Zebrafish (*Danio rerio*) matrilins: shared and divergent characteristics with their mammalian counterparts

Ya-Ping KO*, Birgit KOBBE*, Mats PAULSSON*† and Raimund WAGENER*¹

*Centre for Biochemistry, Medical Faculty, University of Cologne, Joseph-Stelzmann-Str. 52, D-50931 Cologne, Germany, and †Centre for Molecular Medicine, Medical Faculty, University of Cologne, Joseph-Stelzmann-Str. 52, D-50931 Cologne, Germany

We have cloned the cDNAs of the zebrafish (*Danio rerio*) members of the matrilin family of extracellular adaptor proteins. In contrast to mammals, no orthologue of matrilin-2 was found in zebrafish, either by RT (reverse-transcriptase) PCR using degenerated primers or by screening the databases (Ensembl and NCBI); however, two forms of matrilin-3, matrilin-3a and -3b, were present. The identity with the mammalian matrilins is from more than 70% for the VWA (von Willebrand factor A)-like domains to only 28% for the coiled-coil domains of matrilin-3a and -3b. In all zebrafish matrilins we found a greater variety of splice variants than in mammals, with splicing mainly affecting the number of EGF (epidermal growth factor)-like repeats. The exon–intron organization is nearly identical with that of mammals, and also the characteristic AT–AC intron interrupting the exons coding for the coiled-coil domain is conserved. In the matrilin-3b gene a

unique exon codes for a proline- and serine/threonine-rich domain, possibly having mucin-like properties. The matrilin-1 and -3a genes were mapped to chromosome 19 and 20 respectively by the radiation hybrid method. The temporal and spatial expression of zebrafish matrilins is similar to that seen in the mouse. Zebrafish matrilin-4 is highly expressed as early as 24 hpf (h post fertilization), whereas the other matrilins show peak expression at 72 hpf. By immunostaining of whole mounts and sections, we found that matrilin-1 and -3a show predominantly skeletal staining, whereas matrilin-4 is more widespread, with the protein also being present in loose connective tissues and epithelia.

Key words: *Danio rerio* (zebrafish), extracellular matrix, matrilin, von Willebrand factor A-like.

INTRODUCTION

Extracellular matrices are made up by proteoglycans, collagens and non-collagenous proteins that combine in various supramolecular assemblies, giving the tissue its architecture and mechanical properties, as well as influencing cellular differentiation through interactions with cell surface receptors. The proteins are in most cases modular, comprising tandem arrays of domains defined by their sequence and folding, and are sometimes joined into oligomers via coiled-coil α -helices and/or interchain disulphide bonds. The non-collagenous matrix proteins often occur as families of homologous members, and the matrilins form one of the more recently described among these groups (for review see [1]).

The four matrilins belong to a tetralogue protein family [2] and are made up from VWA (von Willebrand factor A)-like and EGF (epidermal growth factor)-like domains, with a C-terminal α -helical coiled-coil oligomerization domain. Subunits of matrilin-1 and -4 trimerize [3,4], whereas matrilin-2 and -3 are able to form tetrameric molecules [5,6]. In addition, matrilin-1 and -3, and perhaps other matrilins, sometimes form hetero-oligomers of variable stoichiometry [6–10].

The mammalian matrilins have been well characterized with regard to their tissue distribution and, in particular, their expression in the developing mouse embryo [4–6,11–14]. Matrilin-1 and -3 are expressed mainly in the cartilaginous tissues of the growing skeleton, whereas matrilin-2 and in particular matrilin-4 have a broader distribution [15].

Matrilin-1 was initially isolated as a cartilage matrix protein associated with aggrecan [16] by an interaction with distinct sites along the chondroitin sulphate-attachment region of the aggrecan core protein [17]. It was further identified as a component of collagen type II-containing filaments found in the pericellular matrix of cultured chondrocytes [18]. Recent results strongly point to a role of matrilins as adaptor proteins in the extracellular matrix. In native supramolecular assemblies containing collagen VI microfibrils and associated extracellular matrix proteins, matrilins were found in a complex with biglycan or decorin bound to the N-terminal domain of collagen VI [19]. Indeed the biglycan–matrilin-1 or decorin–matrilin-1 complex acts as a bridge between collagen VI microfibrils and aggrecan or collagen II molecules. Matrilins also bind to COMP (cartilage oligomeric matrix protein) with high affinity, an interaction which is co-operative and depends on the presence of divalent cations [20].

While point mutations in matrilin-3 have been shown to cause multiple epiphyseal dysplasia [21,22], analysis of matrilin-1, -2 and -3 knockout mice revealed no obvious phenotype [23–26], a fact that may be explained by a redundancy among the members of the matrilin family.

The zebrafish (*Danio rerio*) is a powerful model organism for the study of vertebrate development. The embryos develop rapidly, with all organs having been formed by 72 hpf (h post fertilization). The externally developing embryos are optically clear and are produced in large numbers, therefore large-scale mutagenesis programs can be monitored by simple microscopic

Abbreviations used: cM, centimorgan; dpf, days post fertilization; EBNA, Epstein–Barr nuclear antigen; EGF, epidermal growth factor; EST, expressed sequence tag; hpf, h post fertilization; LAPTM4, lysosome-associated protein transmembrane 4; MIDAS, metal ion-dependent adhesion site; PBT, PBS containing 0.1% Tween; PFA, paraformaldehyde; RBP-L, recombinant binding protein L; RT, reverse transcriptase; TBS, Tris-buffered saline; UTR, untranslated region; VWA, von Willebrand factor A.

¹ To whom correspondence should be addressed (email raimund.wagener@uni-koeln.de).

The nucleotide sequence data reported have been deposited in DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession numbers AJ629191, AJ697718, AJ629312, AJ697887, AJ697888, AJ698458, AJ698459, AJ698730, AJ717297 and AJ698731–AJ698736.

Table 1 Oligonucleotide primer sequences

RH map, radiation hybrid mapping.

Name	Sequence (5' → 3')	Direction	Usage
m1z1	CAGCAGAGTGAAGAACGAAG	Forward	RH map
m1z2	ATGTCAGGGGATTGCGTC	Reverse	RH map
m1z3	CAGTCTGTCTCTGGTTTTGG	Forward	RT-PCR
m1z4	GAACCTCCTGCTTACAGTG	Reverse	RT-PCR
m1z5	CAATGCTAGCAGGTCTGTGTAACACCAAGCCAC	Forward	Expression
m1z6	CAATGCCGGCCGCTTAACCGCACAATGTCTCCCGG	Reverse	Expression
m3az1	TGAAGGTGGAGATTGTCAGG	Forward	RH map
m2z1	AAYACIYAYGAYTAGCIAARGT	Forward	Degenerate primer
m2z2	CYTTICCCATRIATYTCATRTG	Reverse	Degenerate primer
m2z3	CCRGTCATIGWICCYTTICCCAT	Reverse	Degenerate primer
m2z4	GARGAYWSIGRIGGIMGICARGA	Forward	Degenerate primer
m3az2	TCATTTCTCCATTACATGCGTC	Reverse	RH map
m3az3	AGCAAGCAATCACCCGTATC	Forward	RT-PCR
m3az4	TTGCTCACACAACGTGCTGAC	Reverse	RT-PCR
m3az5	TTTACGCAGTCGGAGTGGACAG	Forward	Sequencing
m3az6	GAACAATCTCCACCTTACACC	Reverse	Sequencing
m3az7	CACACAAGCCTGAAGCAGAC	Forward	Sequencing
m3az8	CAATGCTAGCTACAGATTCACAGTGTAGG	Forward	Expression
m3az9	CAATGCCGGCCGCTTAACCGCACAATGTCTCCCGG	Reverse	Expression
m3bz1	ATCCACTCTCGACAGCAAC	Forward	Sequencing
m3bz2	AGGCACGACGAAGATCAAC	Reverse	RT-PCR
m3bz3	TGAGGCGGAATACCAGCAAG	Forward	Sequencing
m3bz4	GGTCAAACTCTCACCGTG	Forward	RT-PCR
m3bz5	CAATGCTAGCAGAGCCCTGCAAGAG	Forward	Expression
m3bz6	CAATGCCGGCCGCTTAACCGCAGAGCGTTTCCC	Reverse	Expression
m4z1	GCATGAGTGTGTGAGTGTCC	Forward	RT-PCR
m4z2	GTATGTCAGTGTGCTATTAGCC	Forward	Sequencing
m4z3	TGGGAATGAACACAGACACAC	Reverse	Sequencing
m4z4	TTGATCTCGTCTTTGCTGTG	Reverse	RT-PCR
m4az5	CAATGCTAGCGTGTAAATCTGGCCCGGTTG	Forward	Expression
m4az6	CAATGCCGGCCGCTTACCGCAGAGCTGTCTTGG	Reverse	Expression

observation of the embryos [27]. A genome-sequencing project will be completed soon, and human diseases that resemble mutations in zebrafish have been extensively analysed [28].

In the present work we study the structure, genetic organization and expression of zebrafish matrilins, with the longer-term goal of using the zebrafish to study matrilin function.

EXPERIMENTAL

RT (reverse-transcriptase)-PCR and EST (expressed sequence tag) clones

RT-PCR was used to clone zebrafish matrilin cDNAs. Primers were designed according to EST and genomic sequences deposited in the databases (Table 1). To reduce mutations in the PCR we used the Expand High Fidelity PCR System (Roche). The full-length matrilin-3a EST clone zeh1241, lacking the EGF-like domain, was obtained from Christopher Ton (Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada) and the full-length matrilin-4 EST clone fc11f02, containing 12 EGF-like domains, was from RZPD (Berlin, Germany).

Gene mapping

The zebrafish matrilin-1 and matrilin-3a genes were mapped with a zebrafish/hamster radiation hybrid panel (Research Genetics) through PCR amplification of fragments with primers m1z1 and m1z2, as well as m3az1 and m3az2 (Table 1), giving 173-bp and 285-bp products from zebrafish genomic DNA respectively. Data were submitted to the Tübingen zebrafish radiation hybrid

database (<http://wwwmap.tuebingen.mpg.de/>), which mapped the matrilin-1 gene at 53.5 to 59.4 cM (centimorgan) and matrilin-3a at 78.8 to 86.4 cM from the top of chromosome 19 and 20 respectively.

Expression and purification of recombinant matrilin-1, -3a, -3b and -4 VWA1 domains

cDNAs coding for zebrafish matrilin-1, -3a, -3b and -4 VWA1 domain were generated by PCR on the full-length cDNA. Suitable primers (Table 1) introduced 5'-terminal NheI and 3'-terminal NotI restriction sites. The cDNA was inserted into the expression vector pCEP-Pu downstream of the sequences encoding the BM-40 signal peptide [29] and an N-terminal His₆-tag [30].

The recombinant plasmids were introduced into HEK-293/EBNA (Epstein-Barr nuclear antigen) cells (Invitrogen) by transfection with FuGENE™ 6 (Roche). Following selection with puromycin (1 mg/ml) the cells were transferred to serum-free medium for harvesting of the recombinant protein. After filtration and centrifugation (1 h, 10000 g), cell culture supernatants containing the N-terminally His₆-tagged matrilin-1, -3a and -4 VWA1 domains were applied to TALON metal affinity columns (1 ml; BD Biosciences). The bound protein was eluted with 0.25 M imidazole in 0.1 M NaCl, 20 mM Tris, 50 mM NaH₂PO₄, pH 8.0, containing 0.2 % sodium azide.

Preparation of antibodies against matrilin-1, -3a and -4

The purified matrilin-1, -3a and -4 VWA1 domains were used to immunize rabbits. The antisera obtained were purified by affinity chromatography on columns with the original antigens coupled to CNBr-activated Sepharose (Amersham Biosciences). The specific antibodies were eluted with 0.1 M glycine, pH 2.5, and the eluate was immediately neutralized through addition of 1 M Tris/HCl, pH 8.8.

Determination of cross-reactivity of the antibody against matrilin-3a with matrilin-3b

Cell culture supernatant of HEK-293/EBNA cells transfected with cDNA coding for zebrafish matrilin-3b VWA1 domain was subjected to SDS/PAGE together with a dilution series of purified zebrafish matrilin-3a VWA1 domain. After transfer on to nitrocellulose the immunoblot was incubated with the affinity-purified zebrafish matrilin-3a specific antibody diluted in TBS containing 5 % low fat milk powder. The bound antibodies were detected by luminescence using peroxidase-conjugated swine anti-rabbit IgG (Dako), 3-aminophthalhydrazide (1.25 mM), *p*-coumaric acid (225 µM) and 0.01 % H₂O₂.

Whole mount immunostaining

Zebrafish larvae [5 dpf (days post fertilization)] were fixed overnight at 4 °C in 4 % PFA (paraformaldehyde) in PBS, pH 7.4, washed in PBT (PBS containing 0.1 % Tween) and finally washed and stored in methanol at -20 °C. To bleach pigment and block endogenous peroxidases, larvae were incubated overnight in 3 ml of 10 % H₂O₂ in methanol, then 10 ml of PBT was added and the incubation continued for a further 16 to 24 h. Larvae were washed in PBT, digested with 2 µg/ml proteinase K for 8 min and fixed again in 4 % PFA for 15 min. After washing, larvae were treated with hyaluronidase (Sigma; 500 units/ml in 0.1 M NaH₂PO₄, 0.1 M sodium acetate, pH 5.0) at 37 °C for 2 h and blocked in 3 % normal goat serum for 2 h. Affinity purified antibodies were applied at appropriate dilutions (matrilin-1 and -3a, 1:1000; matrilin-4, 1:500) and the specimens incubated for 2 h.

The primary antibodies were visualized by consecutive treatment of larvae for 2 h each with biotinylated secondary antibody and a streptavidin–peroxidase conjugate (ABC kit, Vectastain). All antibodies were diluted in 3 % (w/v) normal goat serum in PBT. For colour development, larvae were pre-soaked in diaminobenzidine (0.2 mg/ml PBT) for 30 min and 1 μ l of 0.3 % H₂O₂ solution was added while the larvae were observed under a dissection microscope. For detailed analysis, larvae were post-fixed in 4 % PFA for 15 min, washed in PBT and gradually transferred into 90 % glycerol. Except when indicated, all procedures were carried out at room temperature (20 °C). Occasionally, specimens were kept overnight at 4 °C between steps in the staining procedure.

Immunostaining on sections

Immunostaining was performed on paraffin-embedded sections of 5 dpf and 4-month-old zebrafish. Sections were deparaffinized by two 5-min incubations in xylol. After rehydration in PBS, the sections were digested with hyaluronidase (500 units/ml in 0.1 M NaH₂PO₄, 0.1 M sodium acetate, pH 5.0) at 37 °C for 30 min. After washing, the sections were blocked with 5 % (w/v) BSA in TBS for 1 h and incubated with the affinity-purified antibody overnight at 4 °C. The primary antibodies were visualized by consecutive treatment with biotin–streptavidin–peroxidase-conjugated goat anti-rabbit IgG (Dianova) and alkaline phosphatase-conjugated streptavidin (Dianova) for 1 h each. Antibodies and enzyme-conjugates were diluted in 1 % (w/v) BSA in TBS and the slides developed with Sigma FAST™ Fast Red TR/naphthol AS-MX (Sigma). Immunofluorescence microscopy was performed as described previously [6] using the affinity-purified rabbit antibodies against the zebrafish matrilin VWA1 domains and a Cy3-conjugated affinity-pure anti-rabbit IgG as secondary antibody.

Sequence analysis

Sequences of zebrafish genes and ESTs were obtained from NCBI or from Ensembl. Sequence alignments were obtained using the Pileup program of the Wisconsin Package Version 10.3 (Accelrys Inc., San Diego, CA, U.S.A.). The Phylip software package (version 3.5) was used to construct the phylogenetic tree by the protein parsimony method. Bootstrap resampling assessed the strength of the nodes and the probability of occurrence of a given branching for 100 possible trees was calculated. A consensus tree was derived by the Consense program of the package.

RESULTS

Screening and isolation of zebrafish matrilin clones

In a screen of the zebrafish databases (NCBI and Ensembl) with sequences of mammalian matrilin VWA domains as query, we identified single orthologue genes for matrilin-1 and -4 and two orthologue genes for matrilin-3. The corresponding zebrafish cDNAs were either obtained as EST clones or cloned by RT-PCR.

Matrilin-1

We identified clone BC045465 as the zebrafish orthologue of matrilin-1. It has an open reading frame of 1467 bp, encoding a protein comprising 489 amino acid residues with a signal peptide of 17 amino acid residues, as predicted by a method using neural networks [31]. The mature secreted protein has a calculated M_r of 51 555. It comprises two VWA domains that are connected by a single EGF-like domain followed by an α -helical coiled-coil oligomerization domain, and therefore completely resembles the

Figure 1 Domain structure of zebrafish and mouse matrilins (A) and amino acid sequence of the proline- and serine/threonine-rich domain in zebrafish matrilin-3b (B)

(A) For explanation of symbols, see box below panel. (B) Predicted mucin-type N-acetyl-galactosamine O-glycosylation sites are shaded black. matn, matrilin.

mammalian matrilin-1 (Figure 1A). Furthermore, by RT-PCR using primers mlz3 and mlz4 (Table 1) we also detected an alternatively spliced mRNA that lacks the EGF-like domain (Figure 1A).

Matrilin-3a

Two EST clones, AI353980 and AA566274, were identified as zebrafish matrilin-3a clones. Sequencing of AI353980 revealed a 1991 bp cDNA, encoding a protein comprising 295 amino acid residues. After cleavage of a predicted signal peptide of 22 amino acid residues, the mature secreted protein has a calculated M_r of 32 811. Interestingly, it lacks the two EGF-like domains that in mammals connect the VWA domain and the C-terminal coiled-coil domain (Figure 1A). By RT-PCR we amplified alternatively spliced cDNAs that, in addition, contained sequences coding for three or four EGF-like domains (Figure 1A). The longest cDNA encodes a protein of 480 amino acid residues with a calculated M_r of 53 006. The stretch of amino acid residues N-terminal to the VWA domain is conserved between mammalian matrilin-3 and

zebrafish matrilin-3a, but with fewer positively charged amino acid residues in zebrafish.

Matrilin-3b

In a later screen of the genomic database, a second matrilin-3 gene was identified. In contrast to matrilin-3a, only a single matrilin-3b EST clone coding for a partial sequence (CK693698) could be identified. Therefore, we cloned the matrilin-3b cDNAs by RT-PCR using primers based on the genomic sequence. This yielded four alternatively spliced matrilin-3b cDNAs. The longest splice variant of matrilin-3b has an open reading frame of 1437 bp that codes for a protein comprising 478 amino acid residues. After cleavage of a predicted signal peptide of 22 amino acid residues, the mature secreted protein has a calculated M_r of 50 136. The VWA domains of matrilin-3a and -3b are 80% identical at the amino acid level. Uniquely in a matrilin, a proline- and threonine/serine-rich sequence (Figure 1B) precedes the N-terminal VWA domain in matrilin-3b, which itself is followed by a single EGF-like domain and the C-terminal coiled-coil domain. The long unique N-terminal stretch of amino acid residues also contains a cluster of positively charged amino acid residues (Figure 1A) similar to that in matrilin-3a.

The matrilin-3b variant that lacks the proline- and threonine/serine-rich sequence and the EGF-like domain has the same domain structure as the shortest form of matrilin-3a, containing only the N-terminal positively charged stretch, a single VWA-domain and the coiled-coil domain. In addition, two isoforms exist that lack either the proline- and threonine/serine-rich sequence or the EGF-like domain (Figure 1A). The NetOGlyc server (<http://www.cbs.dtu.dk/services/NetOGlyc/>) predicted that the proline- and threonine/serine-rich sequence contained 33 potential mucin-type N-acetylgalactosamine O-glycosylation sites.

Matrilin-4

At least 15 different EST clones representing zebrafish matrilin-4 were identified and among these EST AI626844 was sequenced. The 3718 bp contain an open reading frame of 2832 bp encoding a protein comprising 944 amino acid residues. After cleavage of a predicted signal peptide of 20 amino acid residues the mature secreted protein has a calculated M_r of 102 576. As in the mammalian matrilin-4, the two VWA domains are connected by EGF-like domains, but here the number of these is between four and twelve. The C-terminal coiled-coil domain follows the second VWA domain (Figure 1A). Using RT-PCR, we searched for alternatively spliced mRNAs and detected a variety, all differing in the number of EGF-like domains, ranging from four to seven (Figure 1A). Another clone (AY570730) harbouring nine EGF-like domains (Figure 1A) was recently published in the databases (Ensembl and NCBI).

Sequence analysis

Phylogenetic trees based on the alignment of VWA domains were constructed and by maximum parsimony (Figure 2), as well as by the neighbour-joining and maximum-likelihood methods (results not shown), the VWA domains of zebrafish are located in the same branches as the respective mouse domains, clearly showing that they are orthologues. We could not find a matrilin-2-specific EST clone and in all cases RT-PCR, using degenerate primers (Table 1) based on conserved sequences from human, mouse and *Fugu* matrilin-2, did not yield a product (results not shown). In addition, there is no orthologue of matrilin-2 present in the most recent version of the zebrafish genomic database or any matrilin-2

Figure 2 Phylogenetic analysis of matrilin VWA domains

VWA domain amino acid sequences of zebrafish (z), *Fugu* (f) and mouse (m) were aligned with the PILEUP program of the GCG package, using the default parameters. The VWA4 domain of human collagen XII $\alpha 1$ (colXII $\alpha 1$) was taken as an outgroup. The aligned sequences were used for the construction of a tree by the PROTPARS program of the PHYLIP package, version 3.5. Bootstrap support values were obtained with 100 replicates and are given at the respective nodes when the values are below 70%. The VWA domains of *Fugu* were derived from the draft sequence of *Fugu* genome, available as *Fugu* assembly release 3. The genomic sequence of the VWA1 domain of *Fugu* matrilin-4 is not yet available. matn, matrilin.

specific genomic trace sequence. Moreover, several genes that have orthologues on zebrafish chromosome 19 flank the matrilin-2 genes in mammalian genomes. In man and mouse the LAPTM4 β (lysosome-associated protein transmembrane 4 β) and RPL30 (ribosomal protein L30) genes flank the matrilin-2 gene, whereas they are direct neighbours in the zebrafish genome (contig BX284622). In apparent contrast with zebrafish, *Fugu* contains a matrilin-2 gene.

The identity of zebrafish VWA domains with their mouse counterparts is 71–72% and the lengths of the VWA domains are strongly conserved. The matrilin-3 A1 domains perfectly fit to the MIDAS (metal ion-dependent adhesion site) motif consensus sequence (DXSXSX_nTX_nD) (Figure 3A), which is in contrast with human and mouse matrilin-3 where the threonine in the MIDAS motif has been exchanged for a serine residue. Minor differences occur in both VWA domains of zebrafish matrilin-1, which lacks a 4-amino-acid residue extension at the N-terminus of VWA2 and one amino acid residue in the loop at the beginning of the fourth central β -strand of VWA1, and in the VWA1 domain of zebrafish matrilin-4, which lacks three amino acid residues at the same loop position.

Phylogenetic analysis did not allow construction of a tree of the zebrafish EGF-like domains with reasonable bootstrap values. Nevertheless, the zebrafish matrilin-4 EGF-like domains 7 and 8 are identical on the protein level and the domains 3, 4, 5, 6, 9, 10 and 11 are nearly identical (Figure 3B), and are probably due to recent duplication events. The identity of the orthologue EGF-like domains is lower than for the VWA domains (Figure 3B) with highest values of 66.7% for the EGF-like domain of zebrafish and mouse matrilin-1, 65% for the EGF-like domain 11 of zebrafish matrilin-4 and EGF-like domain 3 of mouse matrilin-4 and

Figure 3 Amino acid sequence alignments of VWA domains (A), EGF-like domains (B) and coiled-coil domains (C) of zebrafish and mouse matrilins

The sequences were aligned with the PILEUP program of the GCG package, using the default parameters. In (A), the conserved MIDAS are denoted with arrowheads. In (B), the conserved positions of the cysteine residues are marked by arrows. In (C), the positions a and d of the heptad repeats are indicated.

55.8% for EGF-like domain 4 of zebrafish matrilin-3a and the EGF-like domains 1 and 4 of mouse matrilin-3. In contrast with the mammalian matrilin-3 EGF-like domains, in zebrafish the spacer between EGF-like domain 2 and 3, as well as 3 and 4, is extended to 34 and 13 amino acid residues respectively.

Figure 4 Organization of zebrafish matrilin genes

Closed boxes, exons; open box, exon coding for the proline- and serine/threonine-rich domain. SP, A, E and C, exons coding for signal peptide, VWA domain, EGF-like domain and coiled-coil domain respectively. E0a and E0b, exon 0a and 0b of the matrilin-4 gene. In the zebrafish matrilin-1 gene the corresponding supercontigs are given above the map. matn, matrilin.

All zebrafish matrilins contain a coiled-coil α -helix at the C-terminus, as predicted by the COILS program [32]. As for mouse matrilin-3, the agreement with the consensus is the lowest for zebrafish matrilin-3b, whereas matrilin-3a has a higher match. The coiled-coil domains of zebrafish and mouse matrilin-1 show an identity of 67%, whereas it is 48% for matrilin-4 and only 28% for each of matrilin-3a and -3b (Figure 3C).

Zebrafish matrilin genes

The structures of the matrilin genes are well conserved in the zebrafish genome. The zebrafish matrilin-1 gene (Figure 4) consists of 8 exons and the position and phase of the introns is identical with those of man [33], mouse [34] and chicken [35]. There is no contig that contains the complete gene, but all exons are either contained in the contig *Zv4_scaffold1872* (Zebrafish Genome Project) or on trace clones (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>), allowing the determination of the exon-intron borders (Table 2a). By the radiation hybrid method the matrilin-1 gene was mapped to chromosome 19 (53.5–59.4 cM) (Figure 5A). This is in contrast with the data from the Zebrafish Genome Project, which assigned the contig *Zv4_scaffold1872* to chromosome 23. A search for mammalian orthologue genes flanking the zebrafish matrilin-1 gene on chromosome 23 was negative, whereas several genes (*OPRD1*, *SFRS4*, *FABP3*, *PTP4A2*) that map around the matrilin-1 locus on human chromosome 1p35 have orthologous genes that map to zebrafish chromosome 19. In addition, the BAC clone end *bZ87G18.za* that is located upstream of the matrilin-1 gene on the Tübingen map of the zebrafish genome (Figure 5A) matches the contig *Zv4_scaffold1650*, which is assigned to zebrafish chromosome 19 in the Zebrafish Genome Project. Furthermore, this contig contains the zebrafish orthologue for the

Table 2 Splice junctions in the zebrafish matrilin genes

Zebrafish genes: (a) matrilin-1; (b) matrilin-3a (BX649436); (c) matrilin-3b (AL935163); (d) matrilin-4 (AL954861). Exon sequences are in uppercase letters; intron sequences are in lowercase letters. The AT-AC splice sites are in bold. Underlined, G/A exchange in exon 5 of the matrilin-1 gene. *The length of the 5'- and 3'-UTRs are not included. The single letter code is used for amino acids, with the position in parentheses. SigPep, signal peptide; CC1 and CC2, coiled-coil 1 and 2; P/ST, proline- and serine/threonine-rich.

Exon	Domain	Size (bp)	Splice donor	Intron	Size (bp)	Splice acceptor	Codon phase	Amino acid interrupted	Trace identification /GenBank® accession number	
									Donor	Acceptor
1	SigPep	82*	TATGGgtaaga	1	> 1040	ccgtttccagCTG	I	A (28)	483 786 525	105 042 069
2	VWA1A	347	GCAAGgtcaga	2	> 1160	atatgtcagGTG	0	K/V (143/144)	70 110 729	42 718 557
3	VWA1B	223	TTGTGgtggta	3	> 2750	ttcatgttagCGG	I	A (218)	42 718 557	Zv4_scaffold1872
4	EGF	129	CTCAGgtgagt	4	~ 2960	catctctcagCTT	I	A (261)	Zv4_scaffold1872	Zv4_scaffold1872
5	VWA2A	405	GCTAGgttaga	5	~ 1320	tatgttttagGCT	I	G (396)	Zv4_scaffold1872	25 609 863
6	VWA2B	153	TCAAGgtaaat	6	> 825	ttttttgcagAGG	I	E (447)	25 609 863	25 696 578
7	CC1	81	GAAATatctct	7	> 520	taagcatcactGG	I	L (474)	25 696 578	451 980 020
8	CC2	47*								

(b)										
Exon	Domain	Size (bp)	Splice donor	Intron	Size (bp)	Splice acceptor	Codon phase	Amino acid interrupted		
1	SigPep	157*	TGCAGgtaagt	1	2990	caactctcagCTA	I	A (53)		
2	VWA	576	GTGCCgtagagg	2	1790	ctgttccagGTA	I	G (245)		
3	EGF1	126	AAAACgtaaga	3	2564	cttattccagATG	I	H (287)		
4	EGF2	216	TGCCAgtagagt	4	2664	cttatttcagGTG	I	S (359)		
5	EGF3	153	TGCCAgtagagt	5	199	cttattgcagGTC	I	S (410)		
6	EGF4	126	AAAAAgtaaga	6	2457	tgatgtgaagAGG	I	K (452)		
7	CC1	99	CAAGCatacc	7	120	atcgttatacTTG	I	L (485)		
8	CC2	53*	GAATGgtgcgt	8	81	tgctcagcagggt	—	—		
9	3'-UTR	—*								

(c)										
Exon	Domain	Size (bp)	Splice donor	Intron	Size (bp)	Splice acceptor	Codon phase	Amino acid interrupted		
1	SigPep	142*	TGGAGgtaatc	1	1146	ccaacacacagCTC	I	A (48)		
2	P/ST	405	CCCAGgtagac	2	3551	tctcccacagCTC	I	A (193)		
3	VWA	576	CTGTGgtaaga	3	1570	aattttgcagGTT	I	G (385)		
4	EGF1	132	CAACGgtttgt	4	301	tctcttacagAAG	I	E (429)		
5	CC1	102	CAGACataacc	5	196	ctgttcctactGG	I	L (463)		
6	CC2	47*								

(d)										
Exon	Domain	Size (bp)	Splice donor	Intron	Size (bp)	Splice acceptor	Codon phase	Amino acid interrupted		
0	5'-UTR	—	CTGAGgtgagt	0	14215	ccacctacagCAA	—	—		
1	SigPep	74*	ATCAGgttggt	1	2610	tttcccacagGTG	I	G (25)		
2	VWA1	564	CTGCGgttaaaa	2	3312	atttttcagGGA	I	G (213)		
3	EGF1	123	CACACgtgcgt	3	83	cttcctgcagCCA	I	P (254)		
4	EGF2	123	CAAAAgttagt	4	3905	ctgtttccagTGA	I	M (295)		
5	EGF3	123	CACAAGtcagt	5	241	cccggtccagTGA	I	M (336)		
6	EGF4	123	CACAAGtcagt	6	2412	cctgttccagTGA	I	M (377)		
7	EGF5	123	TACAAGtcagt	7	940	ccatatctagTGA	I	M (418)		
8	EGF6	123	TACAAGtcagt	8	3184	ccatatctagTGA	I	M (459)		
9	EGF7	123	TACAAGtcagt	9	2385	tctactgttagTGA	I	M (500)		
10	EGF8	123	TACAAGtcagt	10	257	cccggtccagTGA	I	M (541)		
11	EGF9	123	CACAAGtcagt	11	1615	ctgtttccagTGA	I	M (582)		
12	EGF10	123	TACAAGtcagt	12	744	tgccctcttagTGA	I	M (623)		
13	EGF11	123	CCAGTgtatgt	13	1662	tgtcatgcagCGG	I	S (704)		
14	EGF12	123	TGGAAGtaaatg	14	659	tttctccaagCTT	I	T (745)		
15	VWA2A	414	AGCAGgtatgt	15	8503	tgcglttcagGAA	I	G (883)		
16	VWA2B	153	CGCAGgtgtgt	16	725	tggtgttcagCGG	I	A (934)		
17	CC1	105	gaaacataacc	17	4948	cacattttacTGT	I	L (959)		
18	CC2	47*								

human AK2 (adenylate kinase 2), which is mapped to human chromosome 1p34. Together, this indicates that our mapping of the matrilin-1 gene to chromosome 19 is indeed correct.

The two genes coding for the zebrafish matrilin-3 orthologues are located on different chromosomes (Figure 5B). We mapped the

matrilin-3a gene to chromosome 20 (78.8–86.4 cM) (Figure 5A), which is consistent with the mapping by the Zebrafish Genome Project that can be obtained by a BLAST search on the zebrafish genome server (<http://www.ensembl.org/Multi/blastview>). Two contigs containing the matrilin-3a gene are present in the database,

Figure 5 Zebrafish matrilin gene maps

(A) Linkage map for zebrafish matrilin-1 (*matn1*) and *matn3a*. Arrows indicate the positions of *matn1* and *matn3a*. The Tübingen radiation hybrid map of the zebrafish genome is given on the right with distances in centirays (cR). For an explanation of markers on the map, see [43] and <http://www.map.tuebingen.mpg.de/>. The genetic map, on the left-hand side, shows SSLP (simple sequence length polymorphism) markers, available on the MGH Zebrafish WWW Server (<http://zebrafish.mgh.harvard.edu/>) and distances in cM. (B) Comparative maps of zebrafish chromosomes 13 and 20 and human chromosome 2. The boxes highlight the positions of the matrilin genes. The duplicated zebrafish *matn3a* and *matn3b* are located on two different chromosomes (linkage groups), presenting partial synteny with human chromosome 2, as indicated by the asterisks. The mapping position of the human syntenic genes is indicated in bold. Putative orthologies were determined with information available from the Ensembl zebrafish server (http://www.ensembl.org/Danio_rerio/bblastview) and by searching the human genome using the zebrafish protein sequences as TBLASTN queries (<http://www.ncbi.nlm.nih.gov/BLAST>). Vertical bars indicate the positions of the two contigs containing the zebrafish matrilin 3a gene.

BX942827 and BX649436. The matrilin-3a gene comprises 9 exons (Figure 4 and Table 2b) that have identical positions and intron phases as in mouse [36], except for an additional intron present in the 3'-UTR (3' untranslated region) 14 bp after the stop codon.

The matrilin-3b gene is located on chromosome 13, as determined by a BLAST search on the zebrafish genome server. The matrilin-3b gene present on contig AL935163 is the simplest matrilin gene containing only six exons, i.e. single exons coding for the signal peptide, the VWA and the EGF-like domain and two exons for the coiled-coil domain. An additional exon, unique to the matrilin gene family, is present between the exons coding for the signal peptide and the VWA domain. The positions of the remaining exons and the intron phases are also conserved (Figure 4 and Table 2c). Some genes in the close vicinity of both zebrafish matrilin-3 genes are in synteny to their human and mouse orthologues. The genes *wdr35*, *kcns3* and *nt5c1b*, which map to chromosomes 2 and 12 of man and mouse respectively are located upstream of the zebrafish matrilin-3b gene on chromosome 13. The genes *laptm4a* and *adcy3* are mapped downstream of the zebrafish matrilin-3a gene on zebrafish chromosome 20. Unfortunately, the assembly of the contigs around the matrilin-3a gene is most likely erroneous. The contigs BX942827 and BX649436 each contain the matrilin-3a and the *laptm4a* gene, but are assigned in tandem and therefore the directly neighbouring genes are not clearly defined (Figure 5B). Interestingly, the matrilin-3 gene is the only duplicated gene in this region of the zebrafish genome.

We failed to map the matrilin-4 gene by the radiation hybrid method and also a BLAST search of the zebrafish genome failed to locate this gene, although a contig containing the gene is in the database (AL954861). The gene structure, the position of exons and introns is similar to the mammalian orthologues [37,38] (Figure 4 and Table 2d). Except for the exons coding for the 12 EGF-like domains that are transcribed in the longest splice variant, no additional exon could be detected in the genomic sequence.

As in the human and mouse genome, the zebrafish matrilin-4 gene shows a 5' antiparallel overlap with the gene encoding the transcription factor RBP-L (recombining binding protein L) (Figure 6; EST clone AL918240). In addition we identified an EST clone (BM342533) coding a matrilin-4 cDNA that has an extended 5'-UTR containing also an untranslated exon 0. Thereby we were able to identify the exact positions of the exons, as well as the antisense overlapping sequences of both mRNAs. In zebrafish the overlap is only 5 nt instead of the 71 nt found in mouse, but the positioning of the exons and introns is similar, i.e. exon 1 and parts of exon 2 of the zebrafish RBP-L gene are located in the first intron of the matrilin-4 gene and the first exon of matrilin-4 is located in the second intron and the second exon of the RBP-L gene. In zebrafish an exon 0a, present in the matrilin-4 gene in mouse [38], could not be detected in the databases and only one EST corresponding to the longer variant of RBP-L present in mouse was found. An EST coding for the shorter variant of RBP-L was not present.

As in all known matrilin genes, the last intron, separating the exons coding for the coiled-coil domain, does not follow the GT-AG rule and belongs to the subgroup of U12-type introns having AT-AC at the ends. Their exon intron borders and their branch sites are highly conserved also in the zebrafish matrilin genes (Table 3).

Matrilin expression during development

We studied the differential expression of the four zebrafish matrilin genes at 24, 48, and 72 hpf, as well as in adult fish by RT-PCR

Figure 6 Organization of overlapping parts of mouse and zebrafish matrilin-4 and RBP-L genes

(A) Mouse genome and (B, C) zebrafish genome. (A, B) Filled boxes, exons; (A–C) arrows, general direction of transcription. (C) Genomic sequence of overlapping exons 0 (matrilin-4) and 2 (RBP-L). For RBP-L the exon is given in uppercase and introns in lowercase letters. The amino acid sequence of RBP-L is given below and the matrilin-4 exon is underlined. The overlap between the exons is shaded grey. Matn, matrilin.

Table 3 Zebrafish AT-AC introns

Exon sequences are in uppercase letters; intron sequences are in lowercase letters. The splicing branch point and the AT-AC splice sites are in bold.

Gene	Splice donor	Branch site	Distance (bp)	Splice acceptor
Matrilin-1	AAT at atcctt	ttcct g acct	13	cac TGG
Matrilin-3a	AGC at atcctt	aacct ta cgat	13	tac TTG
Matrilin-3b	GAC at atcctt	ttct ta actgt	11	tac TGG
Matrilin-4	AAC at atcctt	gtcct ta accac	12	tac TGT
Consensus	at atcctt	ttcct tr acyc		yac

(Figure 7). At 24 hpf, matrilin-4 is already clearly expressed, whereas PCR products corresponding to matrilin-3a and -3b were weak and matrilin-1 could be detected only after over-exposure. At 48 hpf matrilin-4 is strongly expressed and matrilin-3a and -3b clearly present, again matrilin-1 could hardly be detected. All matrilins show the highest expression at 72 hpf. In adult fish, mRNAs for matrilin-1, -3a and -4 are clearly present, whereas for matrilin-3b only the shortest splice variant containing the VWA domain and the coiled-coil domain could be detected as a weak band. The splice variants carrying the proline- and threonine/serine-rich stretch of amino acid residues were not found in adult fish.

Figure 7 RT-PCR analysis of matrilin mRNA species expressed during zebrafish development

RT-PCR analysis was performed at 24, 48 and 72 hpf, as well as in adult fish using primer pair m1z3 and m1z4 for matrilin-1, m3az3 and m3az4 for matrilin-3a, m3bz2 and m3bz4 for matrilin-3b, and m4z1 and m4z4 for matrilin-4 (Table 1). As a marker the 1-kb ladder from Gibco-BRL was used. Bands marked with asterisks have been shown by sequencing to be artefactual. In the control sample water was included instead of an cDNA solution. matn, matrilin.

Generation of zebrafish-matrilin-specific antisera

cDNAs encoding the sequences of zebrafish matrilin VWA1 domains of matrilin-1, -3a, -3b and -4 were cloned into the pCEP-Pu vector utilizing the BM-40 secretion signal sequence and an N-terminal His₆-tag [30]. The recombinant plasmids were introduced into HEK-293/EBNA cells and maintained in an episomal form. The recombinant proteins secreted into the cell culture medium were, except for matrilin-3b, subsequently purified by affinity chromatography on a cobalt column. The purified proteins appeared in non-reducing SDS/PAGE mainly as monomeric molecules, but small amounts of higher oligomers could also be detected (results not shown). After reduction single bands with apparent molecular masses in the range expected for monomeric VWA domains were seen (results not shown). The purified proteins were used to immunize rabbits. The antisera obtained were purified by affinity chromatography on columns carrying the

Figure 8 Cross-reactivity of the antibody to matrilin-3a with matrilin-3b

(A) Immunoblot analysis of cell culture supernatant of non-transfected (nt) and matrilin-3b VWA-expressing (3b) HEK-293/EBNA cells and purified matrilin-3a VWA domain at different concentrations (3a) using the antibody specific for matrilin-3a. (B) Ponceau staining of (A) to show protein loading.

original antigens and were shown by ELISA to be highly specific for the matrilin form used for immunization. As the discovery of matrilin-3b is only recent, an antiserum to this protein has not yet been raised. In immunoblot, the antibody to matrilin-3a showed only a marginal cross-reactivity with the matrilin-3b VWA1 domain (Figure 8), despite the high sequence identity.

Matrilins are differentially expressed

Whole mount immunostaining was performed on 5-day-old fish using the affinity-purified matrilin antibodies (Figures 9F–9L). All matrilins are present in the developing skeleton. The matrilin-3a antibody strongly stained Meckel's cartilage, the palatoquadrate, the ceratohyal, the ethmoid plate, the anterior basicranial commissure, the parachordal, the hyosymplectic and the auditory capsule, as well as the basis of the pectoral fin (Figures 9G and 9K). In addition, matrilin-1 was found in the posterior part of the notochord (Figure 9I). In contrast with matrilin-1 and -3a, matrilin-4 showed similar staining intensity in Meckel's cartilage, the ceratohyal and the five ceratobranchials (Figures 9H and 9L). Further, the matrilin-4 antibody stained the eye (Figures 9H and 9L), the skin (Figures 9H and 9L) and the myosepta (Figure 9L). In all fins matrilin-4 staining could be detected in the fin rays (Figure 9L).

In addition, the tissue distribution of zebrafish matrilins was investigated on paraffin sections of 5 dpf and 4-month-old zebrafish. Sectioning of 5 dpf fish (Figures 10A–10C) clearly confirmed the restricted skeletal staining of matrilin-3a seen in whole mount stainings (Figure 10K). In contrast, matrilin-1 could be detected in the notochord and in intestine, albeit after long exposure (Figure 10A). Matrilin-4 is more widespread and could be detected in the eye and in the myoseptum (Figure 10C), as well as in skeletal tissues. The overall expression pattern was not altered in sections of 4-month-old fish, but as the fish were larger a more detailed analysis could be performed (Figure 10). In consecutive sections through the trabecular bone of the skull matrilin-1 (Figure 10G) and matrilin-3a (Figure 10H) showed a similar expression in proliferating, and more strongly in hypertrophic, cartilage, but only matrilin-3a was found in perichondrium (Figure 10H). In contrast, matrilin-4 revealed a strong zonal expression in the proliferating cartilage (Figure 10I). In vertebrae the staining for matrilin-3a was broad and strong (Figure 10L), whereas matrilin-1 and -4 are expressed only around proliferating chondrocytes (Figures 10K and 10M). Uniquely, matrilin-1 shows a staining in the notochord and in the surrounding secondary chordal sheath (Figures 10A, 10B and 10K), whereas matrilin-3a is not present in these tissues (Figure 10D). Matrilin-1 can still be detected in the intestine (Figure 10C) and, as in 5 dpf fish, matrilin-4 was found

Figure 10 Matrilin tissue distribution in 4-month-old zebrafish

(A–F) Immunofluorescence microscopy was carried out on paraffin-embedded tissue sections which were incubated with affinity-purified antibodies against matrilin-1 (A–C), -3a (D) or -4 (E, F), followed by Cy3-conjugated goat anti-rabbit IgG. In the vertebrae (vb), matrilin-1 was detected in the cartilage, in the secondary chordal sheath (sc) (A), the notochord (no) network (B) and in the intestinal epithelium (ep) (C), whereas matrilin-3a was found in the cartilage of the vertebrae (vb) (D). Strong signals for matrilin-4 were found in the adenohypophysis (ad) (E) and in myosepta (ms) (F) throughout the fish. (G–M) Immunohistochemistry was performed on paraffin-embedded tissue sections by staining with affinity-purified antibodies against matrilin-1 (A, D), -3a (B, E) or -4 (C, F) followed by biotin–streptavidin–peroxidase-conjugated goat anti-mouse IgG and alkaline phosphatase-conjugated streptavidin. In the trabecular bone (G–I) and vertebrae (K–M), matrilin-1 is deposited in proliferating (pc) and hypertrophic cartilage (hc) and in the secondary chordal sheath (sc) (G, K); matrilin-3a is expressed throughout the cartilage including proliferating (pc) and hypertrophic cartilage (hc) (H, L) and perichondrium (pe) (H); whereas matrilin-4 is weakly expressed only in the proliferating cartilage (pc) (I, M). (te), tendon. Scale bars represent 100 μ m, except for those in (B, E) which represent 50 μ m. matn, matrilin.

obtained using degenerate matrilin-2-specific primers, indicates that the matrilin-2 gene is truly missing in zebrafish. In addition, it is not found in the finished or draft sequences of the Zebrafish Genome Project, covering 60.9 % of the genome, or in any trace sequence and the databases do not contain any matrilin-2 EST clone. It has been proposed that the four mammalian matrilins evolved through two subsequent duplications of an ancestral matrilin gene [1]. In that case the lack of a matrilin-2 gene in zebrafish would be due to a later elimination. The presence of matrilin-2 genes in *Fugu* (AAH63920; MGC64509 protein) supports this hypothesis. We can also not exclude the existence of duplicated versions of the matrilin-1 and matrilin-4 genes, for the same reasons as given above for matrilin-2. These issues will be completely resolved first upon the completion of the zebrafish sequencing project. A comparison of the domain structures of mammalian and zebrafish matrilins shows a

striking similarity between the longer splice variants of zebrafish matrilin-4, which harbour nine and twelve EGF-like domains, and mammalian matrilin-2, with its ten EGF-like domains. Possibly these long forms of matrilin-4 functionally compensate for the lack of matrilin-2 in zebrafish, in particular, as *Fugu* matrilin-2 lacks the unique sequence present in mammalian matrilin-2 (R. Wagener, unpublished work) and is therefore very similar to matrilin-4.

The zebrafish shows a greater variety of matrilin splice variants than mammals. Predominantly, the alternative splicing of matrilins in zebrafish affects the number of EGF-like domains. In zebrafish, variants of matrilin-1, -3a and -3b occur that completely lack EGF-like domains. Further there is a high identity found between the orthologue VWA domains and the lower identity among EGF-like domains and coiled-coil domains. Together this indicates that the VWA domains, and not the EGF-like domains, are the

principal interaction modules of matrilins and that the EGF-like domains act mainly as spacers. A zebrafish matrilin-3b splice variant contains a sequence rich in proline and threonine/serine residues. This sequence is unique among matrilins in all species studied and is encoded by a single exon. Sequences rich in threonine/serine residues are often targets for a mucin-type *N*-acetyl-galactosamine O-glycosylation, and if such modification really takes place it may drastically influence the physical properties of this matrilin domain.

Duplicated genes are common in zebrafish and are probably the result of a duplication of the whole genome. The teleostii, to which zebrafish belongs, diverged from the tetrapoda, including the mammals, about 450 million years ago, soon after the divergence of actinopterygii from sarcopterygii. It has been proposed that 20% or more of the genes duplicated in the presumed duplication event have survived in zebrafish [39]. Unfortunately, due to our recent discovery of matrilin-3b, just shortly before the completion of this study, we have not yet been able to study the difference between matrilin-3a and -3b in depth. Nevertheless, although the amino acid sequences of the VWA domains in matrilin-3a and -3b are highly conserved (80% identity; 87% similarity), the antibodies we raised against the matrilin-3a VWA domain react only very weakly with matrilin-3b (Figure 8). It is likely that the two forms are no longer redundant and that by the recruitment of the proline- and threonine/serine-rich module, matrilin-3b has evolved a new function that has prevented the loss of this co-orthologue during evolution. Matrilin-3b splice variants carrying the proline- and threonine/serine-rich module were predominantly detected in embryos and not in adult fish. Therefore, such a new function may be of developmental significance.

The temporal and spatial expression pattern of zebrafish matrilins is similar to that of their mouse orthologues [15]. Matrilin-1 and -3 have a skeletal expression, whereas matrilin-4 is more widespread and being additionally expressed in loose connective tissue and epithelia. As in mammals, some expression of matrilin-1 [13] in non-skeletal tissues could be detected, whereas matrilin-3 was found only in cartilage. In both zebrafish and mammals, matrilins are more highly expressed during development with matrilin-4 showing the earliest onset. All zebrafish matrilins are present in cartilage, but their distributions are only partially overlapping, which is again reminiscent of the zonal expression of mouse matrilins in the growth plate and the articular cartilage [15].

The analysis of the contigs that contain the matrilin genes reveals partial synteny with the corresponding human and mouse chromosomes. The duplicated matrilin-3 genes represent end points of syntenic areas present on zebrafish chromosomes 13 and 20, but in this segment only the matrilin-3 gene is present on both chromosomes (Figure 5B), indicating not only the loss of the neighbouring duplicated genes, but also evolutionary chromosome breaks after the duplication of the genome. Alternatively, the duplication of the matrilin-3 gene in zebrafish could be the result of a segmental duplication. When comparing the genomes of human and mouse, segmental duplications are often found associated with syntenic rearrangements [40]. The exon-intron organization of the matrilin genes is identical between mammals and zebrafish, except for the insertion of an intron in zebrafish matrilin-1 and an exon in zebrafish matrilin-3b genes. Also the intron phases are conserved. The conservation of U12-type introns, interrupting the exons coding for the coiled-coil domain, points to the importance of this minor splice form in matrilin genes. It is thought that U12-type introns represent an ancient form of introns that have mostly been exchanged for U2-type introns during evolution [41]. It could be that the splicing of remaining U12-type introns is important for the regulation of the matrilin

mRNA expression, as it was shown in *Drosophila* that the splicing of U12-type introns can be a rate-limiting step in gene expression [42].

In zebrafish, similar to mammals, the genes for matrilin-4 and RBP-L divergently overlap at the 5' end in exonic and intronic sequences and also have an overlapping exon each. We do not yet understand the functional consequences of this overlap. In mouse, where the overlap is significantly longer than in zebrafish on the mRNA level, no evidence for an antisense effect was found [38] and the very short overlap of only 5 nt on the mRNA in antisense orientation in zebrafish further indicates that any potential antisense function is not strongly conserved. The more marked conservation at the gene level may indicate that the overlap region may be the target for transcription factors that act on both the matrilin-4 and the RBP-L gene.

Single knockouts of matrilin-1, -2 and -3 in mouse gave no obvious phenotype [23–26], possibly due to a redundancy in matrilin function. Nevertheless, the presence of highly conserved zebrafish orthologues, showing similar spatial and temporal expression, points to an important biological function of this gene family. Our analysis of zebrafish matrilins will allow future genetic analysis of matrilin function in a second species.

We thank Robert Geisler for support on radiation hybrid mapping, Alexander Reugels for supplying zebrafish, and Neil Smyth and Frank Zaucke for critical reading of this manuscript. This work was supported by grants from the Deutsche Forschungsgemeinschaft (WA 1338/2-2 and WA 1338/2-3) and the Köln Fortune program of the Medical Faculty of the University of Cologne. Y.-P.K. is a student in the International Graduate School in Genetics and Functional Genomics at the University of Cologne, funded by the State of Northrhine-Westphalia.

REFERENCES

- Deák, F., Wagener, R., Kiss, I. and Paulsson, M. (1999) The matrilins: a novel family of oligomeric extracellular matrix proteins. *Matrix Biol.* **18**, 55–64.
- Gibson, T. J. and Spring, J. (2000) Evidence in favour of ancient octaploidy in the vertebrate genome. *Biochem. Soc. Trans.* **28**, 259–264.
- Hauser, N. and Paulsson, M. (1994) Native cartilage matrix protein (CMP). A compact trimer of subunits assembled via a coiled-coil α -helix. *J. Biol. Chem.* **269**, 25747–25753.
- Klatt, A. R., Nitsche, D. P., Kobbe, B., Macht, M., Paulsson, M. and Wagener, R. (2001) Molecular structure, processing, and tissue distribution of matrilin-4. *J. Biol. Chem.* **276**, 17267–17275.
- Piecha, D., Muratoglu, S., Mörgelin, M., Hauser, N., Studer, D., Kiss, I., Paulsson, M. and Deák, F. (1999) Matrilin-2, a large, oligomeric matrix protein, is expressed by a great variety of cells and forms fibrillar networks. *J. Biol. Chem.* **274**, 13353–13361.
- Klatt, A. R., Nitsche, D. P., Kobbe, B., Mörgelin, M., Paulsson, M. and Wagener, R. (2000) Molecular structure and tissue distribution of matrilin-3, a filament-forming extracellular matrix protein expressed during skeletal development. *J. Biol. Chem.* **275**, 3999–4006.
- Kleemann-Fischer, D., Kleemann, G. R., Engel, D., Yates, J. R., Wu, 3rd, J. J. and Eyre, D. R. (2001) Molecular properties of matrilin-3 isolated from human growth cartilage. *Arch. Biochem. Biophys.* **387**, 209–215.
- Wu, J. J. and Eyre, D. R. (1998) Matrilin-3 forms disulfide-linked oligomers with matrilin-1 in bovine epiphyseal cartilage. *J. Biol. Chem.* **273**, 17433–17438.
- Zhang, Y. and Chen, Q. (2000) Changes of matrilin forms during endochondral ossification. Molecular basis of oligomeric assembly. *J. Biol. Chem.* **275**, 32628–32634.
- Frank, S., Schulthess, T., Landwehr, R., Lustig, A., Mini, T., Jenö, P., Engel, J. and Kammerer, R. A. (2002) Characterization of the matrilin coiled-coil domains reveals seven novel isoforms. *J. Biol. Chem.* **277**, 19071–19079.
- Aszódi, A., Modis, L., Paldi, A., Rencendorj, A., Kiss, I. and Bösze, Z. (1994) The zonal expression of chicken cartilage matrix protein gene in the developing skeleton of transgenic mice. *Matrix Biol.* **14**, 181–190.
- Aszódi, A., Hauser, N., Studer, D., Paulsson, M., Hiripi, L. and Bösze, Z. (1996) Cloning, sequencing and expression analysis of mouse cartilage matrix protein cDNA. *Eur. J. Biochem.* **236**, 970–977.
- Mundlos, S. and Zabel, B. (1994) Developmental expression of human cartilage matrix protein. *Dev. Dyn.* **199**, 241–252.
- Segat, D., Frie, C., Nitsche, P. D., Klatt, A. R., Piecha, D., Korpos, E., Deák, F., Wagener, R., Paulsson, M. and Smyth, N. (2000) Expression of matrilin-1, -2 and -3 in developing mouse limbs and heart. *Matrix Biol.* **19**, 649–655.

- 15 Klatt, A. R., Paulsson, M. and Wagener, R. (2002) Expression of matrilins during maturation of mouse skeletal tissues. *Matrix Biol.* **21**, 289–296
- 16 Paulsson, M. and Heinegård, D. (1979) Matrix proteins bound to associatively prepared proteoglycans from bovine cartilage. *Biochem. J.* **183**, 539–545
- 17 Hauser, N., Paulsson, M., Heinegård, D. and Mörgelin, M. (1996) Interaction of cartilage matrix protein with aggrecan. Increased covalent cross-linking with tissue maturation. *J. Biol. Chem.* **271**, 32247–32252
- 18 Winterbottom, N., Tondravi, M. M., Harrington, T. L., Klier, F. G., Vertel, B. M. and Goetinck, P. F. (1992) Cartilage matrix protein is a component of the collagen fibril of cartilage. *Dev. Dyn.* **193**, 266–276
- 19 Wiberg, C., Klatt, A. R., Wagener, R., Paulsson, M., Bateman, J. F., Heinegård, D. and Mörgelin, M. (2003) Complexes of matrilin-1 and biglycan or decorin connect collagen VI microfibrils to both collagen II and aggrecan. *J. Biol. Chem.* **278**, 37698–37704
- 20 Mann, H. H., Özbek, S., Engel, J., Paulsson, M. and Wagener, R. (2004) Interactions between the cartilage oligomeric matrix protein and matrilins. Implications for matrix assembly and the pathogenesis of chondrodysplasias. *J. Biol. Chem.* **279**, 25294–25298
- 21 Chapman, K. L., Mortier, G. R., Chapman, K., Loughlin, J., Grant, M. E. and Briggs, M. D. (2001) Mutations in the region encoding the von Willebrand factor A domain of matrilin-3 are associated with multiple epiphyseal dysplasia. *Nat. Genet.* **28**, 393–396
- 22 Briggs, M. D. and Chapman, K. L. (2002) Pseudoachondroplasia and multiple epiphyseal dysplasia: mutation review, molecular interactions, and genotype to phenotype correlations. *Hum. Mutat.* **19**, 465–478
- 23 Aszódi, A., Bateman, J. F., Hirsch, E., Baranyi, M., Hunziker, E. B., Hauser, N., Böse, Z. and Fässler, R. (1999) Normal skeletal development of mice lacking matrilin 1: redundant function of matrilins in cartilage? *Mol. Cell. Biol.* **19**, 7841–7845
- 24 Huang, X., Birk, D. E. and Goetinck, P. F. (1999) Mice lacking matrilin-1 (cartilage matrix protein) have alterations in type II collagen fibrillogenesis and fibril organization. *Dev. Dyn.* **216**, 434–441
- 25 Ko, Y., Kobbe, B., Nicolae, C., Miosge, N., Paulsson, M., Wagener, R. and Aszódi, A. (2004) Matrilin-3 is dispensable for mouse skeletal growth and development. *Mol. Cell. Biol.* **24**, 1691–1699
- 26 Mates, L., Nicolae, C., Mörgelin, M., Deák, F., Kiss, I. and Aszódi, A. (2004) Mice lacking the extracellular matrix adaptor protein matrilin-2 develop without obvious abnormalities. *Matrix Biol.* **23**, 195–204
- 27 Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P. et al. (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36
- 28 Shin, J. T. and Fishman, M. C. (2002) From zebrafish to human: modular medical models. *Annu. Rev. Genomics Hum. Genet.* **3**, 311–340
- 29 Kohfeldt, E., Maurer, P., Vannahme, C. and Timpl, R. (1997) Properties of the extracellular calcium binding module of the proteoglycan testican. *FEBS Lett.* **414**, 557–561
- 30 Smyth, N., Odenthal, U., Merkl, B. and Paulsson, M. (2000) Eukaryotic expression and purification of recombinant extracellular matrix proteins carrying the Strep II tag. *Methods Mol. Biol.* **139**, 49–57
- 31 Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6
- 32 Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164
- 33 Jenkins, R. N., Osborne-Lawrence, S. L., Sinclair, A. K., Eddy, Jr, R. L., Byers, M. G., Shows, T. B. and Duby, A. D. (1990) Structure and chromosomal location of the human gene encoding cartilage matrix protein. *J. Biol. Chem.* **265**, 19624–19631
- 34 Aszódi, A., Beier, D. R., Hiripi, L., Böse, Z. and Fässler, R. (1998) Sequence, structure and chromosomal localization of *Crtm* gene encoding mouse cartilage matrix protein and its exclusion as a candidate for murine achondroplasia. *Matrix Biol.* **16**, 563–573
- 35 Kiss, I., Deák, F., Holloway, Jr, R. G., Delius, H., Meubst, K. A., Frimberger, E., Agraves, W. S., Tsonis, P. A., Winterbottom, N. and Goetinck, P. F. (1989) Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. Exon/intron organization, unusual splice sites, and relation to alpha chains of $\beta 2$ integrins, von Willebrand factor, complement factors B and C2, and epidermal growth factor. *J. Biol. Chem.* **264**, 8126–8134
- 36 Wagener, R., Kobbe, B., Aszódi, A., Liu, Z., Beier, D. R. and Paulsson, M. (2000) Structure and mapping of the mouse matrilin-3 gene (*Matn3*), a member of a gene family containing a U12-type AT–AC intron. *Mamm. Genome* **11**, 85–90
- 37 Wagener, R., Kobbe, B. and Paulsson, M. (1998) Genomic organisation, alternative splicing and primary structure of human matrilin-4. *FEBS Lett.* **438**, 165–170
- 38 Wagener, R., Kobbe, B., Aszódi, A., Aeschlimann, D. and Paulsson, M. (2001) Characterization of the mouse matrilin-4 gene: a 5' antiparallel overlap with the gene encoding the transcription factor RBP-1. *Genomics* **76**, 89–98
- 39 Postlethwait, J. H., Woods, I. G., Ngo-Hazelett, P., Yan, Y. L., Kelly, P. D., Chu, F., Huang, H., Hill-Force, A. and Talbot, W. S. (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* **10**, 1890–1902
- 40 Bailey, J. A., Baertsch, R., Kent, W. J., Haussler, D. and Eichler, E. E. (2004) Hotspots of mammalian chromosomal evolution. *Genome Biol.* **5**, R23
- 41 Patel, A. A. and Steitz, J. A. (2003) Splicing double: insights from the second spliceosome. *Nat. Rev. Mol. Cell. Biol.* **4**, 960–970
- 42 Patel, A. A., McCarthy, M. and Steitz, J. A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* **21**, 3804–3815
- 43 Geisler, R., Rauch, G. J., Baier, H., van Bebber, F., Bross, L., Dekens, M. P., Finger, K., Fricke, C., Gates, M. A., Geiger, H. et al. (1999) A radiation hybrid map of the zebrafish genome. *Nat. Genet.* **23**, 86–89
- 44 Schilling, T. F., Piotrowski, T., Grandel, H., Brand, M., Heisenberg, C. P., Jiang, Y. J., Beuchle, D., Hammerschmidt, M., Kane, D. A., Mullins, M. C. et al. (1996) Jaw and branchial arch mutants in zebrafish I: branchial arches. *Development* **123**, 329–344

Received 1 September 2004/5 November 2004; accepted 9 December 2004

Published as BJ Immediate Publication 9 December 2004, DOI 10.1042/BJ20041486