCES

1. Dvorak, H. (1986) Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing, *N. Engl. J. Med. 315*, 1650−1659.

2. Liotta, L. A., Steeg, P. S. & Stetler-Stevenson, W. G. (1991) Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation, *Cell 64*, 327−336.
3. Sieweke, M. H. & Bissell, M. J. (1994) The tumor-promoting effect of wounding: a possible role for TGF*β*-induced stroma alterations, *Crit. Rev. Oncog. 5*, 297−311.
4. Basset, P., Belloq, J. P., Wolf, C., Stoll, I., Hutin, P., Limacher, J. M., Podhajcer, O. L., Chenard, M. P., Rio, M. C. & Chambon, P. (1990) A novel metalloproteinase gene specifically expressed in stromal cells of breast carcinomas, *Nature 348*, 699−704.
5. Pyke, C., Ralfkiaer, E., Tryggvason, K. & Danø, K. (1993) Messenger RNA for two type IV collagenases is located in stromal cells in human colon cancer, *Am. J. Pathol. 142*, 359−365.
6. Cullen, K. J., Smith, H. S., Hill, S., Rosen, N. & Lippman, M. E. (1991) Growth factor messenger RNA expression by human breast fibroblasts from benign and malignant lesions, *Cancer Res. 51*, 4978−4985.
7. Sappino, A. P., Schürich, W. & Gabbiani, G. (1990) Differentiation repertoire of fibroblastic cells: expression of cytoskeletal proteins as marker of phenotypic modulation, *Lab. Invest. 63*, 144−152.
8. Scanlan, M. J., Mohan Raj, B. K., Calvo, B., Garin-Chesa, P., Sanz-Moncasi, M. P., Healey, J. H., Old, L. J. & Rettig, W. J. (1994) Molecular cloning of fibroblast activation protein, a member of the serine protease family selectively expressed in stromal fibroblasts of epithelial cancers, *Proc. Natl Acad. Sci. USA 91*, 5657−5661.
9. Garin-Chesa, P., Old, L. J. & Rettig, W. J. (1990) Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody-target in human epithelial cancers, *Proc. Natl Acad. Sci. USA 87*, 7235−7239.
10. Rettig, W. J., Su, S. L., Fortunato, S. R., Scanlan, M. J., Mohan Raj, B. K., Garin-Chesa, P., Healey, J. H. & Old, L. J. (1994) Fibroblast activation protein: purification, epitope mapping and induction by growth factors, *Int. J. Cancer 58*, 385−392.
11. Hegen, M., Niedobitek, G., Klein, C. E., Stein, H. & Fleischer, B. (1990) The T cell triggering molecule Tp103 is associated with dipeptidyl aminopeptidase IV activity, *J. Immunol. 144*, 2908−2914.
12. Niedermeyer, J., Scanlan, M. J., Garin-Chesa, P., Daiber, C., Fiebig, H. H., Old, L. J., Rettig, W. J. & Schnapp, A. (1997) Mouse fibroblast activation protein: molecular cloning, alternative splicing and expression in the reactive stroma of epithelial cancers, *Int. J. Cancer 71*, 383−389.
13. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) DNA sequencing with chainterminating inhibitors, *Proc. Natl Acad. Sci. USA 74*, 5463−5367.
14. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) Enzymatic amplification of *β*-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia, *Science 230*, 1350−1354.
15. McCarthy, L., Hunter, K., Schalkwyk, L., Riba, L., Anson, S., Mott, R., Newell, W., Bruley, C., Bar, I., Ramu, E., Housman, D., Cox, R. & Lehrach, H. (1995) Efficient high resolution mapping of mouse IRS-PCR products, towards integrated genetic and physical mapping of the mouse genome, *Proc. Natl Acad. Sci. USA 90*, 5203−5306.
16. Elango, R., Riba, L., Housman, D. & Hunter, K. (1996) Generation and mapping of Mus spretus strain specific markers for rapid genomic scanning, *Mamm. Genome 7*, 340−343.
17. Johnson, P. & Williams, A. F. (1986) Striking similarities between antigen receptor J pieces and sequence in the second chain of the murine CD8 antigen, *Nature 323*, 74−76.
18. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680−685.
19. Rowe, L. B., Nadeau, J. H., Turner, R., Frankel, W. N., Letts, V. A., Eppig, J. T., Ko, M. S. H., Thurston, S. J. & Birkenmeier, E. H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource, *Mamm. Genome 5*, 253−274.
20. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries, *Proc. Natl Acad. Sci. USA 75*, 4853−4857.
21. Abott, C., Baker, E., Sutherland, G. R. & McCaughan, G. W. (1994) Genomic organization, exact localization, and tissue expression of the human CD26 (dipeptidyl peptidase IV) gene, *Immunogenetics 40*, 331−338.
22. Bernard, A.-M., Mattei, M.-G., Pierres, M. & Marguet, D. (1994) Structure of the mouse dipeptidyl peptidase IV (CD26) gene, *Biochemistry 33*, 15204−15214.
23. Mathew, S., Scanlan, M. J., Mohan Raj, B. K., Murty, V. V. V. S., Garin-Chesa, P., Old, L. J., Rettig, W. J. & Chaganti, R. S. K. (1995) The gene for fibroblast activation protein (FAP), a putative cell surface-bound serine protease expressed in cancer stroma and wound healing maps to chromosome band 2q23, *Genomics 25*, 335−337.
24. Mathew, S., Morrison, M. E., Murty, S., Houghton, A. N. & Chaganti, R. S. K. (1994) Assignment of the DPP IV gene encoding adenosine deaminase binding protein (CD26) to 2q23, *Genomics 22*, 211−212.
25. Yokotani, N., Doi, K., Wenthold, R. J. & Wada, K. (1993) Nonconservation of a catalytic residue in a dipeptidyl peptidase IV-related protein encoded by a gene on human chromosome 7, *Hum. Mol. Genet. 2*, 1037−1039.
26. Walter, R., Simmons, W. H. & Yoshimoto, T. (1980) Proline-specific endo- and exopeptidases, *Mol. Cell. Biochem. 30*, 11−127.
27. Craik, C. S., Choo, Q. L., Swift, G. H., Quinto, C., MacDonald, R. J. & Rutter, W. J. (1984) Structure of two related pancreatic trypsin genes, *J. Biol. Chem. 259*, 14255−14264.