RESEARCH ARTICLE

Cloning of a novel gene, *Cymg1*, related to family 2 cystatins and expressed at specific stages of mouse testis development

Y. XIANG, D. S. NIE and G. X. LU*

Human Reproductive and Stem Cell Engineering Institute, Central South University, Changsha 410078, People's Republic of China

Abstract

We have cloned a novel gene, *Cymg1* (GenBank accession number AY600990), from a mouse testis cDNA library. *Cymg1* is located in 2G3 of mouse chromosome 2. The cDNA includes an open reading frame that encodes 141 amino acid residues. The encoded polypeptide has a cysteine protease inhibitor domain found in the family 2 cystatins but lacks critical consensus sites important for cysteine protease inhibition. These characteristics are seen in the proteins of the CRES subfamily of the family 2 cystatins which are expressed specifically in the reproductive tract. CYMG1 protein shows 44% identity with mouse CRES and 30% identity with mouse cystatin C. Northern blot analysis showed that the *Cymg1* gene was specifically expressed in adult mouse testis. RT-PCR also showed that *Cymg1* was expressed in testis and spermatogonial cells. *Cymg1* expression level varied in the different developmental stages of mouse testis, and were coincidental with spermatogenesis and sex maturation. These results indicate that *Cymg1* may play important roles in mouse spermatogenesis and sex maturation

[Xiang Y., Nie D. S. and Lu G. X. 2004 Cloning of a novel gene, *Cymg1*, related to family 2 cystatins and expressed at specific stages of mouse testis development.*J. Genet*. **83**, 257–263]

Introduction

Although several external factors have been claimed to affect male reproduction, evidences that genetic factors play important roles are being accumulated (Bourrouillou *et al*. 1985; Adams *et al*. 1995). Many genes have been confirmed to be essential for spermatogenesis, but there still are other genes involved in the process. Identification of these new genes and their roles is of great importance in understanding the biology of spermatogenesis (Vogt *et al*. 1996).

Cystatins are a superfamily of reversible competitive inhibitors of C1 cysteine proteases such as plant papain and the mammalian cathepsins B, H and L (Turk and Bode 1991). The superfamily consists of three families, stefins (family 1), cystatins (family 2) and kininogens (family 3). The family 2 cystatins, represented by cystatins C, D, E/M, F, S, SN, SA, are secretory proteins of 13 kDa that contain two characteristic disulphide bonds in the C-terminus; some family members are glycosylated. Several genes, such as *Cres* (cystatin-related epididymal spermatogenic), have been identified to be related to the family 2 cystatins. They have a cysteine protease inhibitor domain but lack critical consensus sites important for cysteine protease inhibition. In addition, these genes are primarily expressed in the reproductive tract, suggesting that they may have evolved to perform tissue-specific functions distinct from those of the typical cystatins (Cornwall and Hsia 2003). Here we report EST (expressed sequence tag) assembly and cloning of a novel gene, *Cymg1*, from a mouse testis cDNA library. *Cymg1* encodes a protein (CYMG1) of 141 amino acid residues which may be a new member of the CRES subgroup of the family 2 cystatins. PCR (polymerase chain reaction), RT-PCR (reverse transcription PCR), Northern blot analysis and expression in *E. coli* M15

*For correspondence. E-mail: lgxdirector@sina.com.cn, lgxdirector@yahoo.com.

Keywords. cystatins; gene cloning; testis; spermatogenesis; sex maturation.

confirmed tissue-specific expression of the gene in mouse testis.

Materials and methods

EST search and gene assembly (*Qian et al. 2002; Stephen et al. 2003*)

Digital differential display (DDD; http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.html), BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) and ExPASy (http://www.expasy.org) software packages were used to search for ESTs in a mouse testis cDNA library. ESTs were assembled using CAP (contig assembly program) Sequence Assembly Machine software (e.g. http://bio.ifom-firc.it/ASSEMBLY/assemble.html). The assembled *Cymg1* gene was sustained by 61 ESTs, two of which came from mouse ovary and brain, and the rest from mouse testis.

Bioinformatics analysis and protein characterization (*Gill and Sanseau 2000; Pruitt et al. 2000*)

Translation into the amino acid sequence from the cDNA was carried out using the ORF Finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) and ExPASy. The protein sequence was analysed using BLAST with the nr database and SMART (Simple Modular Architecture Research Tool; http://smart.embl-heidelberg.de).

RNA isolation (*Xing et al. 2003*)

Fresh mouse testis, epididymis, ovary, kidney, heart, liver, muscle, spleen, lung, cerebellum and cerebrum came from adult C57BL/6 mice; spermatogonial cells (CRL-2053 cells, from American Type Culture Collection, USA) were cultured by our stem cell centre; testes at different developmental stages from BALB/c mice were frozen in liquid nitrogen immediately after dissection and stored at – 80ºC. Total RNA was isolated using the Gentra RNA isolation system (Gentra Systems, USA) according to the protocol provided.

PCR amplification of full-length Cymg1

The primer pair 5′-GGGAGGAGAGGGAAGTCAGA-3′ and 5′-GGTGGTCCAQCTATCTTACAA-3′ was designed according to the full-length *Cymg1* and used in PCR with Advantage 2 DNA polymerase (Clontech, USA) and Marathon-Ready™ cDNA of mouse testis (Clontech) as template. PCR amplification cycles involved initial denaturation at 95°C for 1.5 min; 35 cycles at 94°C for 40 s, 60°C for 40 s and 68°C for 40 s; 68°C for 7 min; and then holding at 4°C. The PCR products were separated on 1.5% agarose gel. These PCR fragments were cloned into pUCm-T vectors and sequenced.

RT-PCR analysis

***Synthesis of first strand cDNA*:** This was done using the RevertAid™ First Strand cDNA Synthesis Kit (Fermentas). Reaction mixture containing 10 µl ribonuclease-free water, 1 µl total RNA (0.23–0.26 µg/µl) and 1 µl oligo(dT) primer (0.5 µg/µl) was incubated at 70°C for 5 min and chilled on ice, and 4 µl 5× reaction buffer, 1 µl ribonuclease inhibitor (20 U/µl) and 2 µl 10 mM dNTP were added. The mixture was incubated at 37ºC for 5 min, and then incubated at 42ºC for 60 min after adding 1 µl RevertAid™ M-MuLV reverse trancriptase (200 U/µl). The reaction was terminated by heating at 70ºC for 10 min, and the mix was then kept at 4°C.

***PCR amplification of Cymg1*:** Using the primer pair 5′-TCTGGAAAGAAAATAGGAACTTGG-3′ and 5′-AAGAAAAGTAAGAGTGGCAAGGTG-3′, PCR was performed to detect *Cymg1* expression in C57BL/6 mouse tissues. PCR conditions were the same as above.

Analysis of Cymg1 expression by RT-PCR (*Luk et al. 2003*): RNA isolated from testes at different developmental stages was reverse-transcribed into cDNA. The amplification reaction was performed in a 20-µl reaction mixture containing 13.55 µl deionized water, 2 µl 10 × PCR reaction buffer, 2 µl 2.5 mM dNTP, 0.4 µl *Taq* DNA polymerase, 0.45 µl primer pair mixture (0.5 µg/µl) and 1.6 µl cDNA, in a PE9600 (PerkinElmer, USA) programmable thermal cycler: initial denaturation at 95ºC for 90 s, and then 35 cycles of 94°C for 40 s, 60°C for 40 s and 72ºC for 40 s. The extension step in the last cycle was at 72ºC for 5 min, and the mixture was finally kept at 4ºC. The PCR products were then separated on a 1.5% agarose gel. *G3pdh* (glucose-3-phosphate dehydrogenase) expression was checked as a control.

Northern blot (*Sato et al. 1993*)

The probe for hybridization was obtained by PCR amplification from total testis cDNA library using the primer pair 5′-GGGAGGAGAGGGAAGTCAGA-3′ and 5′-GGTGGTCCAQCTATCTTACAA-3′. The PCR product, of 780 bp, was purified using a PCR kit (TaKaRa Bio, Japan) according to the protocol provided. The probe was labelled with digoxigenin, according to the protocol provided by Roche (USA). After prehybridization at 65ºC for 1 h in 5 ml Express Hyb™ solution (Clontech), the solution was replaced with fresh Express Hyb™ solution, denatured labelled probe was added, and hybridization carried out at 65ºC overnight, followed by washing with 2× SSC and 0.1% SDS at 65ºC for 10 min three times**,** with 0.1× SSC and 0.5% SDS at 62ºC for 15 min twice, with detection buffer for 5 min, and then with NBT/BCIP overnight.

Results

Cloning and sequencing of the full-length cDNA

The full-length cDNA was successfully cloned using PCR from adult mouse testis (figure 1). The PCR product (780 bp)

was sequenced, and the sequence is identical with the cDNA assembled from ESTs using CAP sequence assembly. The novel gene was submitted to GenBank and was designated as *Cymg1* by the Mouse Genome Nomenclature Committee (MGNC).

Analysis of the sequence of Cymg1

***Conserved splice donor and acceptor dinucleotide sequences*:** The full-length cDNA has 780 bp; GC content is 45.1%. A BLAST search against the mouse chromosome map indicated that *Cymg1* is located at 2G3 of chromosome 2. The gene contains four exons and three introns, and the boundaries between exons and introns are in accordance with the GT-AG rule (table 1).

***cDNA and protein sequences*:** *Cymg1* encodes a 141-amino-acid-residue protein. There is a start codon ATG at nucleotide positions 280 to 282, a stop codon TAA at 703 to 705, and an additional stop codon TGA at 268 to 270 before the start codon ATG. An ACCATGGCC sequence, which is in accordance with the Kozak rule, was found in the start region of the open reading frame (ORF), and a potential polyadenylation signal (AATAAA) at 750 to 754 was found at the 3′ end (figure 2).

Bioinformatics analysis of CYMG1

***Domains of the protein*:** The ORF of *Cymg1* encodes a polypeptide of 141 amino acids with a signal peptide and without a transmembrane region. CYMG1 protein has a theoretical molecular mass of 16.8 kDa and a calculated isoelectric point of 9.0. PSORT WWW Server (http://psort.nibb. ac.jp) analysis showed that there is a 99.9% possibility of locating the protein in cytoplasm. SMART analysis indicated that there is a cystatin domain from amino acid residue 43 to residue 139 (figure 3).

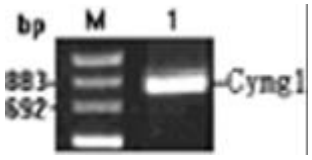

Figure 1. PCR-amplified cDNA of *Cymg1*. M, DNA marker (pUC Mix8); 1, *Cymg1*.

***BLAST homology*:** No gene identical with *Cymg1* was found through BLAST against the nr database. The results showed that the polypeptide encoded by *Cymg1* has 100% identity with that encoded by mouse RIKEN cDNA 1700006C19, and 44% identity and 64% similarity with mouse CRES (figure 4).

***Conserved regions*:** CYMG1 has significant sequence similarity to family 2 cystatins, but lacks some motifs believed to be important for inhibition of cysteine proteases. Comparison (figure 5) of the sequences of CYMG1, CRES and cystatin C showed that, compared with cystatin C, CYMG1 and CRES lack N-terminal glycine and a glutamine_/valine_/glycine (Q_/X_/V_/X_/G) loop segment, which is important for inhibition of cysteine proteases, but retain the C-terminal PW site. In addition, the N-terminal region of cystatin C that is responsible for its tight binding to cysteine proteases (Sutton *et al*. 1999) is poorly conserved in CYMG1 and CRES, suggesting that, like CRES, CYMG1 may also have a function quite different from that of cystatin C.

RT-PCR in different tissues and testis developmental stages

The results (figure 6) show that *Cymg1* (512 bp) is expressed strongly in adult testis but not in other tissues. *G3pdh* expression was seen in all tissues examined. Figure 7 shows that *Cymg1* is expressed in spermatogonial cells. Figure 8 shows that *Cymg1* is expressed at different levels in different developmental stages of mouse testis, while *G3pdh* expression is seen equally in all different developmental stages of mouse testis.

Northern blot

To further investigate the transcript level of *Cymg1* in mouse adult tissues and the transcript length, Northern blot analysis was carried out. The results (figure 9) show that there is a 780-bp transcript only in adult testis.

Table 1. Exon–intron junctions in *Cymg1*.

Exon	Exon size (bp)	5′ splice donor	Intron size (bp)	3′ splice acceptor	Intron
1	152	ATAAAG**gt**a			
2	357	GAGCAG**gt**g	2725	c**ag**CCGTGT	1
3	117	AAAAAG**gt**a	4822	c**ag**ATCACA	2
4	161		1993	t**ag**ATGGTG	3

Uppercase and lowercase letters indicate exon and intron sequences, respectively. The conserved splice donor and acceptor dinucleotide sequences are indicated in bold.

Discussion

It is becoming increasingly clear that multiple genes are responsible for testis development. In this study, a novel testis-specific gene, *Cymg1*, was cloned from a mouse testis library. The encoded amino acid sequence contains a cystatin domain. Northern blot showed that *Cymg1* is 780 bp in length and exclusively expressed in testis and not in other tissues, such as heart, kidney, ovary, epididymis, skeletal muscle, liver and lung. This suggests that *Cymg1* may have tissue-specific functions in testis.

In mouse testis development, at 16 days post-partum there are germ cells in all stages of the first meiotic prophase, and at 20 days post-partum the early type of haploid cells first appear in the seminiferous epithelium (Xing *et al.* 2003). During spermatogenesis, production of sperm occurs within the testicular seminiferous tubules in three phases. First, diploid germ cells, which are primitive spermatogonial cells, self-renew to amplify and produce types A and B spermatogonial cells. Type B spermtogonial cells differentiate into primary spermatocytes. The meiotic division of spermatocytes produces round spermatids. Finally,

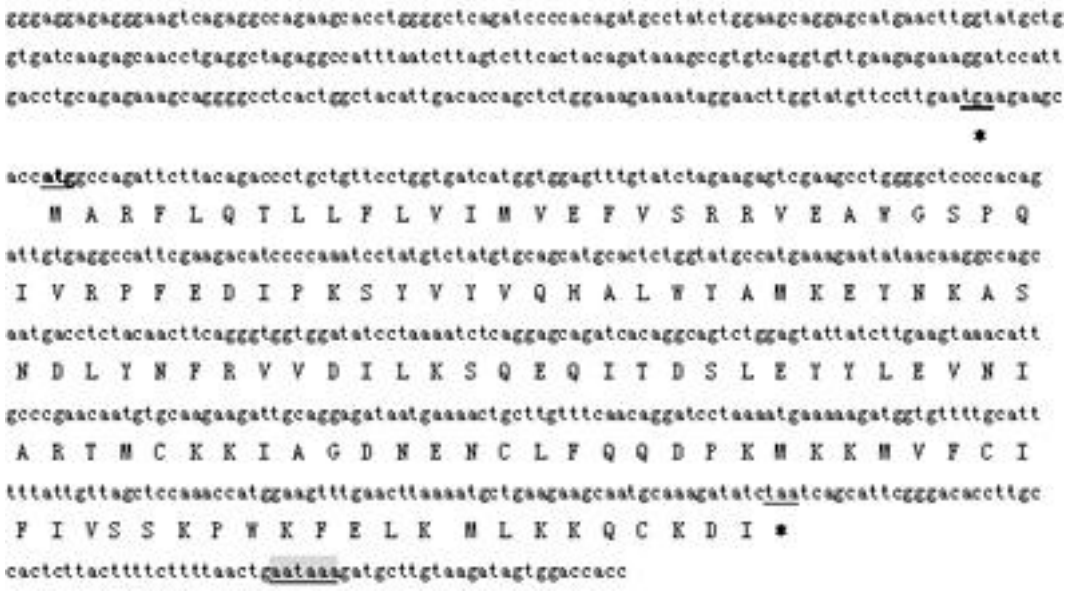

Figure 2. Sequences of *Cymg1* cDNA and of the predicted protein. The start codon is underlined and in bold, polyadenylation signal is underlined and shaded, and the stop codons are underlined and marked with an asterisk.

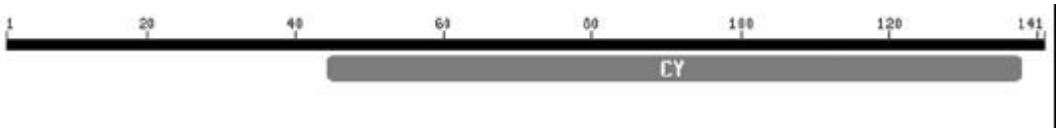

Figure 3. Confidently predicted cystatin domain in CYMG1 protein by SMART analysis.

```
CYMG1   FEDIPKSYVYVQHALNYAMKEYNKASNDLYNFRVVDILKSQEQITDSLEYYLEVNIAKTM 93
        F  I  S   V+  +N+AMKEYNK S D Y F V  IL ++ QITD +EY ++V I+R+
CRES    FGSINISNANVKQCVNFAMKEYNKESEIKYVFLVDKILHAKLQITDRMEYQIDVQISRSN 91

CYMG1   CKKIAGDNENCLFQQDPKMKDMVFCIFIVSSKPWKFELKMLKKQCKDI 141
        CKK   + ENC+ Q+ P+++K + C F+V + PW  E  +L K+CKD+
CRES    CKKPLNNTENCIPQKKPELEKKMSCSFLVGALPWNGEFNLLSKECKDV 139
```

Figure 4. Comparison of sequences of CYMG1 and CRES proteins.

after a series of biochemical and morphological changes, spermatids elongate to form spermatozoa (Luk *et al*. 2003). In mouse, sex maturation begins early, the earliest sperms appear at 5 weeks post-partum, and sex maturation takes place by 6–8 weeks post-partum. In the present study, results of RT-PCR in mouse spermatogonial cells, different developmental stages of testis, and other tissues showed that *Cymg1* is expressed in spermatogonial cells and specifically in adult testis. These results were in accordance with the results of Northern blot analysis that also showed that *Cymg1* is specifically expressed in adult mouse testis. In addition, the RT-PCR results also showed that *Cymg1* expression level varies in the different developmental stages of mouse testis development: low in postnatal first week, steadily increasing in postnatal second to fifth week, highest in postnatal seventh week, and then at postnatal fifth week level in postnatal 13 to 57 weeks. These results indicate that *Cymg1* may play important roles in mouse spermatogenesis and sex maturation.

Studies have shown that *Cres* is primarily expressed in the reproductive tract, is cell specific, and is present in stage-specific germ cells in the testis (Cornwall *et al*. 1992; Cornwall and Hann 1995). CRES protein is present in tissues and luminal fluid of the proximal epididymis as well as in the sperm acrosome, suggesting putative roles in sperm maturation and fertilization (Cornwall and Hann

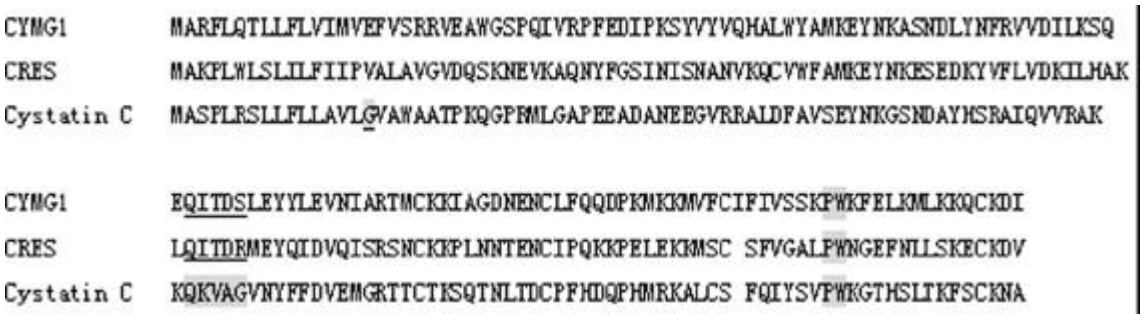

Figure 5. Sequence alignment of CYMG1, CRES and cystatin C proteins. The three highly conserved regions important for inhibition of cysteine proteases are shaded and underlined.

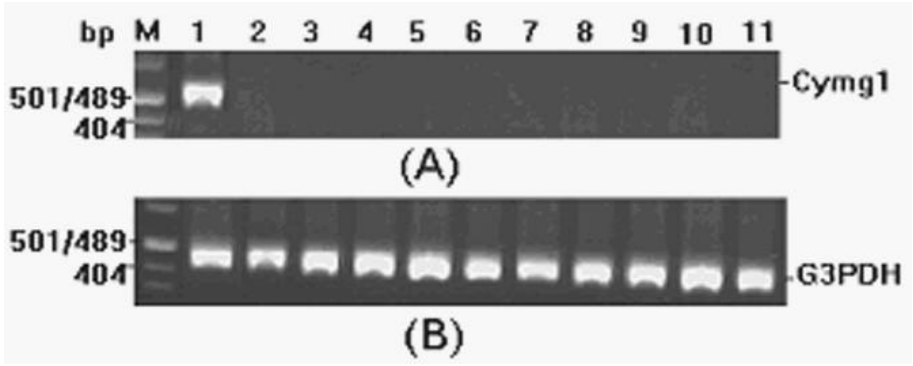

Figure 6. RT-PCR analysis of *Cymg1* expression in mouse tissues. Panel A: M, marker (pUC Mix8); 1, testis; 2, epididymis; 3, ovary; 4, kidney; 5, heart; 6, liver; 7, skeletal muscle; 8, spleen; 9, lung; 10, cerebellum; 11, cerebrum. Panel B: *G3pdh* expression.

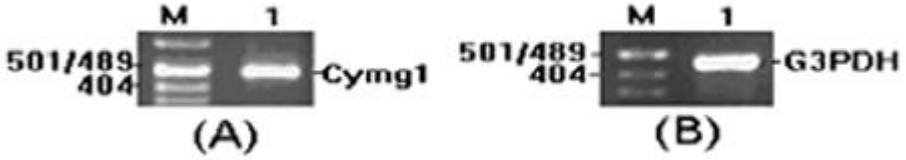

Figure 7. RT-PCR analysis of *Cymg1* expression in CRL-2053 spermatogonial cells. Panel A: M, Marker (pUC Mix8); 1, spermatogonial cells. Panel B shows *G3pdh* expression in the same cells.

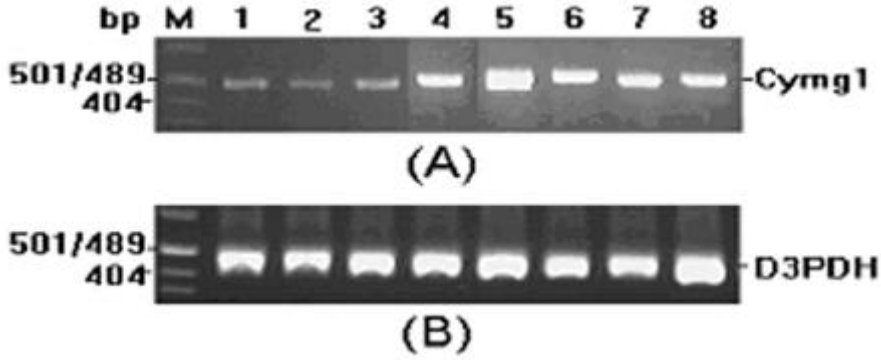

Figure 8. RT-PCR analysis of *Cymg1* expression in different developmental stages of mouse testis. Panel A: M, marker (pUC Mix8); 1 to 8, PCR product from testis of 1, 2, 3, 5, 7, 13, 26 and 57 weeks. Panel B shows *G3pdh* expression in the same tissues.

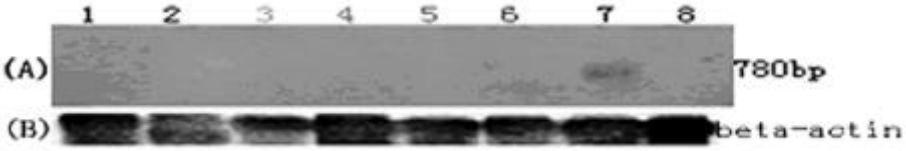

Figure 9. Northern blot analysis of *Cymg1* expression in mouse tissues. Panel A: 1, heart; 2, kidney; 3, skeletal muscle; 4, liver; 5, lung; 6, epididymis; 7, testis; 8, ovary. Panel B shows beta-actin expression in the same tissues.

1995; Syntin and Cornwall 1999). *Cres* mRNA and protein in the anterior pituitary gland gonadotroph cells, and in particular colocalization of CRES with luteinizing hormone beta protein in the secretory pathway were studied (Sutton *et al*. 1999). *Cres* mRNA is present in the corpora lutea of the ovary as well. These data suggested that CRES may perform highly cell-specific and regulative functions that are unlike those of typical cystatins (Cornwall and Hsia 2003). Like *Cres*, the full-length *Cymg1* cDNA includes an ORF that encodes a 141-amino-acid-residue protein with a cysteine protease inhibitor domain but without critical consensus sites important for cysteine protease inhibition. In addition, our present results indicate that CYMG1 may have important roles in mouse spermatogenesis and sex maturation.

In summary, we have cloned a novel gene, *Cymg1*, which is specifically expressed in different developmental stages of mouse testis. This new gene may be a new member of the Cres subgroup of family 2 cystatins and it may play important roles in spermatogenesis and sex maturation.

Acknowledgements

This work was supported by grants from the National Nature Science Foundation of China (No. 30371493) and the Major State Basic Research Development Program of China (973 program) (No. G1999055901).

References

Adams M. D., Kerlavage A. R., Fleischmann R. F., Bult C., Lee N., Kirkness E. *et al*. 1995 Initial assessment of human gene diversity and expression patterns based upon 63 million nucleotides of cDNA sequences. *Nature* **377** (suppl.), 2–174.

Bourrouillou G., Dastugue N. and Colombies P. 1985 Chromosome studies in 952 infertile males with a sperm count below 10 million/ml. *Hum. Genet*. **71**, 366–367.

Cornwall G. A. and Hann S. R. 1995 Transient appearance of CRES protein during spermatogenesis and caput epididymal sperm maturation. *Mol. Reprod. Dev*. **41**, 37–46.

Cornwall G. A. and Hsia N. 2003 A new member of the CRES subgroup of family 2 cystatins. *Mol. Cell. Endocrinol*. **200**, 1–8.

Cornwall G. A., Orgebin–Crist M. C. and Hann S. R. 1992 The CRES gene: a unique testis-regulated gene related to the cystatin family is highly restricted in its expression to the proximal region of the mouse epididymis. *Mol. Endocrinol*. **6**, 1653–1664.

Gill R. W. and Sanseau P. 2000 Rapid in silico cloning of genes using expression sequence tags (ESTs). *Biotechnol. Annu. Rev*. **5**, 25–44.

Luk J. M., Mok B. W., Shum C. K., Yeung W. S., Tam P. C., Tse J. Y. *et al*. 2003 Identification of novel genes expressed during spermatogenesis in stage-synchronized rat testes by differential display. *Biochem. Biophys. Res. Commun*. **307**, 782–790.

Pruitt K. D., Katz K. S., Sicotte H. and Maglott D. R. 2000 Introducing RefSeq and LocusLingk: Curated human genome resources at the NCBI. *Trends Genet*. **16**, 44–47.

Qian J., Zhang X. M., Li X. L., Wang J. R., Li W. F., Wang R. *et al*. 2002 Identification of digital differential expression

patterns of a novel human gene (*UBAP1*) by an expressed sequence tag strategy. *Ai Zheng* **21**, 225–228.

Sato M., Murao K., Mizobuchi M. and Takahara J. 1993 Quantitative and sensitive Northern blot hybridization using PCR-generated DNA probes labeled with digoxigenin by nick translation. *Biotechniques* **15**, 880–882.

Stephen F. M., Barry O. D., Simon J. F., Hugh R. B., Guenole S. and Peter P. D. 2003 Digital extractor: analysis of digital differential display output. *Bioinformatics* **12**, 1594–1595.

Sutton H. G., Fusco A. and Cornwall G. A. 1999 Cystatin-related epididymal spermatogenic protein colocalizes with luteinizing hormone-beta protein in mouse anterior pituitary gonadotropes. *Endocrinology* **140**, 2721–2732.

Syntin P. and Cornwall G. A. 1999 Immunolocalization of CRES (cystatin-related epididymal spermatogenic) protein in the acrosomes of mouse spermatozoa. *Biol. Reprod.* **60**, 1542–1552.

Turk V. and Bode W. 1991 The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett*. **285**, 213–219.

Vogt P. H., Edelmann A., Kirsch S., Henegariu O., Hirschmann P., Kiesewetter F. *et al.* 1996 Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet*. **5**, 933–943.

Xing X. W., Li L. Y., Fu J. J., Zhu W. B., Liu G., Liu S. F. *et al.* 2003 Cloning of cDNA of *TSARG4*, a human spermatogenesis related gene. *Acta Biochim. Biophys. Sin*. **35**, 283–288.

Received 27 August 2004