RESEARCH COMMUNICATION

The human fertilin α gene is non-functional: implications for its proposed role in fertilization

Jennifer A. JURY, Jan FRAYNE and Len HALL*

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

In the guinea-pig, the α subunit of the fertilin complex, a heterodimeric surface membrane glycoprotein found on the head region of spermatozoa, has previously been proposed to mediate membrane fusion with the oolemma plasma membrane during fertilization. Here we describe experiments which indicate that the only fertilin α-like gene in humans is an expressed, but non-functional, pseudogene, possibly derived by genetic recombination between the two fertilin α genes found in some primates. This finding clearly raises questions about the importance and/or role of fertilin α in mammalian fertilization.

INTRODUCTION

Mammalian fertilization involves species-specific recognition of two morphologically distinct gametes, culminating in their fusion. Recognition appears to occur at two levels. After an initial interaction between sperm surface components and the zona pellucida surrounding the oocyte, specific sperm membrane antigens subsequently bind to cognate receptors associated with the oolemma plasma membrane, and gamete membrane fusion ensues. The sperm surface fertilin complex (formerly PH30) has been proposed to play an important role in both oolemma plasma-membrane recognition and subsequent membrane fusion [1]. The fertilin complex was first described in the guinea-pig [2–4], where it was reported to exist as a heterodimeric glycoprotein composed of an α and β subunit, both of which are synthesized in the testis as larger precursors. The fertilin α precursor then undergoes endoproteolytic processing in the testis to produce the mature form, whereas the β subunit is partially processed in the testis, with final processing occurring during epididymal transit, coincident with the acquisition of fertilizing capacity by the spermatozoon [3]. More recently, cDNA sequences encoding fertilin α and β subunits from a number of mammalian species have been reported, including macaque (*Macaca fascicularis*) [5], mouse [6], rat (EMBL database accession numbers Y08616 and X99794) and rabbit [7].

Sequence analyses of the mammalian fertilin subunits have revealed that they belong to a large family of membrane proteins which exhibit sequence similarity to snake-venom haemorrhagic polypeptides (reprolysins) possessing metalloproteinase and disintegrin activities [4–9]. This family of mammalian proteins has been alternatively referred to as the MDC (**M**etalloproteinase-like, **D**isintegrin-like, **C**ysteine-rich) family [5,8,10,11] and the ADAM (**A** **D**isintegrin **A**nd **M**etalloproteinase) family [6,9]. Whilst some MDC transcripts are specific to the male reproductive tract, suggesting functions peculiar to those tissues, others have a broader tissue distribution and may therefore play a more general role [6,10,11].

Several lines of evidence implicate the disintegrin-like domain of the fertilin β subunit in oolemma plasma-membrane recognition. The egg possesses at least one class of surface integrin [12], and short peptides containing an RGD motif (characteristic of many, but not all, disintegrin–integrin interactions) have long been known to inhibit sperm–egg binding *in vitro* [13]. More importantly, short peptides corresponding to the disintegrin-like domain of fertilin β inhibit sperm–egg binding [14,12], arguably by competing with the fertilin complex for its cognate integrin receptor on the egg. Establishing a role for the fertilin α subunit has not been so straightforward. Blobel and co-workers [1] have proposed that fertilin α may be involved in sperm–oolemma membrane fusion by virtue of the presence of a putative fusogenic domain (similar to viral fusion peptides) within the cysteine-rich domain of the subunit of guinea-pig fertilin. Such fusogenic peptide sequences are recognizable by their ability to form α helices with a hydrophobic face. A similar putative fusogenic domain has been identified in meltrin α, another member of the MDC family which is expressed in myoblast cells and has been proposed to function in membrane fusion, resulting in myotubule formation [15]. However, when the corresponding regions of the macaque, rat, rabbit and mouse orthologues of fertilin α are examined [5], the similarity to viral fusion peptides is less convincing. In addition, the recent finding that the fertilin α subunit is expressed in a number of tissues in addition to the testis [10,6], albeit at much reduced levels, casts doubt on a unique role in fertilization.

Recent sequence analysis of full-length cloned cDNAs encoding macaque fertilin α and β subunits, led to the identification of novel α subunit isoforms [5] not previously observed in any other species. These two alternative forms of fertilin α, termed α**I** and α**II**, are identical throughout most of the coding region, but differ (although they are related) at both N- and C-termini (Figure 1). Since a number of independent cDNA clones for each isoform were isolated from the same testis cDNA library, prepared from a single animal, they do not represent polymorphisms of the same gene found in different individuals. Instead, these two

Abbreviations used: MDC; metalloproteinase-like, disintegrin-like, cysteine-rich; RT, reverse transcription.

* To whom correspondence should be addressed.

The nucleotide sequence data reported in this paper appear in the EMBL Nucleotide Sequence Database under the accession number Y09232.

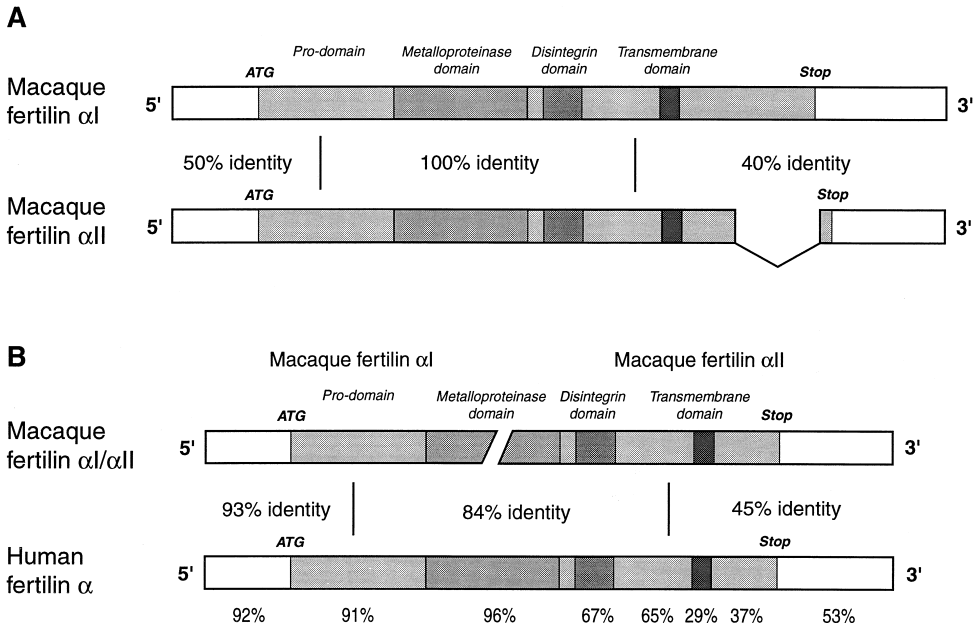

Figure 1 Structural organization of the macaque cDNAs encoding the fertilin αI and αII isoforms and their relationship to the human fertilin α pseudogene

(**A**) Transcripts encoding macaque fertilin αI and αII isoforms possess a common central region and differing, but related, 5′ and 3′ regions with nucleotide identities as shown. (**B**) The human fertilin α pseudogene shows sequence similarity to macaque fertilin αI cDNA at the 5′ end and to macaque fertilin αII at the 3′ end. Nucleotide sequence identities between the 5′ unique, central common, and 3′ unique region of macaque fertilin α and the human pseudogene are indicated between the two sequences. Nucleotide sequence identities between the corresponding 5′ non-coding regions, pro-domains, metalloproteinase-like domains, disintegrin-like domains, cysteine-rich domains, transmembrane domains, cytoplasmic domains and 3′ non-coding regions are given below the human sequence.

different transcripts, which are expressed in approximately equivalent amounts on the basis of Northern-blot analysis with isoform-specific probes [5], suggest the presence of more than one fertilin α gene and/or alternative splicing of the primary transcript. In an attempt to study this further and establish whether a similar situation exists in the human, we embarked on an analysis of the fertilin α structural gene(s).

MATERIALS AND METHODS

Isolation of a human fertilin α genomic sequence

Recombinant phage hybridizing to a full-length, cloned macaque fertilin αI cDNA probe [5] were isolated from a bacteriophage λ-based human genomic DNA library (Stratagene) using standard techniques. Following restriction and Southern-blot analysis of the resulting clones, a region of about 4.5 kb from one recombinant phage was subcloned, sequenced in its entirety on both DNA strands and found to contain a fertilin α pseudogene.

Genomic PCR and identification of fertilin α transcripts

Human testis total RNA was obtained from Clontech; monkey testis total RNA was isolated as described previously [5,16]. Total RNA (2 μg) was used as a template for Expand™ (Boehringer Mannheim) 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) employed 20% of this cDNA and an Expand™ (Boehringer Mannheim) High Fidelity PCR System. Similar conditions were used for PCR amplifications using genomic DNA (1 μg) as a template.

RESULTS AND DISCUSSION

Preliminary experiments (results not shown) utilized a PCR-based approach with macaque, baboon (*Papio cynocephalus anubis*) and human genomic DNA, to identify fertilin α genomic fragments and establish the number and approximate size of intronic sequences. Several different PCR primer pairs, based on macaque fertilin αI and αII cDNA sequences [5], were employed. Surprisingly, all PCR products obtained using macaque genomic DNA were of sizes comparable with those generated from the corresponding cDNA, including those obtained with primers designed to amplify the entire protein coding region (see EMBL database accession no. Y09304). These initial results strongly suggested that the macaque fertilin α genes lack introns (at least within the protein coding region) and that fertilin αI and αII transcripts do not arise by alternative splicing but are products of two different genes, a result which was confirmed by complete sequence determination of these amplified macaque genomic sequences. PCR experiments using the same macaque-based primers and baboon genomic DNA demonstrated that the baboon also possesses intronless fertilin αI and αII genes. However, when similar studies were performed using human genomic DNA as the template, only primer sequences derived from the macaque fertilin αI 5′ unique cDNA sequence or central common cDNA sequence led to the amplification of PCR products, suggesting the presence of a single fertilin gene in the human genome, similar to macaque fertilin αI at the 5′ end, but apparently differing substantially from both αI and αII at the 3′ end. All PCR products derived from human genomic DNA were of a similar size to those predicted from the macaque fertilin αI cDNA sequence, indicating the absence of introns. This absence of introns from fertilin α genes was surprising, since sequence

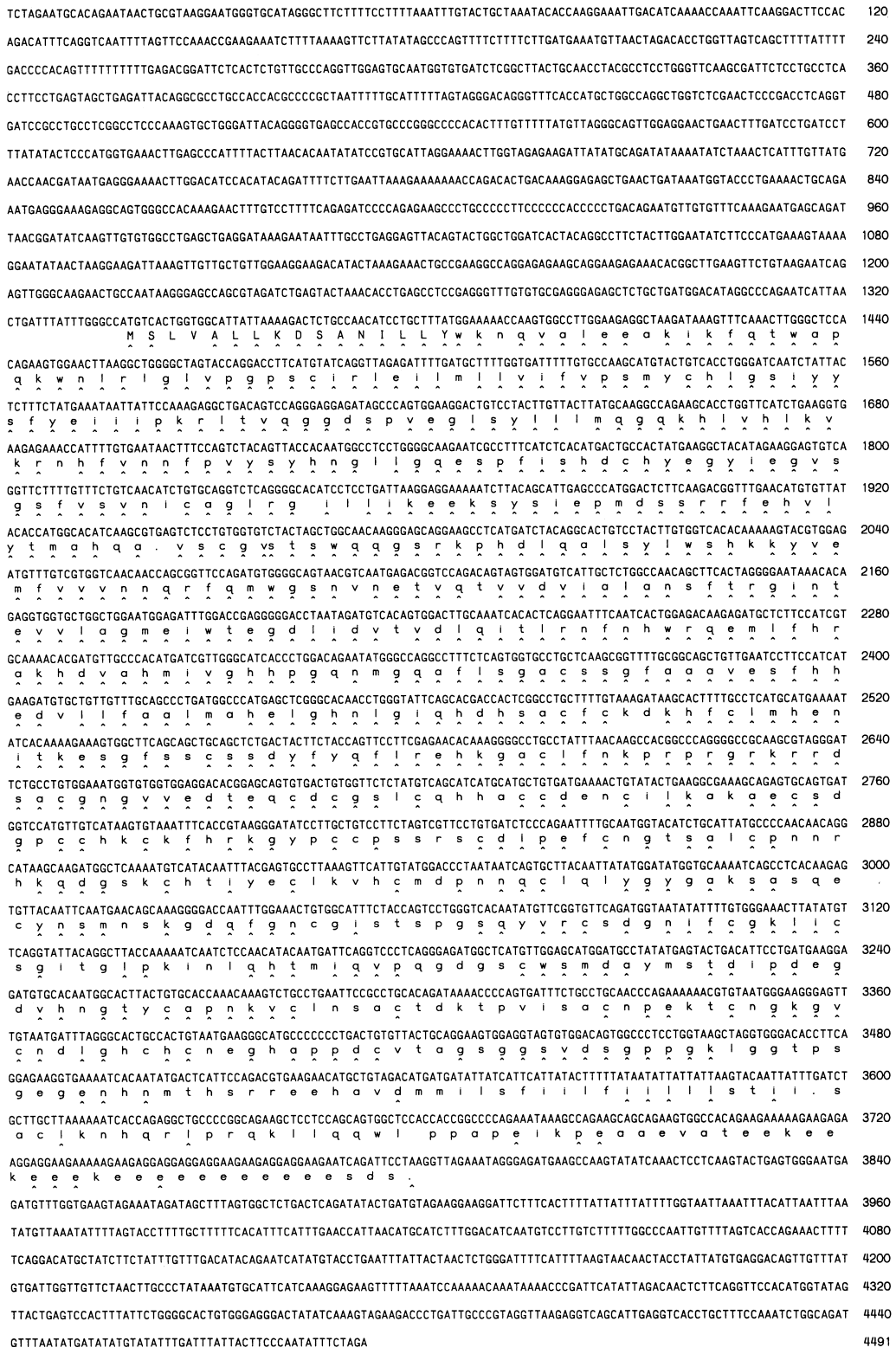

Figure 2 Nucleotide sequence of the human fertilin α pseudogene

The 'coding region' was assigned on the basis of sequence similarity to the macaque fertilin αI and αII orthologues [5]. Amino acid residues shown in upper case (at the 5′ end of the 'coding region') are in-frame; those shown in lower case are out-of-frame owing to the presence of insertions or deletions relative to the macaque sequences. Residues conserved at the corresponding position in the macaque fertilin α sequence [5] are indicated (∧).

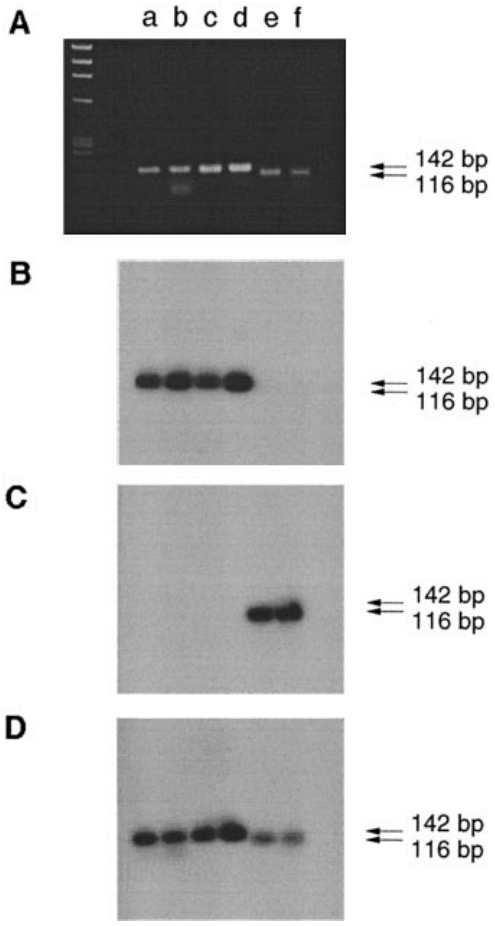

Figure 3 Identification of fertilin α gene sequences and their transcripts in macaque, baboon and human tissues

(**A**) Part of the fertilin α gene (and its cDNA) was amplified by PCR using primers which are totally conserved between the primate and human genes but which flank the 26 bp deletion found within the human gene (see the text). Templates used were (a) macaque, (c) baboon and (e) human genomic DNA, or (b) macaque, (d) baboon and (f) human testis cDNA. Following electrophoresis, PCR products (approx. 50–100 ng) were transferred to a nylon membrane and hybridized under stringent conditions to (**B**) an oligonucleotide from within the 26 bp sequence deleted from the human fertilin α gene, (**C**) an oligonucleotide which straddles the site of deletion (and should therefore only hybridize to the deleted form) and (**D**) an oligonucleotide common to both primate and human fertilin genes.

analysis of the human brain MDC gene [17] revealed the presence of 27 introns and our preliminary data has indicated introns in many other MDC genes expressed in the reproductive tract, including fertilin β.

In an attempt to characterize this human fertilin gene, identified by PCR, further, a human genomic DNA library (Stratagene) was screened with a full-length cDNA probe derived from cloned macaque fertilin αI cDNA [5] (the central common region of which should hybridize to both αI and αII sequences) and a number of positively-hybridizing genomic clones were purified through successive rounds of screening. Preliminary restriction and Southern-blot analyses of these genomic clones indicated that they were not siblings, but were overlapping regions of the same genetic locus. Fragments hybridizing to macaque fertilin αI cDNA probes were confined to a region of about 4.5 kb within the central region of one human genomic clone, and this region was subcloned and sequenced (Figure 2). Comparison of this human fertilin genomic sequence with the macaque fertilin αI and αII cDNAs (see Figure 1) indicate that (i) nucleotides 984–1630 of the human sequence exhibit a high degree of identity (93 %) with the cloned 5′ unique region of macaque fertilin αI (but not αII) cDNA, (ii) nucleotides 1631–3451 of the human sequence show a high degree of identity (84 %) with the central region, common to both macaque fertilin α isoform cDNAs, (iii) nucleotides 3452–4255 of the human sequence exhibit significant, but poorer, identity (45 %) with the cloned 3′ unique region of macaque fertilin αII (but not αI) cDNA, (iv) the human sequence is devoid of introns within those regions corresponding to the macaque fertilin cDNAs, and (v) the human sequenced region includes about 1 kb of 5′ flanking sequence and about 0.23 kb of 3′ flanking sequence. More importantly, the 'coding region' of the human fertilin gene contained a number of small insertions and deletions, as well as termination codons, many of which destroy the expected reading frame. This cloned human fertilin α sequence therefore represents a non-functional pseudogene. To confirm that these deletions and insertions were not acquired during genomic cloning and/or amplification of the resulting recombinants in *E. coli*, a part of the intronless human pseudogene containing several of these changes (nt 1293–2003 in Figure 2) was independently amplified from human genomic DNA and found to possess a sequence identical with that of the library-derived gene.

Subsequent experiments strongly suggested that this human fertilin α pseudogene is the only fertilin α-related sequence in the human genome. First, the human pseudogene sequence contains a deletion of 26 bp within the 'coding region' (after nt 1959 in Figure 2) when compared with the macaque cDNA sequence, in a region of otherwise high sequence identity (98 %). PCR primers flanking this deletion, corresponding to regions which are 100 % conserved in macaque fertilin αI and αII, as well as the human pseudogene (and hence might reasonably be expected to exist in any unidentified functional human fertilin α gene), amplify the expected fragment of 142 bp using macaque or baboon genomic DNA as the template, but only a product of 116 bp (corresponding to the pseudogene sequence) using human genomic DNA (Figure 3). The identity of these PCR products was confirmed after transfer to nylon membranes and hybridization analysis with short oligonucleotide probes specific to functional (macaque-like) and non-functional (human-like) fertilin α sequences (Figure 3). Secondly, using these same primers, fertilin α transcripts (142 bp PCR product) could be readily detected as abundant RNAs in macaque and baboon testis by reverse transcription-PCR (RT-PCR). Similar RT-PCR experiments using human testis RNA isolated from 29 sudden-death victims (with ages ranging from 23 to 65) readily identified fertilin α PCR products of 116 bp, representing transcription of the non-functional pseudogene, with no evidence for the presence of a larger PCR product (142 bp) indicative of a functional fertilin α transcript (Figure 3). Direct sequence determination of this uncloned 116 bp PCR product (derived from human testis RNA) confirmed it to be identical with the corresponding region of the human fertilin α pseudogene, strongly supporting the view that the abundant fertilin α-like transcripts expressed in human testis are derived from this non-functional gene. From these, and additional PCR-based experiments, we conclude that the human genome contains a single, transcribed, but non-functional, fertilin α gene.

It is difficult to reconcile the structures of the two macaque (and baboon) fertilin α genes. Whilst they could have arisen by gene duplication, it is hard to understand why the central common region (which accounts for more than half of the transcript) has remained invariant, whereas the 5′ and 3′ termini

have diverged in sequence considerably (see Figure 1). Perhaps more likely, two divergent fertilin α genes may have undergone relatively recent recombination, resulting in the transfer of the central region of one gene to both genes. However, given this organization in macaques and baboons, it is easy to envisage how a recombination event within the common, central region of the two primate fertilin α genes (particularly if they are tandemly arranged) could lead to the single gene found in human DNA, with a fertilin αI-like 5′ end, a common central region and a fertilin αII-like 3′ end. If such an event produced a small deletion or insertion at the recombination site, a non-functional hybrid gene would result and, in the absence of selective pressure, this pseudogene might then be expected to accumulate additional deletions, insertions and base changes similar to those found in the human sequence. In this respect, it will be particularly interesting and instructive to compare the organization of the fertilin α gene(s) in other primates, particularly the great apes, and this is currently underway.

Clearly the absence of a functional human fertilin α protein casts doubt on its proposed role in fertilization. The widely accepted belief that fertilin exists as a heterodimeric complex of α and β subunits on the sperm surface is based on studies in the guinea-pig, which indicated that affinity-purified fertilin α and fertilin β antibodies were able to identify co-migrating antigens, presumably in a complex, on non-reducing polyacrylamide gels after treatment at room temperature with 0.5% SDS [3]. We have been unable to identify similar fertilin α–β complexes between the macaque subunits under similar experimental conditions, raising the possibility that the observed guinea-pig complex may be an *in vitro* artefact. The recent finding that fertilin α is expressed in a wide range of tissues [10,6], whereas fertilin β appears to be testis-specific, may also raise doubts that fertilin α exists as a fertilin α–β complex *in vivo* or that it plays a primary role in sperm–egg recognition and subsequent membrane fusion. Alternatively, fertilin α may form an α–β complex in some mammalian species where it may play a facilitatory rather than essential role, possibly by promoting sperm–egg membrane fusion, or it is possible that another member of the large and rapidly growing family of testicular MDC proteins may substitute for fertilin α in the human. The idea that some MDC proteins may have degenerate functions would be supported by our recent finding that the tissue distribution and relative abundance of some MDC transcripts apparently vary between mammalian species.

To date, the metalloproteinase-like domains of about 14 different members (excluding species orthologues) of the mammalian MDC protein family are available in sequence databases. Of these, the fertilin α precursor shows one of the best alignments with the catalytic-centre residues of reprolysin-like metalloproteinases (consensus sequence HEXXHXLGXXHD [10,18]), and the metalloproteinase-like domain is the most highly conserved region when the fertilin α precursors of different species are compared. Whilst enzymic activity has yet to be demonstrated in the case of the fertilin α precursor (the metalloproteinase domain is removed and lost from the precursor soon after its synthesis in the testis [3]), it seems feasible, on the basis of its sequence, that it may be an active proteinase. Such a possibility should be borne in mind when attempting to establish a function for fertilin α in both reproductive and non-reproductive tissues.

Received 25 November 1996; accepted 5 December 1996

This work was supported by the Medical Research Council (U.K.) and the Wellcome Trust (U.K.). We thank Rhiannon Murray for assistance with automated DNA sequencing.

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 Primakoff, P., Hyatt, H. and Tredick-Kline, J. (1987) J. Cell Biol. **104**, 141–149
3 Blobel, C. P., Myles, D. G., Primakoff, P. and White, J. M. (1990) J. Cell Biol. **111**, 69–78
4 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
5 Perry, A. C. F., Gichuhi, P. M., Jones, R. and Hall, L. (1995) Biochem. J. **307**, 843–850
6 Wolfsberg, T. G., Straight, P. D., Gerena, R. L., Huovila, A. P., Primakoff, P., Myles, D. G. and White, J. M. (1995) Dev. Biol. **169**, 378–383
7 Hardy, C. M. and Holland, M. K. (1996) Mol. Reprod. Dev. **45**, 107–116
8 Perry, A. C. F., Jones, R., Barker, P. J. and Hall, L. (1992) Biochem. J. **286**, 671–675
9 Wolfsberg, T. G., Primakoff, P., Myles, D. G. and White, J. M. (1995) J. Cell Biol. **131**, 275–278
10 Perry, A. C. F., Jones, R. and Hall, L. (1995) Biochem. J. **312**, 239–244
11 Weskamp, G., Krätzschmar, J., Reid, M. S. and Blobel, C. P. (1996) J. Cell Biol. **132**, 717–726
12 Almeida, E. A. C., Huovila, A.-P. J., Sutherland, A. E., Stephens, L. E., Calarco, P. G., Shaw, L. M., Mercurio, A. M., Sonnenberg, A., Primakoff, P., Myles, D. G. and White, J. M. (1995) Cell **81**, 1095–1104
13 Bronson, R. A. and Fusi, F. (1990) Biol. Reprod. **43**, 1019–1025
14 Myles, D. G., Kimmel, L. H., Blobel, C. P., White, J. M. and Primakoff, P. (1994) Proc. Natl. Acad. Sci., U.S.A. **91**, 4195–4198
15 Yagami-Hiromasa, T., Sato, T., Kurisaki, T., Kamijo, K., Nabeshima, Y. and Fujisawa-Sehara, A. (1995) Nature (London) **377**, 652–656
16 Girotti, M., Jones, R., Emery, D. C., Chia, W. and Hall, L. (1992) Biochem. J. **281**, 203–210
17 Katagiri, T., Harada, Y., Emi, M. and Nakamura, Y. (1995) Cytogenet. Cell Genet. **68**, 39–44
18 Gomis-Rüth, F.-X., Kress, L. F. and Bode, W. (1993) EMBO J. **12**, 4151–4157