The gene for the human tMDC I sperm surface protein is non-functional: implications for its proposed role in mammalian sperm–egg recognition

Jan FRAYNE and Len HALL[1]

Department of Biochemistry, University of Bristol, School of Medical Sciences, Bristol BS8 1TD, U.K.

The sperm surface antigen tMDC I (also known as cyritestin) has been proposed to play a role in sperm–egg binding in the mouse via an interaction between its disintegrin-like domain and an integrin receptor on the oolemma plasma membrane. Here we report the cloning and sequence of human tMDC I transcripts and show that they are non-functional, owing to the presence of a variety of deletions, insertions and in-frame termination codons. The absence of a tMDC I protein is further supported by the lack of immunoreactivity on Western blots of human testis and sperm extracts probed with macaque (*Macaca fascicularis*) and human anti-tMDC I antisera.

INTRODUCTION

The MDC proteins are a novel family of multidomain integral membrane proteins all of which possess a **M**etalloproteinase-like domain, a **D**isintegrin-like domain and a **C**ysteine-rich domain (alternatively referred to as the 'ADAM' family). Several MDC proteins have been postulated to play a role in cell–cell adhesion via binding of the disintegrin-like domain to an integrin receptor.

Fertilin α and β were among the first MDC proteins to be characterized [1]. Both are expressed by spermatogenic cells and are synthesised as large precursor proteins in the testis [2–4]. In the guinea-pig they are reported to exist as a sperm surface heterodimeric complex [2], with fertilin β implicated in sperm–egg binding and fertilin α originally postulated to mediate membrane fusion. However, the recent finding that the human fertilin α gene is non-functional [5] casts some doubt on an essential role in membrane fusion, at least in some species.

More recently, considerable attention has been focused on the role of an additional member of the MDC protein family, tMDC I (also known as cyritestin), in sperm–egg membrane binding and fertilization [6,7]. In rodents, tMDC I is exclusively expressed by spermatogenic cells [8–10], with transcripts initially expressed by spermatocytes [9,10], but stored for about 4 days before translation [9], indicating a degree of post-transcriptional regulation. Following synthesis, tMDC I is transported to the forming acrosomal vesicle, via the Golgi apparatus, to become part of the acrosomal membrane [11], where it does not become exposed to the sperm surface until after the acrosome reaction [9], consistent with a role in oolemma plasma-membrane binding.

Like fertilin α and β, tMDC I undergoes extensive post-translational modification, with endoproteolytic removal of the N-terminal portion, including the pro- and metalloproteinase-like domains, leading to a type I integral membrane protein retaining the extracellular disintegrin domain, on the surface of mature, fertilization-competent sperm in the cauda epididymidis [6,9,11]. The disintegrin domain of tMDC I, which is highly conserved among the MDC proteins, contains a typical XCD integrin-binding motif, implying that tMDC I may interact with an integrin on the egg plasma membrane. Although recent studies have indicated that fertilin β interacts with a β1 integrin on the egg plasma membrane [12], whether tMDC I binds to the same or a different integrin has yet to be addressed.

Evidence for a physiological role for tMDC I in sperm–egg binding stems from *in vitro* studies using peptide mimics of the disintegrin loop. Such peptides significantly inhibited both sperm–egg binding [6] and fertilization rates [6,7] in a murine system, and were found to be even more potent inhibitors than peptides corresponding to the integrin-binding region of mouse fertilin β [6]. Similarly, antibodies raised to the disintegrin loop of murine tMDC I dramatically decreased sperm–egg binding *in vitro* [6], indicating that tMDC I does appear to play a role in sperm–egg interactions in this species, making it an attractive potential target for contraceptive intervention (e.g. immunocontraception) in humans. With this in mind we have cloned and sequenced human tMDC I transcripts from a number of individuals. Surprisingly, a range of differently processed tMDC I transcripts were identified, all of which contained three common features (one deletion and two internal in-frame stop codons) which would render these transcripts incapable of producing a functional protein.

MATERIALS AND METHODS

Reverse-transcription (RT-) PCR

Human testis RNA was obtained from Clontech Laboratories (Palo Alto, CA, U.S.A.) and was a pool of total RNA isolated from 29 sudden-death victims (aged 23–65 years). Human heart, skeletal muscle, uterus and prostate total RNA was extracted as described previously [10].

Total RNA (5 μg) was used as a template for Expand™ (Boehringer-Mannheim, Lewes, East Sussex, U.K.) reverse-transcriptase-directed cDNA synthesis using oligo(dT)$_{12–18}$ as a primer, as recommended by the manufacturer. Subsequent PCR amplification (1 min at 94 °C; 2 min at 58 °C; 1 min at 72 °C; 30 cycles) used 10% of this cDNA and Expand™ (Boehringer-Mannheim) High Fidelity PCR system. The resulting PCR products were resolved on low-melting-temperature agarose gels, appropriate bands excised, purified and cloned into pGEM®-T Easy plasmid vector (Promega U.K., Southampton, U.K.).

Abbreviations used: RT-PCR, reverse-transcription PCR; EST, expressed sequence tag.

[1] To whom correspondence should be addressed (e-mail L.Hall@bris.ac.uk).

The nucleotide sequence data reported in this paper appear in the EMBL, Genbank™ and DDBJ Nucleotide Sequence Databases under the accession number AJ005372.

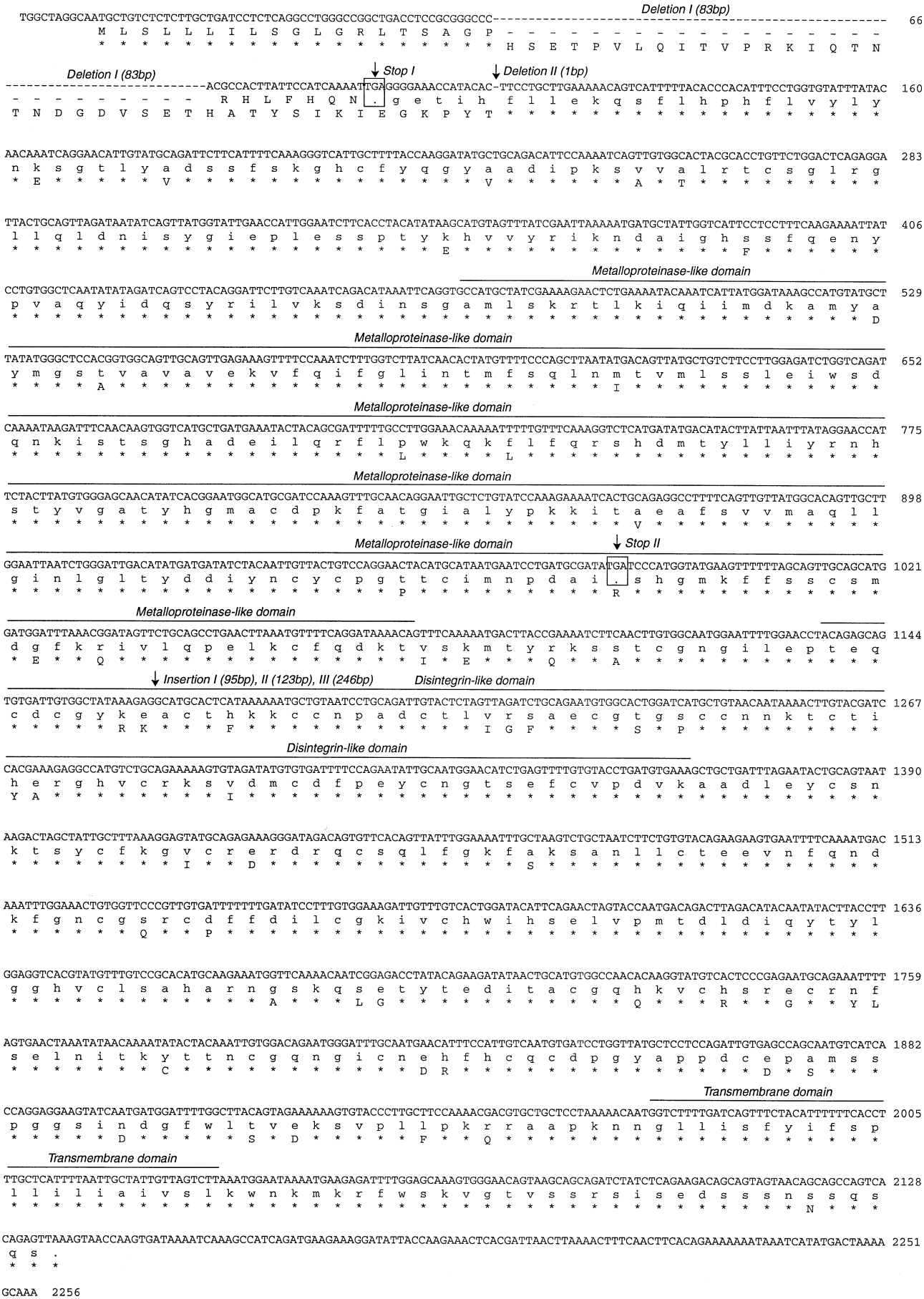

Figure 1 For legend see opposite.

Cloned DNA inserts were completely sequenced on both DNA strands using an ABI 377 automated DNA sequencer and a custom primer walking strategy.

Western-blot analysis of human tMDC I

Macaque (*Macaca fascicularis*) and human testis extracts were prepared by homogenization of tissues in PBS containing 1 mM 4-(2-aminoethyl)benzenesulphonyl fluoride (ICN Biomedicals Ltd., Thame, Oxfordshire, U.K.), followed by incubation on ice in the same buffer containing 1% (v/v) Triton X-100 (Sigma Chemical Co., Poole, Dorset, U.K.). Homogenates were then centrifuged at 10000 ***g*** and supernatants collected.

Plasma-membrane-enriched fractions of macaque caudal-epididymidal sperm and human ejaculated sperm were prepared by detergent extraction with 1% (v/v) Triton X-100 in PBS and vortex mixing [13]. 4-(2-Aminoethyl)benzenesulphonyl fluoride at 1 mM was included in the buffer.

Proteins (100 μg/lane) were separated by electrophoresis under reducing conditions on 12% (w/v) polyacrylamide gels containing SDS, then electroblotted on to PVDF membranes (PolyScreen™; NEN Life Science Products, Brussels, Belgium). Blots were probed for 1 h with either affinity-purified polyclonal antisera (J102 or J103, from two independent rabbits) raised against a recombinant protein containing the region of macaque tMDC I from the end of the disintegrin-like domain through to the natural C-terminus [10], or with polyclonal antisera (H101 or H102) raised against a synthetic peptide (WNKMKRFWSK-VGTVSSRSIS; conjugated to a purified protein derivative of tuberculin) derived from the putative C-terminal cytosolic domain of human tMDC I (see Figure 1). After extensive washes the blots were then incubated with horseradish peroxidase-conjugated swine anti-rabbit IgG (Dako Ltd., High Wycombe, Bucks., U.K.) for a further 1 h. Detection of bound antibodies was achieved by enhanced chemiluminescence (Amersham ECL®; Amersham International, Amersham, Bucks.,U.K.) and exposure of membranes to Hyperfilm® (Amersham).

RESULTS

Cloning human tMDC I transcripts

Initially, a 1kbp fragment encoding the disintegrin-like and cysteine-rich domains of human tMDC I was amplified by PCR using primers based on the macaque orthologue [14] and human testis cDNA as the template. DNA sequence analysis of this cloned PCR product confirmed a high degree of sequence identity with macaque tMDC I. Subsequent PCR reactions utilized primers based on this cloned human tMDC I cDNA fragment and macaque-based primers derived from the 5′ and 3′ non-coding regions. In this way the entire coding region of human tMDC I cDNA was amplified (as a series of extensively overlapping fragments), cloned and sequenced.

Sequence analysis of human tMDC I cDNA clones

DNA sequence analysis of the initial cloned 1 kbp PCR product soon established that two populations of tMDC I cDNA were present in roughly equal proportions. One population showed the expected high degree of sequence identity (96%) with macaque tMDC I throughout its entire length; the other population was identical in sequence, except for the presence of a 95 bp insertion within the region encoding the disintegrin-like domain. Sequence-database analyses indicated that this insertion represented part of an *Alu* repeat sequence (88% identity). Subsequent PCR reactions using cDNA derived from four individuals demonstrated that tMDC I transcripts with and without the 95 bp *Alu* repeat insertion were present in the same individual, suggesting that they arise by alternative splicing, rather than representing a difference between individuals.

Examination of the deduced coding sequence of both of the above classes of tMDC I transcript identified an in-frame termination codon (TGA) in both cases (compared with CGA, arginine, in macaque tMDC I), located towards the C-terminal end of the metalloproteinase-like domain (see Figure 1, *Stop II*).

Subsequent DNA sequence analysis of further cloned PCR products encompassing the entire 'coding region' of human tMDC I confirmed the presence of this in-frame termination codon in all sequenced clones (15 clones from 10 independent PCR reactions). Furthermore, GenBank EST (expressed sequence tag) database searches identified a 441 bp human EST (AA421475) which overlapped this region and also contained the in-frame termination codon.

Analysis of cloned human tMDC I PCR products containing the N-terminal coding region identified (i) an 83 bp deletion (compared with the macaque orthologue; see Figure 1, *Deletion I*) followed by (ii) a second, in-frame, termination codon (Figure 1, *Stop I*), and (iii) a 1 bp deletion (Figure 1, *Deletion II*), in all clones examined (ten clones from six independent PCR reactions; see Figure 1). Attempts to translate these human tMDC I transcripts would yield little more than the signal peptide.

In addition to the 83 bp and 1 bp deletions and two termination codons (present in all cloned human tMDC I transcripts) and the 95 bp *Alu* repeat insertion (in nine clones from six independent PCR reactions), a number of additional variations were observed in some, but not all, tMDC I cDNA clones (Figure 2). These included: (i) a 123 bp *Alu* repeat insertion in place of the 95 bp *Alu* repeat insertion (in two clones from two independent PCR reactions), (ii) a single clone containing a 316 bp deletion within the region encoding the end of the pro-domain and the beginning of the metalloproteinase-like domain and (iii) a single clone containing a 178 bp deletion in the pro-domain encoding region, a 72 bp deletion in the metalloproteinase-like domain and a 246 bp insertion (containing the 95 bp *Alu* repeat sequence) in place of the 95 bp *Alu* repeat sequence.

Tissue distribution of human tMDC I transcripts

In rodents, transcripts for MDC I are exclusively expressed by spermatogenic cells and cannot be detected in other tissues, even using sensitive RT-PCR approaches. By contrast, in the macaque, a similar RT-PCR-based approach indicated that although tMDC I is abundantly expressed in the testis, it is also expressed in a number of other tissues, including skeletal muscle, uterus and ovary, albeit at very much decreased levels [10]. The

Figure 1 Nucleotide sequence of human tMDC I cDNA

The complete nucleotide sequence of human tMDC I cDNA derived from a series of overlapping PCR fragments is shown together with the deduced amino acid sequence. The macaque tMDC I protein sequence [14] is given underneath the human sequence for comparison, with identical residues shown as asterisks. All sequenced transcripts contain deletions of 83 bp (*Deletion I*) and 1 bp (*Deletion II*), as well as two in-frame termination codons (*Stop I*, *Stop II*; boxed). The site of insertion of a number of *Alu* repeat-containing inserts, found in some tMDC I transcripts, is also indicated (*Insertions I, II* and *III*).

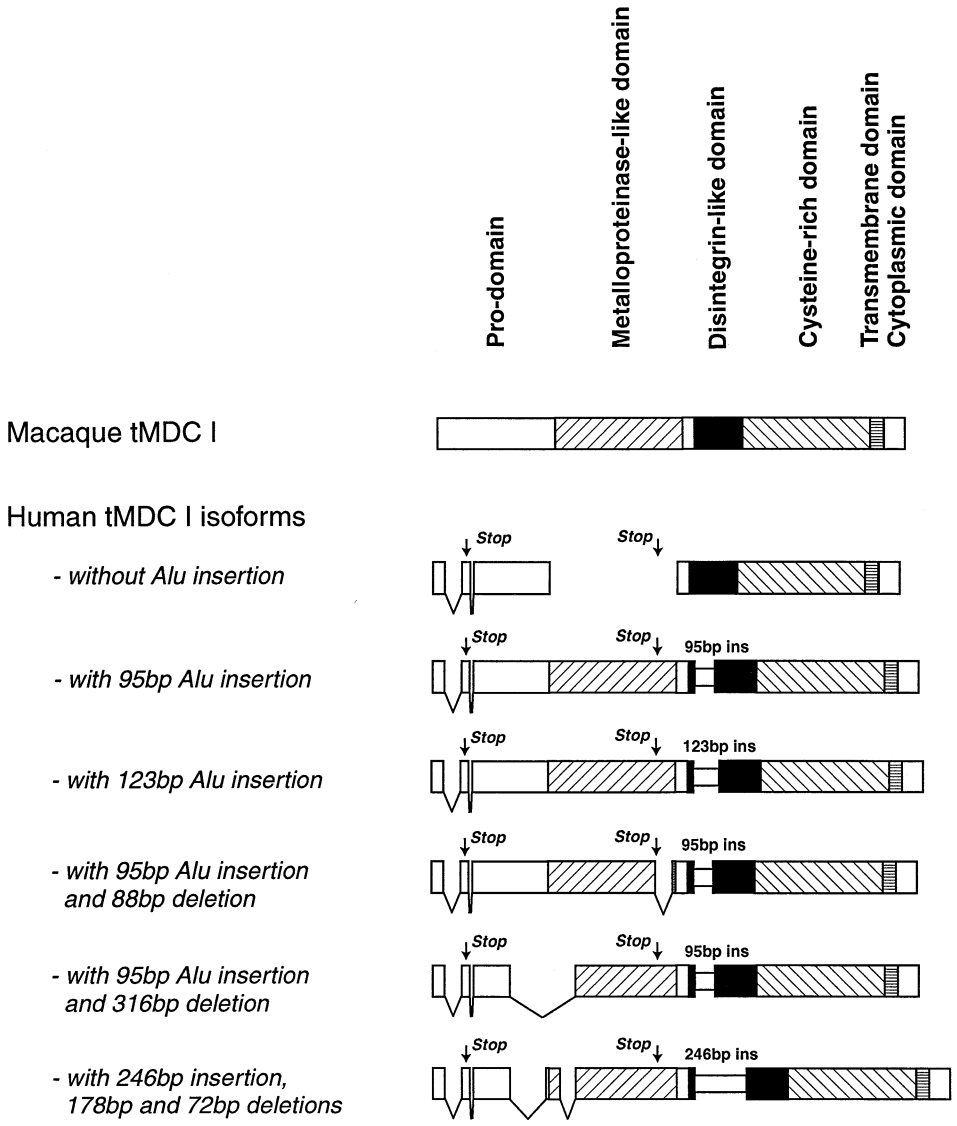

Figure 2 Diagrammatic representation of human tMDC I transcripts

The domain organization of a series of cloned human tMDC I transcripts is shown. Positions of deletions, insertions and in-frame termination codons (relative to the macaque tMDC I orthologue) are indicated, together with the number of independent clones for each type of transcript.

physiological significance of such low levels of transcript, which were not detectable by Northern-blot analysis of total RNA, although difficult to assess, indicated important species differences.

In the present study we have again used RT-PCR with specific primers designed to amplify a PCR product of approx. 400 bp encoding the cysteine-rich domain of human tMDC I. As can be seen in Figure 3, specific transcripts for human tMDC I were expressed abundantly in the testis and at lower levels in all other tissues examined (heart, prostate, skeletal muscle and uterus). These data are similar to those obtained for the macaque, although it would appear from the very limited number of human tissues available that tMDC I transcripts may be even more widely expressed in humans; macaque tMDC I transcripts were not detected in a number of tissues, including heart. In view of our finding that tMDC I transcripts are non-functional in the human, it is tempting to speculate that this has led to a loss of the tight tissue-specific expression observed in rodents. Such an hypothesis might also question the functionality of tMDC I in the macaque where tissue specificity is also not restricted to the testis.

Western-blot analysis of tMDC I

On the basis of transcript analysis, it would appear that humans do not express a tMDC I protein. This was investigated further by Western-blot analysis of membrane preparations from human testis and ejaculated spermatozoa, using four tMDC I-specific antisera; two (J102 and J103) raised against a middle/C-terminal region of the macaque tMDC I protein with approx. 90% overall amino acid sequence identity with human tMDC I, and two (H101 and H102) raised against a C-terminal synthetic peptide derived from the deduced human tMDC I protein

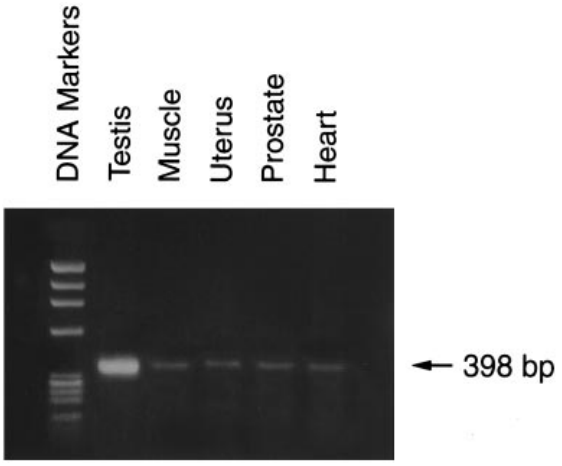

Figure 3 Tissue distribution of human tMDC I transcripts

Human tMDC I transcripts were detected by RT-PCR as described in the Materials and methods section using total RNA from testis, skeletal muscle, uterus, prostate and heart as template and primers derived from the cysteine-rich domain (which is identical in all tMDC I transcripts; see Figure 2). Parallel PCR reactions using glyceraldehyde-3-phosphate dehydrogenase primers were carried out to ensure equivalence of cDNA concentration in each sample.

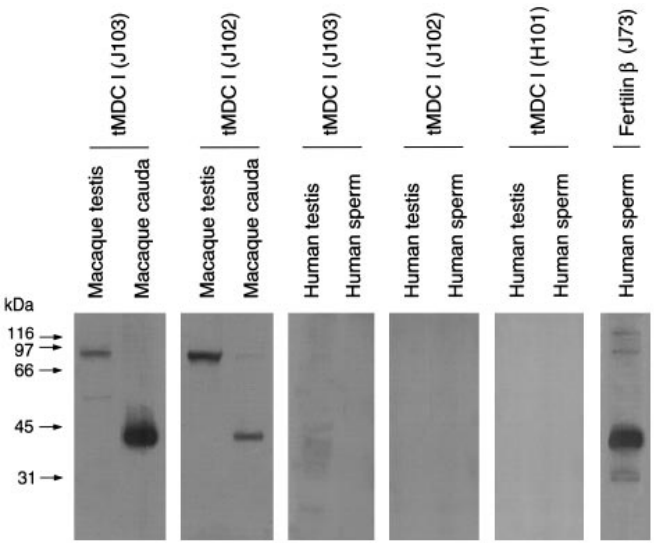

Figure 4 Western-blot analysis of tMDC I

Macaque and human testis and sperm membrane extracts were separated on SDS/12%-PAGE gels, electroblotted and probed with two macaque anti-tMDC I polyclonal antisera (J102 and J103) or a polyclonal antiserum (H101) raised against a synthetic C-terminal peptide derived from the deduced human tMDC I sequence. A second antiserum (H102) raised against the same peptide in a different rabbit gave identical results (results not shown). The human sperm membrane extract was also probed with a macaque fertilin β antiserum to check its integrity.

sequence in a region possessing maximum disparity to other MDC family members. No immunoreactivity was observed with these human samples (Figure 4), even using the sensitive technique of enhanced chemiluminescence. In contrast, antisera J102 and J103 readily detected a protein of approx. 86 kDa in extracts of macaque testicular cells and of approx. 46 kDa in extracts of macaque cauda-epididymidal spermatozoa (Figure 4). To verify the integrity of the human protein preparations, the same human spermatozoa membrane extracts were probed with an antiserum raised against a similar region of macaque recombinant fertilin β [3]. This antiserum detected a protein of the expected size on Western blots (Figure 4).

DISCUSSION

DNA sequence analysis of cloned human tMDC I cDNAs has identified a number of different transcripts, all of which are non-functional owing to the presence of a series of insertions, deletions and in-frame termination codons when compared with its macaque orthologue. Interestingly, all of the insertions and deletions map to the known positions of exon/intron boundaries in the related mouse fertilin β [15] and human brain MDC [16] genes, suggesting that these tMDC I variants arise by alternative splicing. In particular, the 83 bp deletion found in all sequenced human tMDC I transcripts, 19 amino acid residues in from the N-terminal coding region, corresponds precisely to one exon in the mouse fertilin β gene. PCR amplification of human testis cDNA, using primers which flank this 83 bp deletion, produced a single intense band (corresponding to transcripts containing the deletion) when analysed by agarose-gel electrophoresis, with no evidence for any transcripts lacking this deletion. The absence of this 83 bp exon and the associated in-frame termination codon just downstream from this deletion, preclude the synthesis of an active tMDC I protein in the human. Western-blot analyses using antisera raised against macaque tMDC I, or a deduced human tMDC I C-terminal peptide, failed to demonstrate immunoreactivity in the human, providing further strong evidence for the absence of tMDC I protein in the human.

Recently, Adham and co-workers [17] have reported sequences for four human tMDC I transcripts (which they call *CYRN1*), which all contain the same 83 bp deletion that we observe (Figure 1, *Deletion I*) at the beginning of the coding region. However, all four *CYRN1* transcripts apparently lack the 1 bp deletion (Figure 1, *Deletion II*) and the second in-frame termination codon within the metalloproteinase-like domain (Figure 1, *Stop II*) and contain an additional deletion of 2 bp within the region encoding the putative signal peptide. None of these features are observed in any of the many clones we have sequenced. As a result of these differences, two of the four sequences reported by Adham and co-workers [17] apparently contain an open reading frame roughly comparable in size with that found in other tMDC I species orthologues, leading those authors to propose that a functional protein is expressed in the human testis. However, we believe that these sequences are either erroneous or represent rare transcripts which are not found in the majority of individuals. The 2 bp deletion within the signal peptide region, if correct, would cause a frameshift, resulting in a protein with an alternative signal peptide sequence and one which does not contain the features normally expected of a functional signal peptide. This 2 bp deletion, as well as the absence of the 1 bp *Deletion II* (see Figure 1) are not observed in any of the independent clones, isolated from many individuals, which we report here. In addition, two of the four *CYRN1* cDNA clones and four out of four characterized *CYRN1* PCR products (also described by Adham et al. [17]), all contained deletions or insertions associated with the disintegrin-like domain, the region implicated in egg binding.

Our cDNA was synthesized from a pool of testicular RNA isolated from 29 males, and a total of 10 tMDC I clones (from six independent PCR reactions) which span this region have been sequenced. All contain the same sequence. The sequence reported by Adham et al. [17] was derived from a human testis cDNA library, but it is not clear whether this was derived from a single

individual. Although they have deposited sequence data for four different *CYRN1* cDNA clones in the EMBL database (X89654, X89655, X89656 and X89657), somewhat surprisingly all four sequences start and end at exactly the same nucleotide. It is therefore unclear if all four clones have been independently sequenced in their entirety. Finally, the second in-frame termination codon we observed within the metalloproteinase-like domain of human tMDC I cDNA (in 12 clones) is independently confirmed in an EST (AA421475) in the GenBank EST database. In conclusion, taken together with our Western-blotting data, we believe that our cDNA sequence analyses indicate that the majority of human males do not express a functional tMDC I protein.

The question therefore arises as to whether an alternative MDC protein could fulfil the role of rodent tMDC I in humans. In this context it is noteworthy that although there is only a single murine tMDC I gene (*cyrn1*; [18]), the human genome appears to contain two closely related tMDC I-like genes; *CYRN1* and *CYRN2* [17]. Of these, human *CYRN1* is most closely related to the macaque and rodent tMDC I gene, suggesting that this is the true orthologue of the macaque and rodent sequences. Only 540 bp of *CYRN2* sequence data is available (all exon, from a genomic clone; EMBL accession number Y10615; [17]) and, using RT-PCR with primers based on this sequence, Adham and co-workers [17] have suggested that *CYRN2* transcripts are expressed in human testis, but not thymus, placenta or kidney. Using similar primers we initially obtained similar PCR products with human testis cDNA (results not shown), but these were subsequently found to be derived from low levels of genomic DNA contamination in the commercial RNA used for cDNA synthesis. In the light of further experiments, which included extensive DNase treatment of the RNA template, we conclude that expression of the human *CYRN2* gene is barely detectable, if at all, and is certainly at a considerably lower level than the expression of tMDC I (*CYRN1*). It therefore seems unlikely that the *cyrn*2 gene, which is absent in rodents, replaces the non-functional tMDC I (*CYRN1*) gene in the human.

The finding of a non-functional human MDC gene, implicated in fertility in rodent models, is not restricted to tMDC I. As mentioned in the Introduction, the human fertilin α gene, although expressed, is similarly non-functional, due to the presence of insertions, deletions and an in-frame termination codon [5]. However, fertilin α and tMDC I belong to a growing family of related proteins, many of which are expressed in the testis, some exclusively in spermatogenic cells. Although the function of many of these MDC proteins has yet to be determined, the presence of a conserved disintegrin-like domain would imply similar roles in cell–cell interactions; such proteins could possibly replace fertilin α and tMDC I in the human. Alternatively, the plethora of MDC proteins expressed on mature sperm may represent a family of redundant oolemma-binding molecules, whereby the loss of some does not prevent the overall process of sperm–egg binding.

Received 5 June 1998/19 June 1998; accepted 23 June 1998

This work was supported by the Wellcome Trust (U.K.). We thank Rhiannon Murray for assistance with automated DNA sequencing and Nick Bishop for assistance with cDNA cloning.

REFERENCES

1 Blobel, C. P., Wolfsberg, T. G., Turck, C. W., Myles, D. G., Primakoff, P. and White, J. M. (1992) Nature (London) **356**, 248–252
2 Blobel, C. P., Myles, D. G., Primakoff, P. and White, J. M. (1990) J. Cell Biol. **111**, 69–78
3 Frayne, J., Jury, J. A., Barker, H. L., Perry, A. C. F., Jones, R. and Hall, L. (1998) Mol. Human Reprod. **4**, 429–437
4 Perry, A. C. F., Gichuhi, P. M., Jones, R. and Hall, L. (1995) Biochem. J. **307**, 843–850
5 Jury, J., Frayne, J. and Hall, L. (1997) Biochem. J. **321**, 577–581
6 Yuan, R., Primakoff, P. and Myles, D. G. (1997) J. Cell Biol. **137**, 105–112
7 Linder, B. and Heinlein, U. A. O. (1997) Dev. Growth Differ. **39**, 243–247
8 Wolfsberg, T. G., Straight, P. D., Gerena, R. L., Huovila, A.-P. J., Primakoff, P., Myles, D. G. and White, J. M. (1995) Dev. Biol. **169**, 378–383
9 Linder, B., Bammer, S. and Heinlein, U. A. O. (1995) Exp. Cell Res. **221**, 66–72.
10 Frayne, J., Jury, J. A., Barker, H. L. and Hall, L. (1997) Mol. Reprod. Dev. **48**, 159–167
11 Forsbach, A. and Heinlein, U. A. O. (1998) J. Exp. Biol. **201**, 861–867
12 Evans, J. P., Kopf, G. S. and Schultz, R. M. (1997) Dev. Biol. **187**, 79–93
13 Jones, R. (1986) J. Reprod. Fertil. **77**, 435–449
14 Barker, H. L., Perry, A. C. F., Jones, R. and Hall, L. (1994) Biochim. Biophys. Acta **1218**, 429–431
15 Cho, C., Turner, L., Primakoff, P and Myles, D. G. (1997) Dev. Genet. **20**, 320–328
16 Katagiri, T., Harada, Y., Emi, M. and Nakamura, Y. (1995) Cytogenet. Cell Genet. **68**, 39–44
17 Adham, I. M., Kim, Y., Shamsadin, R., Heinlein, U. A. O., von Beust, G., Mattei, M.-G. and Engel, W. (1998) DNA Cell Biol. **17**, 161–168
18 Lemaire, L., Johnson, K. R., Bammer, S., Petry, P., Ruddle, F. H. and Heinlein, U. A. O. (1994) Genomics **21**, 409–414