J Mol Evol (1996) 43:263–269

JOURNAL OF MOLECULAR EVOLUTION

Evolution of Disintegrin Cysteine-Rich and Mammalian Matrix-Degrading Metalloproteinases: Gene Duplication and Divergence of a Common Ancestor Rather Than Convergent Evolution

Ana M. Moura-da-Silva,[1–3] R. David G. Theakston,[2] Julian M. Crampton[1]

[1] Wolfson Unit of Molecular Genetics, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK
[2] Venom Research Unit, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK
[3] Laboratorio de Imunopatologia, Instituto Butantan, Sāo Paulo, Brazil

Received: 26 December 1995 / Accepted: 16 April 1996

Abstract. The evolution of the Metalloproteinase Disintegrin Cysteine-rich (MDC) gene family and that of the mammalian Matrix-degrading Metalloproteinases (MMPs) are compared. The alignment of snake venom and mammalian MDC and MMP precursor sequences generated a phylogenetic tree that grouped these proteins mainly according to their function. Based on this observation, a common ancestry is suggested for mammalian and snake venom MDCs; it is also possible that gene duplication of the already-assembled domain structure, followed by divergence of the copies, may have significantly contributed to the evolution of the functionally diverse MDC proteins. The data also suggest that the structural resemblance of the zinc-binding motif of venom MDCs and MMPs may best be explained by common ancestry and conservation of the proteolytic motifs during the divergence of the proteins rather than through convergent evolution.

Key words: Metalloproteinase — Disintegrin — Evolution — Venom — Phylogeny — Gene duplication

Introduction

Matrix-degrading metalloproteinases (MMPs) are a group of related enzymes involved in extracellular degradation of the protein components of the connective tissue. They are involved in the tissue remodeling which occurs during embryogenesis, development, and wound healing. They may also be involved in certain diseases such as arthritis, periodontitis, and tumor metastasis. The proteolytic activity of these enzymes is dependent on the presence of a consensus zinc-binding motif in their catalytic domain (Blundell 1994). The same zinc-binding motif is present in certain toxins of viper venoms which induce hemorrhage in the prey or victims of snake bite. They act on similar substrates to MMPs present in the basement membrane of the endothelium (Bjarnason and Fox 1994). Both snake venom and matrix metalloproteinases are classified in the Astacin family of zinc metalloproteinases based on the structure of the catalytic domain and zinc-binding motif; however, the sequence and structure of the surrounding areas are variable (Blundell 1994).

Other components of viper venoms that interfere with hemostasis are the disintegrins. These comprise a family of RGD-containing peptides with high affinity for the platelet glycoprotein GP IIb/IIIa integrin receptor (Gould et al. 1990). They prevent platelet aggregation by inhibiting the binding of fibrinogen and von Willebrand factor to platelets (Huang et al. 1989).

Interestingly, like MMPs both hemorrhagic toxins and disintegrins are coded by cDNAs that predict zymogen molecules containing a large precursor pro-peptide domain followed by a metalloproteinase domain. These precursor proteins differ from MMPs in having a car-

Correspondence to: J.M. Crampton

boxy-terminal disintegrin domain of varying length which replaces the hemopexin domain present in MMPs (Paine et al. 1992). The pro-peptide domain is apparently involved in the inactivation of the proteolytic activity prior to the secretion of the enzymes by venom gland cells. This process is thought to occur through a cysteine switch mechanism, as also proposed for the MMPs (Woessner 1991), and is correlated with the sequence PRCGV in MMPs, and with PKMCGV in the snake toxins. The metalloproteinase domain has a zinc-binding sequence, HEXXHXXGXXH, characteristic of the Astacin family described above (Blundell 1994). The disintegrin domain is capable of binding to the platelet integrins through the RGD motif that is classically represented in venom disintegrins (Huang et al. 1989). Large hemorrhagins contain a region of sequence similarity with the disintegrins, with a substitution of the RGD motif with E/DCD at the same position, and a cysteine-rich extension of the carboxy-terminus (Paine et al. 1994).

It has recently become apparent that a number of cysteine-rich proteins with similar structural organizations occur in the mammalian male reproductive tract. The best characterized of these are Fertilin (formerly PH-30), a guinea pig sperm surface protein involved in sperm–egg fusion (Blobel et al. 1992), and the androgen-regulated epididymal apical protein-I, EAP-I (Perry et al. 1992). These proteins include a carboxy-terminal extension linked to a venom-like long hemorrhagin sequence comprising an epidermal growth factor (EGF) repeat, a transmembrane domain, and a cytoplasmatic tail (Weskamp and Blobel 1994). Representatives of this protein family were also detected in other tissues, including the MS-2 antigen present on the surface of certain lineages of macrophages (Yoshida et al. 1990) and the protein predicted by the MDC gene, which has been associated with tumor suppression (Emi et al. 1993).

All the mammalian proteins possess a non-RGD disintegrin domain, characteristic of the large hemorrhagins. The zinc-binding motif is present only in EAPs, Fertilin α, and the monocyte surface antigen metalloproteinase domains, and is not yet fully correlated with the function of these proteins. The cysteine switch motif is absent in all the mammalian pro-domains. These cysteine-rich mammalian proteins plus the snake venom hemorrhagins and disintegrins make up the metalloproteinase, disintegrin, cysteine-rich (MDC) family (Fig. 1).

The evolutionary aspects of the complex MDC gene family are still poorly understood. Paine et al. (1994) suggested that accelerated evolution may apply to the snake toxins, specifically with regard to their high variability occurring mainly in the metalloproteinase domain. They also suggested that the generation of the RGD disintegrins might be associated with a deletion in the DNA region that codes for the cysteine-rich domain in long hemorrhagins. Recently, Wolfsberg et al. (1993) suggested, based on phylogenetic trees, that the individual domains of the MDCs may have been assembled before the divergence of the members of the family. This last report does not, however, adequately explain the close functional similarity between the venom metalloproteinases and MMPs as opposed to the mammalian MDCs. We now report an analysis of MMP and MDC sequences undertaken precisely to address this issue. The aim was to study the evolutionary aspects of the MDCs in comparison to MMPs in order to provide new insights into the evolution of these gene families.

Materials and Methods

The sequences used in this paper were obtained by scanning GenBank, Swiss-Prot, and PIR databases using the BLAST program (Altschul et al. 1990). The sequences of the following proteins, predicted from cDNA sequences, were used: Human Matrilysin (gb L22524), collagenase 3 (gb X75308), and tumor suppressor gene product MDC (gb D17390); mouse cyritestin (pir S18968) and monocyte surface antigen (pir A60385); rat Stromelysin-2, MMP-10 (sp P07152), and epididymal protein, EAP (pir S28259); guinea pig Fertilin α (gb Z11719) and β (gb Z11720); monkey epididymal protein, EAP (pir S28258), tMDC-I (gb X76637) and II (gb X77619); Rhodostomin (pir S33792) from *Calloselasma rhodostoma* snake venom; Atrolysin E (pir A43296), B (pir S41608), C (pir S41609), and Catrocollastatin (gb U21003) from *Crotalus atrox* venom; Trigramin (pir A30065) from *Trimeresurus gramineus* venom; Jararhagin (pir S24789) from *Bothrops jararaca* venom; Trimucin (pir S43125) from *Trimeresurus mucrosquamatus* venom; a hemorrhagin (gb U18234) from *Agkistrodon contortrix* venom; Halystatin (gb D28870) from *Agkistrodon halys* venom; EcH I (gb X78970), and EcH II (gb X78971) from *Echis pyramidum leakeyi* venom. Because of recent taxonomic reorganisation, it is likely that *Tr. gramineus* and *A. halys* referred to here now correspond to *Tr. stejnegeri* and *A. blomhoffi brevicaudus*, respectively (D.A. Warrell, personal communication).

Percentage similarity and frequency of codon usage have been calculated using the GCG software, bestfit, and codon usage program (Genetics Computer Group 1991). Phylogenies were constructed using the Clustal V program (Higgins et al. 1992). Progressive alignments used the multiple alignment algorithms described by Higgins and Sharp (1989), with a fixed gap penalty of 10 and the Dayhoff PAM 250 protein weight matrix (Dayhoff et al. 1978). Alignments were finally refined by eye and differ very slightly from those generated by computer. The phylogenetic trees according to distances were generated from the above alignments using a neighbor-joining method (Saitou and Nei 1987) and rooted using the Thermolysin sequence (pir M21663) from *Bacillus stearothermophilus.* The degree of error was calculated for each branch by bootstrapping (Felsenstein 1985) and the values are shown in the tree. Trees were constructed using the whole protein sequence, as well as for the pro-protein, metalloproteinase, and disintegrin domains. In all cases phylogenies constructed with sites where gaps occurred in any one sequence of the alignment being deleted and not deleted were compared.

Results and Discussion

Sequence Similarity Between MDCs and MMPs Appears to Be Restricted to the Functional Motifs

This study has been carried out using the precursor sequences predicted by the cDNAs coding for three MMPs,

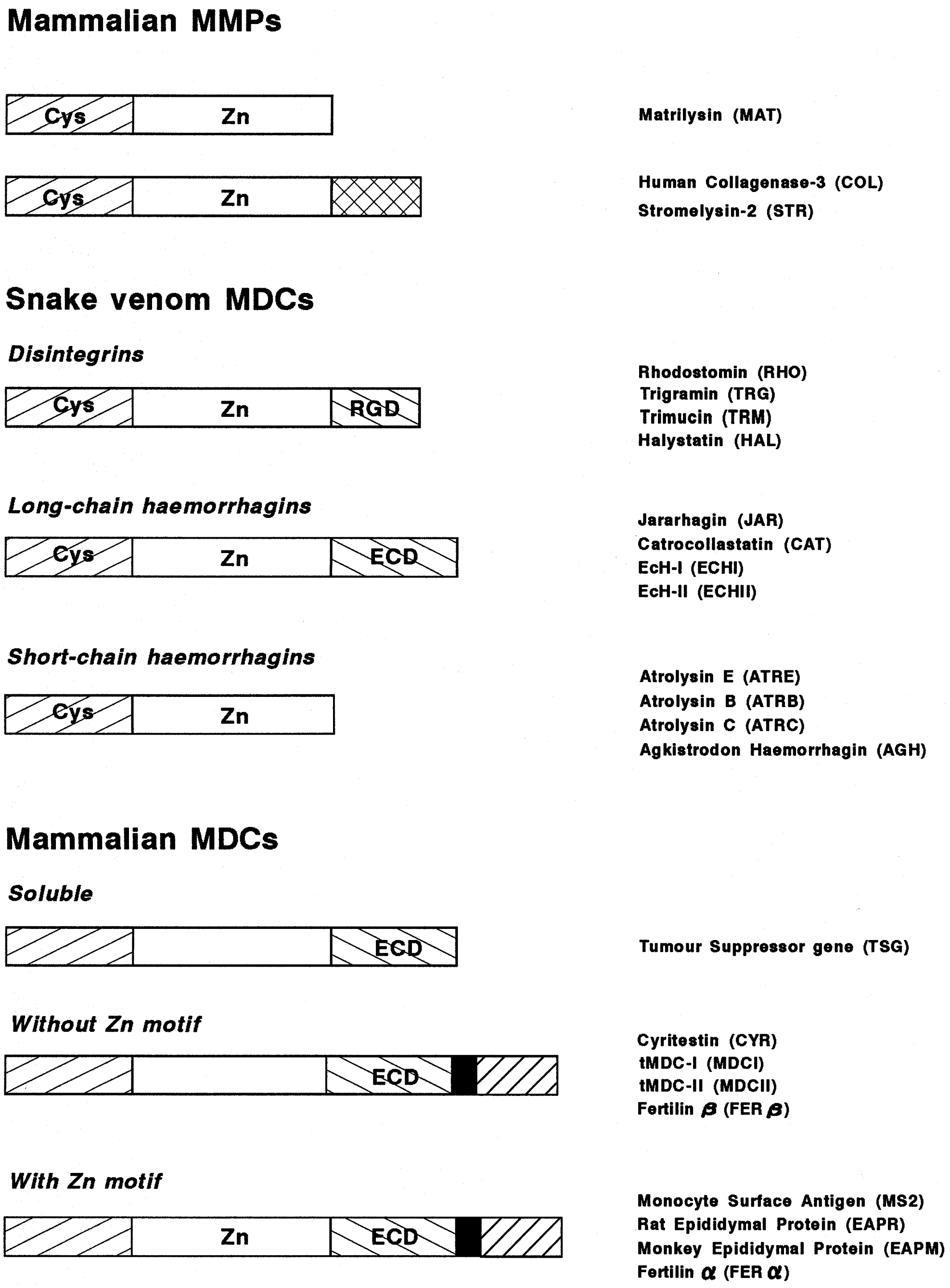

Fig. 1. Schematic representation of mammalian matrix-degrading metalloproteinases (MMPs) and snake venom and mammalian proteins belonging to the metalloproteinase-like, disintegrin-like, cysteine-rich (MDC) gene family: (▨) Pro-peptide domain; (□) metalloproteinase domain; (▧) disintegrin domain; (▩) hemopexin; (■) EGF repeats; (▨) transmembrane domain. The presence of the cysteine-switch and zinc-binding motifs is represented by Cys and Zn, respectively. RGD and ECD represent the potential integrin binding peptides in the disintegrin domain.

nine mammalian MDCs, and 12 snake venom MDCs. The criteria used to select these sequences was based on a wider representation of the functional diversity among these proteins and also a number of representatives from different genera/species of snakes. Figure 2 shows two representative fragments extracted from the alignment of the whole precursor sequences, containing, respectively, the cysteine-switch and zinc-binding motifs. The similarity between snake and mammalian MDCs is recognized by the conservation of most cysteines and other residues throughout all the MDC sequences (Fig. 2, stars). Some conserved positions among snake toxins are also shared with few, but not all, mammalian MDCs (Fig. 2, dots). Therefore, MDCs may be considered a family of proteins with high sequence similarity. However, the comparison of MDCs with MMPs reveals little sequence similarity (below 45%), this being confined to the zinc-binding and cysteine switch motifs, the latter being observed only in MMPs and snake venom MDCs. These observations also apply to the conformation of

Table 1. Frequency of codon usage for the essential residues for the catalytic activity in cDNAs regions coding for the zinc-binding motif (Zn) and the whole precursor molecules (WP) of MMPs and proteolytic MDCs[a]

		His			Glu			Gly	
		WP	Zn		WP	Zn		WP	Zn
AGH	cat	0.87	1.0	gag	0.5	1.0	ggg	0.16	—
	cac	0.13	—	gaa	0.5	—	gga	0.44	—
							ggt	0.28	—
							ggc	0.12	1.0
JAR	cat	0.68	1.0	gag	0.36	1.0	ggg	0.09	—
	cac	0.32	—	gaa	0.64	—	gga	0.51	—
							ggt	0.19	—
							ggc	0.21	1.0
TRG	cat	0.82	1.0	gag	0.33	1.0	ggg	0.13	—
	cac	0.18	—	gaa	0.67	—	gga	0.48	—
							ggt	0.16	—
							ggc	0.23	1.0
ECHI	cat	0.69	1.0	gag	0.38	1.0	ggg	0.13	—
	cac	0.31	—	gaa	0.62	—	gga	0.49	—
							ggt	0.15	—
							ggc	0.23	1.0
EAPM	cat	0.69	1.0	gag	0.32	—	ggg	0.16	1.0
	cac	0.31	—	gaa	0.68	—	gga	0.56	—
							ggt	0.14	—
							ggc	0.14	—
FERα	cat	0.27	0.33	gag	0.46	1.0	ggg	0.24	—
	cac	0.73	0.67	gaa	0.54	—	gga	0.31	—
							ggt	0.27	1.0
							ggc	0.18	—
COL	cat	0.67	0.33	gag	0.41	1.0	ggg	0.14	—
	cac	0.33	0.67	gaa	0.59	—	gga	0.29	—
							ggt	0.31	1.0
							ggc	0.26	—
MAT	cat	0.60	0.33	gag	0.22	—	ggg	0.19	—
	cac	0.40	0.67	gaa	0.78	1.0	gga	0.33	—
							ggt	0.14	1.0
							ggc	0.33	—

[a] Abbreviations as in Fig. 1

these proteins. A comparison of the crystal structure of Adamalysin II, a snake venom metalloproteinase, with the crayfish Astacin reveals some topologically equivalent residues (Gomis-Ruth et al. 1993). However, only the active site environment, comprising the zinc-binding consensus region and the active site basement, appears to exhibit identical conformation (Table 1).

The Evolutionary History of MDCs

Evolutionary trees have been constructed using the alignments of the complete precursor proteins (Fig. 3) or the regions comprising the distinct domains of the sequences described above (data not shown). The tree constructed using the complete sequences (Fig. 3) shows two primary divergent groups comprising the MMPs and MDCs, respectively. The MDC cluster is apparently monophyletic and the sequences are distributed mainly according to their function. The first group contains the sperm proteins related to Fertilin, and the second group includes the EAPs. The tumor suppressor gene and the monocyte surface antigen are located in the first and second groups, respectively. Snake venom sequences are also distributed according to function, the first group representing the long-chain hemorrhagic toxins, the second clustering the RGD-disintegrins, and the third enclosing the short hemorrhagins. Clearly, some of these sequences are quite dissimilar, and it is important to note that the clustering of the different molecules in the tree remains essentially the same even when all sites containing alignment gaps are removed from the data set during tree construction. In this situation, the analysis to some extent favors the combined pro- and metalloproteinase domains of the molecules.

Trees corresponding to the pro-, metalloproteinase or disintegrin domains were also constructed using the alignments corresponding to the relevant fragment of each toxin. The same clustering characteristics are suggested by the trees generated using whole sequences (Fig. 3) or independent domains (data not shown). The only exceptions were the long-chain hemorrhagins, Jararhagin and Catrocollastatin, from pit vipers, which appear to cluster preferentially with other pit viper se-

```
          ...156                                                                                     248...
MAT          GKLSPYIMEIMQKPRCGV------PDVAEYSLMP--NSPKW-----HSRIVTYRIVSYTSDLPRIVVDQIV---------------------
STR          GKLDSNTVEMMHKPRCGV------PDVGGFSTFP--GSPKW-----RKNHISYRIVNYTLDLPRESVDSAI---------------------
COL          GKLDDNTLDVMKKPRCGV------PDVGEYNVFP--RTLKW-----SKMNLTYRIVNYTPDMTHSEVEKAF---------------------
TSG          GKLRGNPHSFAALSTCQGLHGVFSDGNLTYIVEPQEVAGPWGAPQGPLPHLIYRTPLLPDPLGCREPGCLFAVPAQSAPPNRPRLRRKRQVR
FERα         GYIEGASSSFVSVSACSGLRGILIKENTSYGIEPILSSQR-------FEHVLYT---MARQAPVSCR-ASAKDSQAVSTSWQQGSRKPHSVQ
FERβ         GHIEGFPTSLASISTCAGLRGLLQFETVSYGIEPLKSSIG-------FEHVIYP---V-KHDNEKSQ-YLKKSINV--K---NVVYKI----
MDCI         GYVADIPKSAVTLRTCSGLRGLLQLDNISYGIEPLESSPT-------YEHVVYR---I-KNDAIGHF-SFQENYPV--AQYIDQSYRILVKS
MDCII        GYVAGIPNSLVTLSVCSGLRGTMQLKNISYGIEPMEAVSG-------FIHKIYE---E-KFADTNI---LLEENDT--YSWFNSEYQVRKSS
CYR          GHAAEIPVSTVTLSTCSGLRGLLQLENITYGIEPLESSAT-------FEHILYE---I-KNNKIDYS-PLKENFAN--SEQESQSYRILVKP
MS2          GHVEGYEGSAASISTCAGLRGFFRVGSTVHLIEPLDADEE-------GQHAMYQAKHLQQKAGT----CGVKDTNLNDLGP--RALEIYRAQ
EAPR         GSIIHEFDSAASISTCNGLRGFFRVNDQRYLIEPVKYSDE-------GDHLVFKY-NVKAPYATNYS-CEGLNFTKKSTLIDAKIIE----E
EAPM         GSIVHEYDSAASISTCNGLRGFFRVNDQRYLIEPVKYSDE-------GEHLVFKY-NPRVPYVANYS-CTELNFTRKTVPGDTESEG----D
EcHI         GRIQNDADSTASMSACNGLKGYFMLRGETYLIEPLKIPDS-------EAHAVYKYENVEKEDEAPKM-CGVTQTNWES-DELKKASQLVATS
EcHII        GRVQNDAHSSASISACNGLKGHFKLQGETYFIEPLKIPDS-------EAHAVYKYENIEKEDQAPKM-CGVTHTNWESDEPIKEASRLVASS
JAR          GRIENDADSTASISACNGLKGYFKLQRETYFIEPLKLPDS-------EAHAVFKYENVEKEDEAPKM-CGVTQ-NWKSYEPIKKASQLAFTA
CAT          GRIENDADSTASISACNGLKGHFKLQGEMYLIEPLKLPDS-------EAHAVYKYENVEKEDEALKM-CGVTQ-NWESYEPIKKASQLVVTA
RHO          GRIHNDADSTASISACNGLKGHFKLQGETYFIEPMKLPDS-------EAHAVFKYENIEKEDESPKM-CGVTETNWESDEPIKKVSQLNLNH
TRM          GRIHNDADSTASISACDGLKGYFKLQGETYPIEPLELSDS-------EAHAVFKYENVEKEDEAPKM-CGVTQ-NWESDESIKKASQLYLTP
TRG          GRIENDADSTASISACDGLKGHFKLQGEMYLIEPLELSDS-------EAHAVFKYENVEKEDEPPKM-CGVTQ-NWESYESTKKASQLNVTP
HAL          GRVENDADSTASLSACDGLKAHFKIQGEMYLIEPLEVSDT-------DAHAVFKYENVEKEDEPPKM-CGVTQ-NWESYESTKKASQLNVSP
AGH          GRIQNDADSIASISACNGLKGHFKLQGEMYLIEPLELSDS-------EAHAVFKYENVEKEDEAPKI-CGVTQ-NWESYEPIKKASQLNLNY
ATRE         GRIENDADSTASISACNGLKGHFKLQGEMYLIEPLKLSDS-------EAHAVFKLKNVEKEDEAPKM-CGVTQ-NWESYEPIKKASDLNLNP
ARTB         GRIENDADSTASISACNGLKGHFKLQGEMYLIEPLELSDS-------EAHAVFKYENVEKEDEAPKM-CGVTQ-NWESYEPIKKASDLNLNP
ATRC         GRIENDADSTASISACNGLKGHFKLQGELYLIEPLELSDS-------EAHAVFKLENVEKEDEAPKM-CGVTQ-NWESYEPIKKASDLNLNP
             * .    .* ...*.*.** * .      ...**    .          . .    .       ... ....
```

```
          ...363                               +---------------+                         441...
MATR         GPGLGG---DAHFDKDEYWTDDEDAGVNFLFAAT|HEFGHSLGL---SHS|SVPGTV------------------------
STRM         GPGFYG---DAHFDDDEKWSLGPS-GTNLFLVAA|HELGHSLGL---FHS|NNKESL------------------------
COLG         GPNYGG---DAHFDDDETWTSSSK-GYNLFLVAA|HEFGHSLGL---DHS|KDPGAL------------------------
TSG          GAAYVGGICSLSHGGGVNEYGN---MGAMAVTLA|QTLGQNLGMMWNKHR|SSAGDCKCPDIWLG--CIMEDTGFY--LPR
FERa         GQAFLNGACSSGFAAAVEAFHHEDALLS-AALLV|HELGHNLGI---RHD|-HS-ACVCRDKH---SCLMQENITEESG--
FERb         GAIFQGMICNTSYGGGIALHSKTITLDSFGVILV|QLLSVSMGI---AYD|-NADLCRCRGA----ICLMSPEAVFSSGMK
MDC1         GATYHGMACDPKFATGIALYPKKITVEAFSVVMA|QLLGINLGL---TYD|-DIYNCYCPGP----TCIMNPDAIRSHGMK
MDC2         GATFPGQVCNKDFAAAVALYPEGLSLESYTVIIV|QLLGLNLGL---TYD|-KTDTCHCSGD----VCTMTPKAVYSGGVK
CYRT         GATYHGMACNPNFTAGIALHPKTLAVEGFAIVLS|QLLGINLGL---AYD|-DVYNCFCPGS----TCIMNPSAIRSQGIK
MS-2         GLAKVSALC-SRHSGAVNQDHSKNSIGV-ASTMA|HELGHNLGM---SHD|EDIPGCYCPEPREGGGCIMTE-SIGSKFPR
EAPr         GIAYPGGICQTLRSCSVVKDLLPDVNII-GNRMA|HQLGHSLGM---RHD|D-FP-CTCPL---GK-CVMGA---GSIPAI
EAPm         GISYPAGMCLPYYSTSIIKDLLPDTNII-ANRMA|HQLGHNLGM---QHD|E-FP-CTCPS---GK-CVMDS-D-GSIPAL
EcH1         GIARMRGMCSPSNSVGVIQDYCKNYLLV-AITMA|HELGHNLGM---DHD|N--GNCNCPD---TS-CIMSA-VAGPEPVF
EcH2         GLRDVSSMCQATRSVGVVQDHSPTVRAV-AVTMA|HEMGHNLGM---SHD|G--NHCNC-G---ANSCIMAA-VLRNPAPE
JARH         GYAYIGSMCHPKRSVGIVQDYSPINLVV-AVIMA|HEMGHNLGI---HHD|T--GSCSC-G---DYPCIMGP-TISNEPSK
CATR         GLAYVGSMCHPKRSTGIIQDYSEINLVV-AVIMA|HEMGHNLGI---NHD|S--GYCSC-G---DYACIMRP-EISPEPST
RHOD         GKAYLDSICDPERSVGIVQNYHGITLNV-AAIMA|HEMGHNLGV---RHD|G--EYCTCYG---SSECIMSS-HISDPPSK
TRIM         GWAYVGRMCDEKYSVAVVKDHSSKVFMV-AVTMT|HELGHNLGM---EHD|D-KDKCKC-D---T--CIMSA-VISDKQSK
TRIG         GRAPVGGMCDPKRSVAIVRDHNAIVFVV-AVTMT|HEMGHNLGM---HHD|E--DKCNC-N---T--CIMSK-VLSRQPSK
HALS         GRAPVGGMCDPKRSVAIVRDHNAILFIV-AVTMT|HEMGHNLGM---HHD|E--DKCNC-N---T--CIMSK-VLSRQPSY
AGKH         GLAYVGTMCDPKLSTGVVEDHSKINFLV-AVTMA|HEMGHNLGM---RHD|T--GSCSC-G---GYSCIMSP-VISDDSPK
ATRe         GRAYIGGICDPKRSTGVVQDHSEINLRV-AVTMT|HELGHNLGI---HHD|T--DSCSC-G---GYSCIMSP-VISDEPSK
ARTb         GRAYTSSMCNPRKSVGIVKDHSPINLLV-GVTMA|HELGHNLGM---NHD|G--DKCLR-G---ASLCIMRP-GLTPGRSY
ATRc         GLAPLGTMCDPKLSIGIVQDHSPINLLM-GVTMA|HELGHNLGM---EHD|G--KDCLR-G---ASLCIMRP-GLTKGRSY
             *       *          .            ..|.. *...*     ..|     * .        *.*
                                               +---------------+
```

Fig. 2. Segments extracted from the alignments of the MMP and MDC complete precursor sequences comprising the cysteine-switch and zinc-binding motifs (*bold*). The positions corresponding to residues conserved among all MDCs (*) or among snake toxins and at least two mammalian MDCs (•) are indicated. The zinc-binding motif and its homologue in nonproteolytic MDCs are *boxed.* The *numbers* correspond to the amino acid position in the consensus sequence of the original alignment. *Abbreviations* as in the legend to Fig. 1.

quences in metalloproteinase and pro-domain alignments, and Fertilin α, whose disintegrin domain does not cluster with other sperm proteins.

The apparent monophyletic distribution of MDC proteins and their independent domains suggests that both mammalian and snake venom proteins have evolved from a common ancestor gene (already assembled as the multidomain structure) both by speciation and by gene duplication. The evolution of a gene family with different functions is currently associated with rapid amino acid divergence among duplicated copies of the genes, thus increasing the functional diversity of the gene family (Ohta 1994). Certainly in the case of snakes, such diversity would seem to be of real benefit because it may result in rapid variation in venom toxicity which could broaden the spectrum of available prey (Daltry et al. 1996). In some cases, a single amino acid change in these molecules could significantly alter their toxicity in venom, and an accumulation of such changes may account for a rapid amino acid divergence. An example of this is the functionally distinct members of venom phospholipases A_2 (Moura-da-Silva et al. 1995) and members of the serine proteinase family (Creigton and Darby 1989), which are thought to have arisen through gene duplication followed by divergence of the copies through positive Darwinian selection causing sequence hypervariability. However, serine proteinases present a slightly different evolutionary history since the different domains of the molecules appear to have evolved as independent units rather than as an already-assembled block (Ikeo et al. 1995).

With regard to points in the tree at which gene duplications may have occurred, it seems reasonable to suggest that gene duplications generating the α and β chains

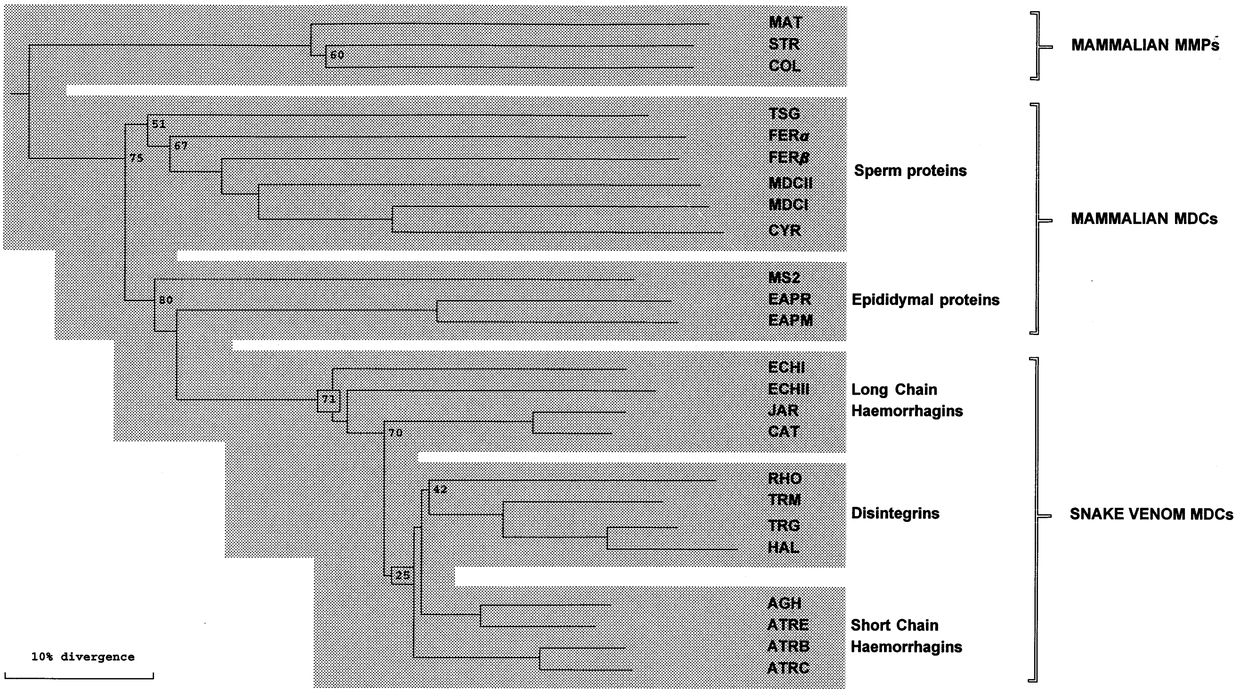

Fig. 3. Phylogenetic tree of the MDC proteins: All amino acid positions predicted by the cDNA regions coding for the whole protein have been included in the analysis. The *horizontal scale* is proportional to the calculated distance. The *numbers* indicated at nodes represent the percentage confidence limits by bootstrapping. Only those values less than 95% are indicated. *Abbreviations* as in the legend to Fig. 1.

of Fertilin and RGD and non-RGD venom disintegrins may have occurred. Gene duplications may also have taken place between EAPs and Fertilin and between short-chained hemorrhagins and RGD-disintegrins. The generation of the different venom MDC proteins has been correlated with post-translational proteolytic processing or RNA splicing (Bjarnason and Fox 1994). However, if these were the only mechanisms operating, a taxonomically related duplication would be expected, at least for the trees generated with the sequences comprising the metalloproteinase and pro-domains. The distribution of sequences showed by the trees does not eliminate the possible occurrence of different processing of RNA and translated proteins; however, if this is the case, alternative processing may be occurring on products generated by different gene copies.

The trees also suggest that the venom toxins arose relatively late during the evolution of MDCs. According to this data, venom metalloproteinases have appeared from a common ancestor gene only after mammals and reptiles diverged; copies of the gene having evolved in snakes to become venom toxins and, perhaps, proteins with some function in the male reproductive tract. However, no work has yet been carried out to identify MDC proteins in snake sperm cells or in the epididymis.

MDCs and MMPs: Evolutionary Conservation of the Proteolytic Motifs or Sequence Convergence?

With regard to the evolutionary history of the MDCs, how can one explain the similarity of the zinc-binding and cysteine-switch motifs of venom MDCs with MMPs? The similarity mainly confined to the proteolytic motifs has already been shown for microbial and mammalian metalloproteinases (Jongeneel et al. 1989), and Woessner (1991) considered that convergent evolution could explain such similarity. Convergent evolution could also be applied to explain the structural resemblance of the proteolytic motifs between snake venom MDCs and mammalian MMPs. Considering this possibility, the zinc-binding motif may have arisen after a gene duplication, during the early divergence of the reproductive tract MDCs. One copy might have accumulated point mutations introducing the histidines, conferring a selective advantage and resulting in the proteolytic MDCs, which became fully functional in the venon proteins. The other unmutated copy evolved as the nonproteolytic MDCs. In favor of this hypothesis, two out of the three histidines critical for the zinc-binding replace very conserved residues in nonproteolytic MDCs, glutamine, and tyrosine. A single point mutation is needed to change the triplet codon that codes for both amino acids to histidines. The third histidine shares a very variable position in nonproteolytic MDCs, where substitutions might be expected. The cysteine-switch motif may have arisen later due to the need to regulate the proteolytic activity. The mammalian proteolytic MDCs containing a cysteine in the same position of the pro-domain that might possess functional activity, but lack other postulated consensus regions between MMPs and venom MDCs that consist of four residues, PXCGV (Fig. 2).

However, we have to consider a second possibility in which a common ancestor gene to MDCs and MMPs may have carried the proteolytic motifs which were later lost in the nonproteolytic MDCs. This hypothesis is supported by the similarity between microbial and mammalian metalloproteinases. De Souza and Brentani (1993) suggest that the similarity of the zinc-binding motifs arose in these molecules by conservation of sequences present in ancestral genes, while other domains of the molecule, such as the hemopexin-like domain, could be a recent acquisition of the eukaryotic metalloproteinases. The topology of the phylogenetic tree presented in this paper would support the hypothesis that MMPs and MDCs are derived from a common ancestor bearing the zinc-binding motif. The recent assembly of the disintegrin domain on MDCs generated a new functional possibility for the molecules, possibly independent of the proteolytic activity. The selective pressure on mammalian MDCs could therefore be attributed to a cell-matrix/cell-cell adhesion function, thus explaining the loss of the proteolytic motif. On the other hand, the main toxicity of certain viper venoms seems to be proteolytic. This observation would suggest a high selective pressure, and hence conservation, of the proteolytic motifs in the venom MDCs, despite the marked divergence of the remaining parts of the molecules, presumably enhanced by positive Darwinian selection. This hypothesis therefore explains why only the proteolytic motif appears to be shared between MMPs and MDCs in relation to both their sequence and conformation (Fig. 2).

In conclusion, the functional diversity of the MDC proteins may have been generated by gene duplication and divergence of common ancestor genes. The functional proximity of venom MDCs and MMPs can best be explained by evolutionary conservation of the proteolytic motifs rather than by sequence convergence.

Acknowledgments. This research was funded by the Medical Research Council and the Wellcome Trust. J.M.C. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Sciences, and A.M.M.D.S. is the recipient of a Research Development Award in Tropical Medicine from the Wellcome Trust.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tools. J Mol Biol 215:403–410

Bjarnason JB, Fox JW (1994) Hemorrhagic metalloproteinases from snake venoms. Pharmacol Ther 62:325–372

Blobel CP, Wolfsberg TG, Turck CW, Myles DG, Primakoff P, White JM (1992) A potential fusion peptide and integrin ligand domain in a protein active in sperm-egg fusion. Nature 356:248–252

Blundell TL (1994) Metalloproteinase superfamilies and drug design. Nat Struc Biol 1:73–75

Creigton TE, Darby NJ (1989) Functional evolutionary divergence of proteolytic enzymes and their inhibitors. Trends Biochem Sci 14: 319–324

Daltry AA, Wüster W, Thorpe RS (1996) Diet and snake venom evolution. Nature 379:537–540

Dayhoff MO, Schartz RM, Orcutt BC (1978) Atlas of protein sequence and structure, vol 5, supplement 3. NBRF, Washington, p 345

De Souza SJ, Brentani R (1993) Sequence homology between a bacterial metalloproteinase and eukaryotic matrix metalloproteinases. J Mol Evol 36:596–598

Emi M, Katagiri T, Harada Y, Saito H, Inazawa J, Ito I, Kasumi, F, Nakamura Y (1993) A novel metalloprotease/disintegrin-like gene at 17q213 is somatically rearranged in two primary breast cancers. Nat Genet 5:151–157

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Genetics Computer Group (1991) Program manual for the GCG package, Version 7. 575 Science Drive, Madison, WI 53711, USA

Gomis-Ruth FX, Kress LF, Bode W (1993) First structure of a venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases EMBO J 12:4151–4157

Gould RJ, Polokoff MA, Friedman PA, Huang TF, Holt JC, Cook JJ, Niewarowski S (1990) Disintegrins—a family of integrins inhibitory proteins from viper venoms. Proc Soc Exp Biol Med 5:168–171

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. Comput Appl Biosci 5:151–153

Higgins DG, Bleasby AJ, Fuchs R (1992) Cluster V: improved software for multiple sequence alignment. Comput Appl Biosci 2:189–191

Huang TF, Holt JC, Kirby EP, Niewiarowski S (1989) Trigramin—primary structure and its inhibition of von Willebrand factor binding to glycoprotein IIB/IIIA complex on human platelets. Biochemistry 28:661–666

Ikeo K, Takahashi K, Gobori T (1995) Different evolutionary history of kringle and protease domains in serine proteases: a typical example of domain evolution. J Mol Evol 40:331–336

Jongeneel CV, Bouvier J, Bairoch A (1989) A unique signature identifies a family of zinc-dependent metallopeptidases. FEBS Lett 242: 211–214

Moura-da-Silva AM, Paine MJI, Diniz MRV, Theakston RDG, Crampton JM (1995) The molecular cloning of a phospholipase A_2 from *Bothrops jararacussu* venom: evolution of venom Group II PLA_2's may imply gene duplication. J Mol Evol 41:174–179

Ohta T (1994) On hypervariability at the reactive center of proteolytic enzymes and their inhibitors. J Mol Evol 39:614–619

Paine MJI, Desmond HP, Theakston RDG, Crampton JM (1992) Purification, cloning and molecular characterization of a high molecular weight haemorrhagic metalloprotease, jararhagin, from *Bothrops jararaca* venom. J Biol Chem 267:22869–22876

Paine MJI, Moura da Silva AM, Theakston RDG, Crampton JM (1994) Cloning of metalloproteinase genes in carpet viper (*Echis pyramidum leakeyi*): further members of the metalloproteinase/disintegrin gene family. Eur J Biochem 224:483–488

Perry ACF, Jones R, Barker, PJ, Hall L (1992) A mammalian epididymal protein with remarkable sequence similarity to snake venom hemorrhagic peptides. Biochem J 286:671–675

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Weskamp G, Blobel CP (1994) A family of cellular proteins related to snake venom disintegrins. Proc Natl Acad Sci USA 91:2748–2751

Woessner JF (1991) Matrix metalloproteinases and their inhibitors in connective-tissue remodelling. FASEB J 5:2145–2154

Wolfsberg TG, Bazan JF, Blobel CP, Myles DG, Primakoff P, White JM (1993) The precursor region of a protein active in sperm-egg fusion contains a metalloproteinase and disintegrin domain: structural, functional, and evolutionary implications. Proc Natl Acad Sci USA 90:10783–10787

Yoshida S, Setoguchi M, Higuchi Y, Akizuki S, Yamamoto S (1990) Molecular cloning of a cDNA encoding MS2 antigen, a novel cell surface antigen strongly expressed in murine monocytic lineage. Int Immunol 2: 586–591