

# Analysis of transcripts encoding novel members of the mammalian metalloprotease-like, disintegrin-like, cysteine-rich (MDC) protein family and their expression in reproductive and non-reproductive monkey tissues

Anthony C. F. PERRY,\* Roy JONES† and Len HALL\*‡

\*Department of Biochemistry, University of Bristol, School of Medical Sciences, Bristol BS8 1TD, U.K. and †Department of Development and Signalling, BBSRC Babraham Institute, Babraham, Cambridge CB2 4AT, U.K.

A number of sequence-related, cysteine-rich proteins containing metalloprotease-like and disintegrin-like domains (the MDC protein family), at least one of which has been shown to play a role in egg recognition during fertilization, are abundantly expressed in the mammalian male reproductive tract. In this paper we report the cloning and sequence analysis of three closely related isoforms of a novel member of this family which are expressed not only in the testis, but also in the liver, albeit at a

lower level. Using a PCR-based approach we also demonstrate the presence of transcripts encoding additional, novel, disintegrin-containing proteins, in the liver and epididymis. We conclude that while some members of the MDC family are specific to the reproductive tract, suggesting functions peculiar to those tissues, others have a broader tissue distribution and may therefore play a more general role in integrin-mediated cell–cell recognition, adhesion or signalling.

## INTRODUCTION

In 1992 we reported the first full-length cDNA sequences of a member of what has now become known as the metalloprotease-like, disintegrin-like, cysteine-rich (MDC) family of mammalian proteins. These initial sequences encoded the rat and monkey homologues of EAP I (Epididymal Apical Protein I), an androgen-regulated 89 kDa transmembrane protein synthesized exclusively in the caput region of the epididymis with a predominant localization on the apical surface of its principal epithelial cells [1]. A remarkable feature was the finding that two regions of the protein EAP I showed significant homology to habu (*Trimeresurus flavoviridis*) venom haemorrhagic protein HR2A [2] (a zinc proteinase with basement membrane-hydrolysing activity [3]) and a Habu venom disintegrin [4] (a small, cysteine-rich, integrin ligand which binds to the glycoprotein GPIIb-IIIa receptor on platelets, thereby inhibiting fibrinogen binding and hence platelet aggregation). In addition, a number of snake-venom mosaic proteins or snake-venom precursors containing both protease and disintegrin domains within the same polypeptide chain have been characterized, including HR1B [5], jararhagin [6], atrolysin a [7], rhodostomin precursor [8], trigramin precursor [9] and pro-trimucin (J. H. Tsai, unpublished work), all of which have a number of domains similar to those found in EAP I.

More recently, full-length cDNA sequences of a number of EAP I-related MDC proteins have been described in the male reproductive tract. These include the  $\alpha$  and  $\beta$  subunits of the guinea-pig [10] and monkey [11] fertilin complex (formerly known as PH-30 antigen), monkey tMDC I [12], mouse cyritestin [13] (the mouse homologue of tMDC I) and monkey tMDC II [14]. Unlike EAP I, all of these additional MDCs appear to be expressed in the testis. The mature fertilin complex is a hetero-

dimeric sperm-surface transmembrane glycoprotein, comprising  $\alpha$  and  $\beta$  subunits that are synthesized in the testis as large precursors, with the  $\beta$  subunit undergoing final epididymal processing co-incidental with the acquisition by spermatozoa of full fertilizing ability [15]. It has been proposed that the fertilin complex plays an important role in recognition of the oolema plasma membrane by spermatozoa during fertilization and possibly also in subsequent sperm–oolema membrane fusion [16,17]. A number of preliminary observations point towards the importance of the disintegrin domain of the fertilin  $\beta$  subunit in its proposed egg recognition function: the egg is known to possess at least one type of integrin [18] and short peptides mapping to the disintegrin domain of fertilin  $\beta$  inhibit sperm–egg fusion, possibly by competition with the fertilin complex for its integrin receptor on the egg [18,19].

In contrast to our growing knowledge of the role of the fertilin complex in fertilization, virtually nothing is known about the functions of the other MDC family members in the male reproductive tract, although their abundance in the reproductive tract suggests they are involved in aspects of reproductive function. However, the recent finding of a related transcript in human brain [20] provided the first indication that mammalian MDC family members may not be restricted to reproductive tissues in mammalian species. This novel MDC cDNA sequence of unknown function was isolated from a human cerebellar cDNA library, but was also demonstrated, by reverse-transcriptase PCR (RT-PCR), to be expressed in a range of nervous tissues as well as several endocrine and reproductive tissues (mammary gland, adrenal gland, thyroid, pancreas, testis and ovary). However, unlike the other characterized members of the mammalian MDC family, this new MDC apparently lacked a C-terminal transmembrane domain, suggesting that it would be a secreted, rather than an integral, membrane protein. In addition,

Abbreviations used: MDC, metalloprotease-like, disintegrin-like, cysteine-rich; EAP I, epididymal apical protein I; RT-PCR, reverse-transcriptase PCR.

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported in this paper appear in the EMBL Nucleotide Sequence Database under the accession numbers: X87205, X87206 and X87207 (monkey tMDC IVa, IVb and IVc cDNAs respectively).

database searches reveal a further, unpublished member of the MDC family (EMBL accession number Z21961), identified in bovine brain and reported in the database documentation to be an active metalloproteinase, and the presence of disintegrin-containing transcripts in mouse lung and C2 muscle cells have also recently been demonstrated [21].

We now report the cloning and complete sequence determination of a series of very closely related, but distinct, novel MDC transcripts expressed in the testis and, at lower levels, in the liver. This suggests a function which is not confined to reproductive tissues. We have extended this finding using a redundant oligonucleotide-directed RT-PCR approach, and report the partial sequences of a number of additional novel disintegrin-containing MDC transcripts expressed in the liver and epididymis. Taken together with earlier findings, it is now clear that the mammalian MDCs form a very large, growing family. While some members are tissue-specific (e.g. EAP I and fertilin  $\beta$ ), suggesting functions peculiar to those tissues, others have a broader tissue distribution and may therefore play a more general role in integrin-mediated cell-cell or cell-matrix interactions or signalling.

## MATERIALS AND METHODS

### Materials

All chemicals and reagents were of AnalaR grade or higher. *Taq* DNA polymerase was supplied by Boehringer Mannheim, Lewes, E. Sussex, U.K. Other DNA-modifying enzymes, oligo(dT)<sub>12-18</sub> and dephosphorylated *Sma*I-cut pUC18 were from Pharmacia, Milton Keynes, U.K. Oligo(dT)-cellulose was obtained from Collaborative Biomedical Products. Nitrocellulose membranes for library screening and Hybond-N for Northern blot analysis were respectively from Schleicher and Schüll and Amersham International, Amersham, Bucks., U.K. [ $\alpha$ -<sup>32</sup>P]dCTP (3000 Ci·mmol<sup>-1</sup>) and a random 14-mer DNA labelling kit were provided by DuPont-NEN, Stevenage, U.K.

### Isolation of RNA and Northern blot analysis

Fresh tissues (testis, liver, kidney and epididymis) were obtained *post mortem* from a sexually mature monkey (*Macaca fascicularis*), flash frozen in liquid N<sub>2</sub> and stored at -70 °C until required for RNA extraction. Total RNA was extracted from fresh samples (approx. 2 g) using an SDS/proteinase K-based method and, where appropriate, enriched for polyadenylated RNA by oligo(dT)-cellulose affinity chromatography. Northern blot analysis of RNA was performed as previously described [11]. Reprobing of filters with an  $\alpha$ -actin cDNA confirmed the integrity of RNA samples and enabled standardization of the lane loadings.

### RT-PCR amplification of sequences

*M. fascicularis* testis or epididymis polyadenylated RNA (1  $\mu$ g), or liver or kidney total RNA (2  $\mu$ g) were used as templates for AMV reverse transcriptase-directed cDNA synthesis using oligo(dT)<sub>12-18</sub> as a primer. Amplification of disintegrin domains employed 6.25% of the cDNA from each synthesis in a PCR reaction containing the following degenerate oligonucleotide primers: 5'-RSD GAR SAG TGT GAY TGT GG-3' (corresponding to conserved amino acid residues T/GEECDCC near the N-terminus of the disintegrin domain) and 5'-GCA AWW TTC WGG RAR RTC RCA-3' (corresponding to conserved amino acid residues CDF/LPEF/YC near the C-terminus of the disintegrin domain). PCR parameters were: 94 °C for

1.5 min; cool to 50 °C over 1.5 min; hold at 50 °C for 2 min; 72 °C for 2 min; 30 cycles. For tMDC IV-isoform-specific reactions, an annealing temperature of 58 °C and an elongation time of 2.5 min at 72 °C were utilized, other parameters remaining the same as above. Primer pairs used to amplify specific tMDC IV isoforms were: 5'-TGCTGTACAAACTGCAGCTTCT-3' and 5'-GAATGAAACATGTTCCCTGCAGC-3' (corresponding to tMDC IVa and b only) and 5'-CTCCAACAGCCTGCACATTTGC-3' and 5'-AGCCTTATACACATTGGTACTC-3' (corresponding to tMDC IVc only). Where appropriate, products were cloned into pUC18 and their sequences determined using a DuPont Genesis 2000 automated sequencer.

### Identification of monkey tMDC IV cDNA clones

The construction of an *M. fascicularis* cDNA library has been described previously [11]. Approx.  $1.4 \times 10^5$  independent cDNA clones were transferred to four nitrocellulose filters (22 cm  $\times$  22 cm) and screened at high stringency [11] for those whose cDNA inserts hybridized to a monkey tMDC IV-cloned PCR fragment that had been gel-purified and labelled to a high specific activity with <sup>32</sup>P. Hybridization was verified by secondary screening and clones purified following tertiary screening under identical hybridization conditions. Plasmid DNAs from positive clones identified in this way were then isolated and the cDNA inserts of several, including the largest, subjected to sequence analysis using a DuPont Genesis 2000 automated sequencer. Sequences were determined on both strands using a custom primer walking strategy and were compiled and aligned using the LASERGENE suite of programs (DNASTAR, West Ealing, London, U.K.).

## RESULTS AND DISCUSSION

### Clones for monkey tMDC IV reveal three closely related isoforms

In an attempt to identify novel members of the MDC family of transcripts we designed redundant PCR primers to conserved regions which flank the disintegrin-like domains of EAP I and the  $\alpha$  and  $\beta$  subunits of fertilin. These primers were then used to carry out PCR reactions with cDNA derived from monkey testis polyadenylated RNA (see the Materials and methods section). Subsequent cloning and sequence determination of a number of these PCR products led to the identification of several novel disintegrin-containing testicular transcripts. Full-length cDNA clones for three of these transcripts {tMDC I [12], tMDC II [14] and tMDC III (A. C. F. Perry, R. Jones and L. Hall, unpublished work)} have been isolated and characterized. Here we describe the isolation of cDNA clones for a fourth class of transcript which we shall refer to as tMDC IV.

Approx.  $1.4 \times 10^5$  independent *M. fascicularis* testis cDNA clones were screened under conditions of high stringency for their ability to hybridize to the cloned monkey PCR product corresponding to the tMDC IV disintegrin domain, and approx. 23 strongly positive clones were identified and purified during subsequent rounds of screening. Following determination of the insert size, the sequences of the entire cDNA inserts for several of the larger recombinant clones were determined. Preliminary alignment of these sequence data indicated that the sequenced clones fell into three distinct, but very closely related, classes which we have called tMDC IVa, tMDC IVb and tMDC IVc. The differences between these classes do not represent polymorphisms between individual animals since the cDNA library used for their isolation was derived from a single testis.

The complete nucleotide and deduced amino acid sequences of monkey tMDC IVa (derived from five completely overlapping,

```

1   TTTCTTTTCCCATCTCTTTGGATGACTATCCCTGGGCTGGGGCATGAGGTCAGCAGAGGCCAGTCACCCCTTAGGGCCCCGCTTGTGCTGGGGCTTGGGCACCTCTGGCTCCA
      1           M R S A E A P V T L R A P V L L L G L W A L L A P           10           20
120 GTCTGGTGTCTCAAGGCCGTCCTTGTGGCACTACGCATCCTCTGAGGTGGTATGCCAGGAAGGAGCACCACATGGCAAAGGCGTTTCAGTTTCCGGCTGGCTGTCTACAGCGTG
      V W C S Q G R P L W H Y A S S E V V I P R K E T H H G K G V Q F P G W L S Y S L           30           40           50           60
240 CGTTTTGGGGTCAAAGACACGTCGTTACATGCGGAGAAAACACCTTTTGGCCAGACATCTGCTGATGACAACCTCAGGATGACCAAGGGGCTTGCAGGTGGATGACCCCTACATC
      R F G G Q R H V V H M R R K H L L W P R H L L M T T Q D D Q G A L Q V D D P Y I           70           80           90           100
360 CCTCCAGACTGCTACTACCTGGCTACCTGGAGGAGTGCCTCTGTCCATGGTCACCGTCGACACGTGCTATGGGGCCCTCAGAGGCATCATGAAGTGGAGACCTTGCCTACGAAATC
      P P D C Y Y L G Y L E E V P L S M V T V D T C Y G G L R G I M K L D D L A Y E I           110          120          130          140
480 AAACCCCTCCAGATTCCCGCAGGTTTGAACATGTTGTTTCTCAGATAGTGGCCGAGCCCAACGCAACGGGGCCACATTTAGAGATGGTGACAATGAGGAGACAGACCCCTGTCTCT
      K P L Q D S R R F E H V V S Q I V A E P N A T G P T F R D G D N E E T D P L F S           150          160          170          180
600 GAAGCAATGACAGCATGAATCCAGGATATCTACTTTGCTGATAGTTCTCATAGAGGCAATATAAAGGCCACGTTCAATGTTCCAACACATATTATCGTATATATGGCAATATTACA
      E A N D S M N P R I S T L L Y S S H R G N I K G H V Q C S N T Y Y R I Y G N I T           190          200          210          220
720 ACTTGTTACATAGCGGTGGTCCAGATGTTGAGTGCATTGACAGCATTGTTCAAATATTGATCTGCGGCTACTATATTTATCTTTGACCATATATAATGTTGTCGACCCAGCTCCTGTG
      T C Y A S R C C R S D C R L T P G S I C H L G D C C T N C S F S A Q G T L C R P I           230          240          250          260
840 AATCAATATGGAGTAAATAGTGGCATGTTACTATTTTAAAGAACCTTTTTGATACTTTTCGTGTTTCATCCACTACTTATTAAGAAGGCCACATGAATCTAACTATGAA
      N Q Y G A N S A M F T Y F K R T F F D T F R V H S S T L L I K E A P H E S N Y E           270          280          290          300
960 CCACAAAGTATAGCTCTGTACACATTTAGGCTATTACACTTGGTACTCTAGGCAGACATTATTTATGGTAGCCGTCATAACACCCAGACACTGATGAAGTATGGGTGGGAG
      P Q R Y S F C T H L G L L H I G T L G R H Y L L V A V I T T Q T L M R S M G V E           310          320          330          340
1080 TATGATGACAACCTACTGCACATGTCAGAGAAGGGCCTTCTGCAATTTGACAGGATCTCTGGGATAACAGATGCGTTCAGTAACTGTTCTTATGGACATACACAAAATGTTTGTAAAT
      Y D D N Y C T C Q R R A F C I M Q R Y P G I T D A F S N C S Y G H T Q N C F V N           350          360          370          380
1200 TCATCCAGTGCCTTTTTGAAGCACTCGCTCCAGTGCATAACGAAAGCATAACAATGGTTCGCTGTGGAACCTAATCGTGAAGGGAGGAGGAATGTGACTGTGGCTCCTTCAAGCAG
      S S Q C V F E A L A P V H N E S I T M V R C G N L I V E G R E E C D C G S F K Q           390          400          410          420
1320 TGTTATGCCAGTCTGTTGCTGCCAAGTACTGTCGCCTCACACCGGGAGCATCTGTCTCTAGGAGACTGCTGTACAACCTGCAGCTTCTCCGCACAAGGGACTCTGTCAGACCTATC
      C Y A S R C C R S D C R L T P G S I C H L G D C C T N C S F S A Q G T L C R P I           430          440          450          460
1440 CAAAATATATGTGACCTTCCAGAGTACTGTCATGGGACCACCTTGACATGCCCCAGACCTTTATTTGCAAGATGGAACCCCGTGCATGAAGAAGGCTACTGCTATCATGGAACTGC
      Q N I C D L P E Y C H G T T L T C P P D L Y L Q D G T P C T E E G Y C Y H G N C           470          480          490          500
1560 ACTGACCGCAATGTCTCTGCAAGGCGATCTTGGTGTGAGTCTAAGGATGCTCCTGGGACTGCTATGACATCAATCTTGAAGTACCAGTTCGGACATGTACTCGAGAACAATCA
      T D R N V L C K A I F G V S A K D A P G D C Y D I N L E S H R F G H C T R E Q S           510          520          530          540
1680 GCTCTCAGTACCAGGCTGTGTAGGAATAGATAAGTTTGTGGAAGACTGCAGTGTACCAATGTGACCCATCTTCCCGGCTGCAGGAACATGTTTCATTCCTCACTCAATAAGAAGA
      A L S Y Q A C V G I D K F C G R L Q C T N V T H L P R L Q E H V S F H H S I R R           550          560          570          580
1800 GGGTTTCAGTGTTTTGGATGGATGAACACCGTGAACAGACACAACCTGATGTTGGCGTGTGATAGAGGCCTCCTGTTCTCGTGAATGGCCTGATAACACCCAGTGAATGATG
      G F Q C F G L D E H R A T D T T D V G R V I D G T P C S R G M A C N N T Q C N V           590          600          610          620
1920 GCTATCACTTCACTGAACTACGACTGTTACCCTCAGAAGTGCAGTTACAAGGGGTCTGCAACACAGAAGGAACCTGCCATTGCCATATAGGCTGGATCCTCCACGGTGCCTAAGAAGA
      A I T S L N Y D C Y P Q K C S Y K G V C N N R R N C H C H I G W D P P R C L R R           630          640          650          660
2040 GGTATTGGTGGGAGTGTGACAGCGGGCCACCTCCAAGAAGAACCTGTCGGTCAACAAGCCATCAGGCAGTGTGATCTGAGAGTGGTCTTTGGTGTATTTACACCTTCAATAAT
      G I G G S V D S G P P P R R T R S V K Q S H Q A V L Y L R V V F G R I Y T F I I           670          680          690          700
2160 GCACTGCTCTTTGGGATGGCCACAATGTGCGAAGTATCAGGACCACCGCGTTAAGGAAGTAAACGCTACTGACCCAGAATAAGACAAAATTCAGCCTGTAAGATACAGAGAATATAAC
      A L L F G M A T N V R T I R T T A V K E V T A T D P E .           710          720          730          732
2280 AGCAAAATCTATGGAACAGGATCAGGGGAAAGGATGGCAAAGCTCAAGTCCACATTTCTGAAATGATGCTCACTCTGAAATAAATCTTCAAAAACACAAAAA 2387
    
```

**Figure 1 Nucleotide and deduced amino acid sequences of monkey tMDC IVa**

non-sibling clones) are shown in Figure 1. This transcript encodes a hypothetical translation product of 732 amino acid residues (with a predicted molecular mass of 82694 Da) and a domain organization (N-terminal domain, metalloprotease-like domain, disintegrin-like domain, cysteine-rich domain and transmembrane domain) very similar to those of other MDC proteins (results not shown). However, the region of tMDC IV predicted to correspond to the zinc-binding consensus and protease catalytic centre for snake-venom HR2A (Figure 2) lacks the zinc-binding histidine and catalytic glutamate residues, features shared by tMDCs I–III and fertilin  $\beta$ . These members of the MDC family are therefore unlikely to bind zinc in this region, or to exhibit proteolytic activity.

The nucleotide sequences of tMDC IVb (derived from three completely overlapping, non-sibling clones) and tMDC IVc (derived from a single clone) respectively exhibit about 90% and 93% nucleotide sequence identity to tMDC IVa (results not

shown). The deduced amino acid sequences of tMDC IVa–c are aligned in Figure 3. It can be seen that many of the nucleotide differences between the three isoforms result in amino acid substitutions and that tMDC IVc is a C-terminally truncated isoform which terminates within the protease-like domain. This truncation apparently arises from a deletion of the tetranucleotide TTAT corresponding to nucleotide position 1026–1029 in tMDC IVa (Figure 1), resulting in an in-frame termination codon 3–6 nucleotides downstream of this deletion. Since the sequence of tMDC IVc was derived from a single cDNA clone, the apparent tetranucleotide deletion was checked and confirmed following sequence analysis of RT-PCR products obtained using tMDC IVc-specific primers. It is therefore clear that the deletion in tMDC IVc is not a cloning artefact and that these transcripts would potentially encode a considerably truncated protein product. However, even if the tetranucleotide TTAT is inserted into the tMDC IVc sequence, the resulting deduced translation

<i>Active protease consensus</i>	<i>M - H E - G H - L G - - H D</i>
Snake venom HR2A	M T H E I G H N L G M E H D
Monkey fertilin $\alpha$	M V H E L G H N L G I Q H D
Human MADM	F A H E V G H N F G S P H D
Monkey EAP I	M A H Q L G H N L G M Q H D
Monkey tMDC I	M A Q L L G I N L G L T Y D
Monkey tMDC II	I V Q L L G L N L G L T Y D
Monkey tMDC III	I A Q L L G L N V G L T Y D
Monkey tMDC IV	T T Q T L M R S M G V E Y D
Human brain MDC	L A Q T L G Q N L G M M W N
Monkey fertilin $\beta$	L A Q L L S L S M G I P Y D

**Figure 2** Alignment of the region corresponding to the protease catalytic centre of snake-venom haemorrhagic protein HR2A with the metalloprotease-like domains of MDC proteins

The region corresponding to the metalloprotease catalytic centre of HR2A [2] is shown aligned with the corresponding region in the metalloprotease-like domains of monkey EAP I [1], fertilin  $\alpha$  and  $\beta$  subunits [11], tMDC I [12], tMDC II [14], tMDC III (A. C. F. Perry, R. Jones and L. Hall, unpublished work), tMDC IV (this paper), human brain MDC [20] and human MADM (P. Glynn, unpublished work; EMBL accession number Z48579). The extended consensus sequence corresponding to the active-site helix of snake-venom proteinase adamalysin II and thermolysin [24] is also shown. Putative zinc-binding histidine residues and the active-site glutamate residue are underlined.

product would still be C-terminally truncated at a position midway between the disintegrin domain and the transmembrane domain of tMDC IVa. Interestingly, this would correspond within a few residues to the C-terminal end of human brain MDC [20], which lacks the transmembrane-containing C-terminal extension found in all other characterized mammalian MDCs. Given the significantly truncated nature of the tMDC IVc deduced amino acid sequence, together with the presence of additional changes distal to the tetranucleotide deletion site (which would also result in premature termination), it seems likely that tMDC IVc does not represent a functional transcript but rather a transcribed, processed pseudogene.

**Tissue distribution of tMDC IV transcripts**

Total RNA from *M. fascicularis* testis, epididymis, kidney and liver was analysed for tMDC IV transcripts by Northern blotting using a probe which, although derived from cloned tMDC IVa cDNA, should detect tMDC IVa, b and c transcripts due to the extremely high level of sequence identity between these three forms. Although a single, readily detectable transcript (of a size consistent with the nearly full-length cloned tMDC IVa cDNA) was observed in the testis, no hybridizing transcripts were detected in the epididymis, kidney or liver, suggesting a steady-state level in these latter tissues at least 30-fold lower than that in the testis (results not shown).

In view of the limited sensitivity of Northern blots when using total RNA, attempts were made to identify tMDC IV transcripts in the liver by PCR, using liver cDNA as template. Primer pairs intended to be specific to tMDC IVa/b or tMDC IVc (i.e. corresponding to regions poorly conserved with respect to all other sequenced MDCs) were used in PCR reactions with cDNA derived from monkey liver total RNA. PCR products of the expected size were obtained with both primer pairs; collectively these data indicate that the cryptic tMDC IVc transcripts, as well as tMDC IVa and/or b, co-exist in both testis and liver, and that the steady-state level of total tMDC IV transcripts is greatly reduced in the liver compared with the testis.

**Identification of disintegrin-containing transcripts in the liver, kidney and epididymis**

The finding of transcripts encoding tMDC IV isoforms in the liver prompted us to look for additional novel MDC transcripts in non-reproductive tissues. At the same time we wished to see whether MDC transcripts additional to EAP I are expressed in the epididymis. To this end, we used the same degenerate PCR primers which yielded a part of tMDC IV, to amplify disintegrin domains of sequences present in monkey liver or kidney total RNA or epididymis polyadenylated RNA. Subsequent cloning and sequence determination of these PCR products led to the

<i>Monkey tMDC IVa</i>	MRSAEAPVTLRAPVLLGLWALLAPVWCQGRPLWHYASSEVVI	PKETHHGKGVQFPGWLSYSLRFGGQRHVHMRKHLWPRHLLMTTQDDQGLQVDDPYIPPD
<i>Monkey tMDC IVb</i>	.....	.....A.....
<i>Monkey tMDC IVc</i>	.....	.....T.....
		Protease-like domain
	CYYLGLYEEVPLSMVTVDTCYGGRLGIMKLLDLDLAEIKPLQDSRRFEHVVSQIVAEPNATGPTFRDGDNEETDPLFSEANDSMNPRISTLLYSHRGNIKGHVQCSNTYYRIYGNITTCYIAVVOMFS	.....L-PS.....SKE.....
	.....	.....L-PS.....S.....D.....SKE.....
		Protease-like domain
	VIDSIVQNIIDLRYIYLLTIYVNRDPAPVNYGANSAMFTYFKRTFFDTRFVHSSSTLLIKEAPHESNYEPQRYSFCTHLLGHLHIGTLGRHYLLVAVITTTQTLMRSMGVEYDDNYCTCORRAFIMORY	.....I-R.....SR.....D-RYR-PILN-R.....D.....R.....S-S-I.....D-L-V-HI-L-A-YK-C-TS.....
	.....T-W.....NH.....Q-RVR-AMFT-K.....E.....R.....S-S-I.....D-Y-HI-L-A-YK-R-TS.....	
		Disintegrin-like domain
	PGITDAFNSCSYGHTONCFVNSQCVFEALAPVHNESITMRCGNLIVEGEECDGCSFKQCYASRCCRSOCDRLTPGSIHLGDCCTNCSFSAQGTLCRPIQNICDLPEYCHGTTTLTCTPPDLYLQDGT	.....S.....STGR-I.....T-Y-TM-T-L.....K.....NH-Q-HF.....AO.....Y.....F.....
	PGM.....S.....I.....	.....SEM-G-L-NSAR-I-A-Y-TM-T-L.....K.....SH-Q-HF.....PL.....H.....F.....
		Protease-like domain
	PCTEEGYCYHGNCITDRNVLCKAIFGVSAKDAPGDCYDINLESHRFHGCTREQSALSQACVGDKFCRGLQCTNVTHLPRLQEHVSFHHSIRRGFCQFGLDEHRATDITDVGVRIDGTPCSRGMACNN	.....E-E-N-RART-IA-E-CAL.....V-LQ-HVSFHHSVRGGF.....H-AD-IF.....
	.....E-E-N-QEQN-IR-Q-GVG.....M-ME-RQMCILLEKYR.	
		Transmembrane domain
	TQCVAITSLNYDCYPQKCSYKGCNRRNCHCHIGWDPRLRRIGGSVSDGPPRRRTRSVKSHOAVLYLRVVFGRITYTFIALLFGMATNVRTIRTTAVKEVTATDPE.	.....L-A.....Q-S.....V.....A-IL-T-EK-V.....
	S---A-T---G---H-E---HR.....	

**Figure 3** Alignment of monkey tMDC IV isoforms

Dashes represent sequence identity to tMDC IVa. No clones containing the entire coding region of tMDC IVb were obtained; the largest tMDC IVb clone starts at a position corresponding to amino acid residue 21 of tMDC IVa. Protease-like, disintegrin-like and transmembrane domains are indicated.

```

Consensus          -E-CDCG-----CC---C-----C-G-CC---C-----CR-----D---C-G---C-----

Habu venom disintegrin
GEECDGSPS-----NPCCDAATCKLRPQAQCADGLCCD--QCRFKKKTGICRIA---RGDFPDDRCTGLSNDPCRWNLD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Trigramin precursor
GEDCDGSPA-----NPCCDAATCKLIPGAQCGEGLCCD--QCSFIEEGTVCRIA---RGDLDDYCNGRSAGCPRNPFHA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Habu venom HR1B
GEECDGSPENCQ---YQCCDAASCKLHSWVKCESGECCD--QCRFRTAGTECRAAES-ECDIPE-SCTGQSADCPDRFHRN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey eMDC clones 27 & 32
geecdsgTSEECT---NICCDAKTCKIKAGFOCTSGECCD--KQCFKKAGMVC RPAKD-Ecdlpe-ic
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey IMDC clones 11 & 12
peecdsgTSEECT---NICCDAKTCKIKAGFOCTSGECCD--KQCFKKAGMVC RPAKD-Ecdlpe-fc
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey EAP I
GEECDGPPQECT---NPCCDAHTCVLKP GFTCAEGECCD--SCQIKKAGSICRPAED-ECDFPE-MCTGHSPACPKDQFRVN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey eMDC clone 6
peqcdsgPLKHC A-KD-PCCLS-NCTLDGSTCAFGLCCK--DCKFLPSGEVCRKEAN-Ecdfpe-nc
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey kMDC clones 3 & 12
teecdsgS--AC-HLD-PCCDP-TCTLKEHAEC SHGLCCL--DCTFRRKGFLCRPTQD-Ecdlpe-yc
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey fertilin  $\alpha$ 
TEEDCGS--AC-HLD-PCCDP-TCTLKEHAEC SHGLCCL--DCTFRRKGFLCRPTQD-ECDLPE-YCDGSSAECPADSYKQD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey fertilin  $\beta$ 
GEECDGTQNCFL LGAKCDDTATCRFKAGSNAEGPCCE--NCLFMSQERVCRPSFD-ECDLPE-YCNGTSASC PENHFQIT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey tMDC II
GEECDGKNDKNTH--KLCCDALTCRLKDNAQCGSGDCCS--KDCFKPANTICRKSVDVECFTE-FCNGSYPYCLLD TYVRD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey tMDC I
TEQDCGYRKA CTF--KKCCNPADCTLIGFAECGSGPCNNKTCTIYARGHVCRKSID-MCDFPE-YCNGTSEFCVDPVKAAD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey tMDC III
NEECDGKNGKECQF--KKCCDYNTCKLKG SVKCGSGPCCTSK-CELSIVGTPCRKSVDECFDTE-YCNGTSSDCVPD TYALN
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey tMDC IVa
REECDGFSFKQCYA--SRCCRS D-CRLTPGSICHLGDCCT--NCSFSAQGLCRPIQN-ICDLPE-YCHGTTLTCPDLYLQD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey IMDC clone 15
reecdsgSFKQCYA--SRCCRS D-CRLTPGSICHLGDCCT--NCSFSAQGLCRPIQN-Icdlpe-fc
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey tMDC IVb
KEECDGFSFKQCYA--NHCCQSD-CHFTPGSICHLGDCCT--NCSFSAQGLCRPIQN-ICDLPE-YCYGTTLTCPD FYLQD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey tMDC IVc
KEECDGFSFKQCYA--SHCCQSD-CHFTPGSICHLGDCCT--NCSFSP LGTLCRPIQN-ICDLPE-YCHGTTLTCPD FYLQD
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey IMDC clones 8 & 5
reecdsgSFKQCYA--SHCCQSD-CHFTPGSICHLGDCCT--NCSFSP LGTLCRPIQN-Icdlpe-yc
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey IMDC clone 10
reecdsgSLKDCAN--DRCCETS-CILSLG SVCNTGLCCR--ECKYAAPGVVCRDLGG-Icdfpe-ic
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey IMDC clones 13 & 6
geecdsgSL--CQR--HPCCDEN-CILKVAECSDGPCC H--NCKFQPKGYSCRPSSD-Scdfpe-ic
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Monkey IMDC clone 3
teqcdsgSVQQCEQ--DACLLN-CTLP GAA-CAFGLCCK--DCKFMPSGELCRQEVN-Ecdfpe-ic
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Human bMDC
GEECDGSVQECSRAGNCKK--CTLTHDAMCSDGLCCR--RKYEPRGVSCREAVN-EC DIAE-TCTGDSSQCPPLHLKL

```

**Figure 4** Alignment of the disintegrin-like domain of MDC proteins

The disintegrin domains of monkey EAP I [1], fertilin  $\alpha$  and  $\beta$  subunits [11], tMDC I [12], tMDC II [14], tMDC III (A. C. F. Perry, R. Jones and L. Hall, unpublished work), tMDC IV (this paper), human brain MDC [20], and a series of cloned, liver-, kidney- and epididymis-derived, PCR products (IMDCs, kMDCs and eMDCs; see the Results section) are aligned, together with habu venom disintegrin [4] and the disintegrin domains of the trigramin precursor [9] and Habu venom HR1B [5]. Residues in lower-case type represent regions corresponding to the degenerate PCR primers and are therefore unconfirmed.

identification of six different disintegrin-containing liver transcripts, of which one corresponded to tMDC IVa (IMDC clone 15), one to tMDC IVc (IMDC clones 5 and 8), with the remaining four being novel (IMDC clone 3; IMDC clones 6 and 13; IMDC clone 10; IMDC clones 11 and 12). Similar analyses of the cloned epididymal PCR products identified two different disintegrin-containing epididymal transcripts, the first (eMDC clones 27 and 32) corresponding to one of the four novel liver transcripts (IMDC clones 11 and 12), the other (eMDC clone 6) encoding a novel transcript. Surprisingly, analysis of two independent, cloned kidney PCR products (kMDC clones 3 and 12) revealed them to contain a sequence identical to that found in the monkey fertilin  $\alpha$  subunit [11], indicating that transcripts encoding this sperm-surface antigen are not exclusively confined to the testis. These PCR-derived sequences are aligned in Figure 4 together with the disintegrin domains of all characterized monkey MDCs, human brain MDC and representative snake-venom components. The complete conservation of the eight cysteine residues present in snake-venom disintegrins and known to be involved in disulphide bonds in the native venom proteins strongly imply that all of these disintegrin domains share a similar tertiary

structure. These data indicate that the repertoire of MDCs expressed in the liver, epididymis and testis in part overlap, suggesting that the pattern of cognate MDC integrin receptor expression may also be conserved between these tissues. Interestingly, integrin  $\beta 1$  subunits are not only expressed in hepatocarcinoma cells [22], but are apparently present in Sertoli cells, where they do not co-localize with their potential ligand, laminin [23]. It is thus conceivable that the repertoire of MDC ligands peculiar to the liver and testis each mediates a characteristic set of interactions between MDC-presenting cells and those expressing integrin  $\beta 1$  subunits.

This work was supported in part by the Medical Research Council (U.K.) and the Agricultural and Food Research Council (U.K.). Automated DNA sequencing facilities were provided by the BBSRC-funded University of Bristol Molecular Recognition Centre.

## REFERENCES

- Perry, A. C. F., Jones, R., Barker, P. J. and Hall, L. (1992) *Biochem. J.* **286**, 671–675

- 2 Miyata, T., Takeya, H., Ozeki, Y., Arakawa, M., Tokunaga, S. and Omori-Satoh, T. (1989) *J. Biochem. (Tokyo)* **105**, 847–853
- 3 Ohsaka, A., Just, M. and Habermann, E. (1973) *Biochim. Biophys. Acta* **323**, 415–428
- 4 Musial, J., Niewiarowski, S., Rucinski, B., Stewart, G. J., Cook, J. J., Williams, J. A. and Edmunds, L. H. (1990) *Circulation* **82**, 261–273
- 5 Takeya, H., Oda, K., Miyata, T., Omori-Satoh, T. and Iwanaga, S. (1990) *J. Biol. Chem.* **265**, 16068–16073
- 6 Paine, M. J., Desmond, H. P., Theakston, R. D. and Crampton, J. M. (1992) *J. Biol. Chem.* **267**, 22869–22876
- 7 Hite, L. A., Jia, L.-G., Bjarnason, J. B. and Fox, J. W. (1994) *Arch. Biochem. Biophys.* **308**, 182–191
- 8 Au, L. C., Huang, Y. B., Huang, T. F., The, G. W., Lin, H. H. and Choo, K. B. (1991) *Biochem. Biophys. Res. Commun.* **181**, 585–593
- 9 Neeper, M. P. and Jacobson, M. A. (1990) *Nucleic Acids Res.* **18**, 4255
- 10 Wolfsberg, T. G., Bazan, J. F., Blobel, C. P., Myles, D. G., Primakoff, P. and White, J. M. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10783–10787
- 11 Perry, A. C. F., Gichuhi, P. M., Jones, R. and Hall, L. (1995) *Biochem. J.* **307**, 843–850
- 12 Barker, H. L., Perry, A. C. F., Jones, R. and Hall, L. (1994) *Biochim. Biophys. Acta* **1218**, 429–431
- 13 Heinlein, U. O. A., Wallat, S., Senftleben, A. and Lemaire, L. (1994) *Dev. Growth Differ.* **36**, 49–58
- 14 Perry, A. C. F., Barker, H. L., Jones, R. and Hall, L. (1994) *Biochim. Biophys. Acta* **1207**, 134–137
- 15 Blobel, C. P., Myles, D. G., Primakoff, P. and White, J. M. (1990) *J. Cell Biol.* **111**, 69–78
- 16 Primakoff, P., Hyatt, H. and Myles, D. G. (1985) *J. Cell Biol.* **101**, 2239–2244
- 17 Blobel, C. P., Wolfsberg, T. G., Turck, C. W., Myles, D. G., Primakoff, P. and White, J. M. (1992) *Nature (London)* **356**, 248–252
- 18 Fusi, F. M., Vignalli, M., Busacca, M. and Bronson, R. A. (1992) *Mol. Reprod. Dev.* **31**, 215–222
- 19 Fusi, F. M., Vignalli, M., Gailit, J. and Bronson, R. A. (1993) *Mol. Reprod. Dev.* **36**, 212–219
- 20 Emi, M., Katagiri, T., Harada, Y., Saito, H., Inazawa, J., Ito, I., Kasmuni, F. and Nakamura, Y. (1993) *Nature Genetics* **5**, 151–157
- 21 Weskamp, G. and Blobel, C. P. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2748–2751
- 22 Patriarca, C., Roncalli, M., Gambacorta, M., Cominotti, M., Coggi, G. and Viale, G. (1993) *J. Pathol.* **171**, 5–11
- 23 Palombi, F., Salanova, M., Tarone, G., Farini, D. and Stefanini, M. (1992) *Biol. Reprod.* **47**, 1173–1182
- 24 Gomis-Rüth, F.-X., Kress, L. F. and Bode, W. (1993) *EMBO J.* **12**, 4151–4157