Biochem. J. (1992) 281, 203–210 (Printed in Great Britain)

Structure and expression of the rat epididymal secretory protein I gene

An androgen-regulated member of the lipocalin superfamily with a rare splice donor site

Milena GIROTTI,* Roy JONES,† David C. EMERY,* William CHIA*‡ and Len HALL*§
*Department of Biochemistry, University of Bristol School of Medical Sciences, Bristol BS8 1TD, U.K., and †Department of Molecular Embryology, Institute of Animal Physiology and Genetics Research, Cambridge Research Station, Babraham, Cambridge CB2 4AT, U.K.

The complete rat epididymal secretory protein I (ESP I) gene was isolated from a genomic library constructed in bacteriophage λ Charon 4A. The complete nucleotide sequence of the gene and its immediate 5′ and 3′ flanking sequences were determined. Interesting features include the presence of a rare, but functional, splice donor site (...GC) and the presence of a putative androgen-receptor-binding element. A detailed analysis of ESP I regulation was carried out after castration and subsequent testosterone treatment, demonstrating the requirement for androgens. Efferent-duct ligation and cryptorchidism, on the other hand, had no effect on the steady-state concentrations of ESP I transcripts. Comparison of the exon/intron organization of the ESP I gene with those of members of the lipocalin superfamily provides strong support for a common ancestral origin.

INTRODUCTION

Following spermatogenesis in the testis, mammalian spermatozoa pass into the epididymis, an androgen-dependent organ composed of a single long convoluted tubule. During their passage along the epididymal duct, spermatozoa undergo a complex variety of maturational changes to their cell surface and intracellular organelles, which culminate in the acquisition of forward motility and the ability to recognize and penetrate the zona pellucida of the egg (see Brooks, 1981*a*; Eddy, 1988). It is well established that this functional maturation of the spermatozoa is dependent on sustained androgenic support to the epididymis, and that many of the maturational events are mediated by androgen-regulated epididymal, secretory proteins that either associate with the spermatozoal membrane (Brooks & Tiver, 1984; Smith *et al.*, 1986; Vernon *et al.*, 1987) or mediate changes in the glycosylation patterns of spermatozoal integral membrane proteins (Voglmayr *et al.*, 1985; Hamilton *et al.*, 1986).

One rat epididymis-specific protein, which appears to exist in two forms and has been studied by a number of groups, has been variously called B and C protein (Brooks, 1981*b*), 18.5 kDa protein (Brooks *et al.*, 1986), 18–19 kDa protein (Hamilton, 1981) and 18 kDa protein (Moore *et al.*, 1990). As this is now one of a growing number of partially characterized epididymal proteins (several of similar size), we have named it 'epididymal secretory protein I' (ESP I) to avoid confusion.

Many studies have demonstrated that rat ESP I is an androgen-regulated epididymis-specific secretory protein that is a major component of the epididymal fluid; the two forms (B and C) probably representing different levels of post-translational modification (Brooks *et al.*, 1986). Early work suggested that this protein possessed α-lactalbumin-like activity *in vitro* (Hamilton, 1981), but this is unlikely to be of any physiological significance (Holpert & Cooper, 1990; Moore *et al.*, 1990). Sequenced ESP I cDNA clones are available (Brooks *et al.*, 1986; Walker *et al.*, 1990; Moore *et al.*, 1990) and have been used to establish that the transcription of this gene is androgen-regulated, epididymis-specific and confined to the caput region of the epididymis (Brooks *et al.*, 1986; Brooks, 1987*a*; Walker *et al.*, 1990). Furthermore the deduced amino acid sequence has suggested that this protein may belong to the lipocalin superfamily (Ali & Clark, 1988; Godovac-Zimmermann, 1988).

We now describe the isolation and complete nucleotide sequence of the rat ESP I gene and its immediate flanking regions. Features of interest include the presence of a rare, but functional, splice donor site (...GC) and the presence of a putative androgen-receptor-binding element in the 5′ flanking region. A more detailed analysis of the regulation of ESP I transcription was carried out after castration and subsequent testosterone treatment, efferent-duct ligation and cryptorchidism. Finally, comparison of the exon/intron organization of ESP I with that of other members of the lipocalin superfamily provides further strong evidence for a common ancestral origin.

MATERIALS AND METHODS

Materials

Restriction endonucleases, avian-myeloblastosis-virus reverse transcriptase, polynucleotide kinase, RNAase-free DNAase, alkaline phosphatase and deoxynucleotides were obtained from Pharmacia, Milton Keynes, Bucks., U.K. Klenow fragment of DNA polymerase I, T4 DNA ligase and Hybond-N transfer membrane were purchased from Amersham International, Amersham, Bucks., U.K. [α-^{32}P]dATP (> 800 Ci/mmol) and [γ-^{32}P]ATP (> 3000 Ci/mmol) were obtained from Du Pont–NEN, Stevenage, Herts., U.K. Proteinase K was from Boehringer, Lewes, East Sussex, U.K., low-melting-point agarose was from

Abbreviation used: ESP I, epididymal secretory protein I.

‡ Present address: Institute of Molecular and Cell Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 0511.

§ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDJB Nucleotide Sequence Databases under the accession number X59831.

Gibco–BRL, Paisley, Scotland, U.K., and Ampli*Taq* from was Perkin–Elmer/Cetus, Beaconsfield, Bucks., U.K. All other chemicals were of AnalaR grade from BDH Chemicals, Poole, Dorset, U.K., or the purest grade available.

Recombinant plasmids

The isolation and characterization of prE23, a recombinant plasmid containing a cDNA copy of rat ESP I, has been described previously (Walker *et al.*, 1990; Moore *et al.*, 1990). The recombinant plasmid pAM91, containing a mouse actin cDNA insert, was obtained from Humphries *et al.* (1981).

Isolation of the ESP I gene from a rat genomic library

A genomic DNA library containing a partial *Hae*III digest of adult female Sprague–Dawley rat liver genomic DNA, cloned by using *Eco*RI linkers into bacteriophage λ Charon 4A, was obtained from Clontech, Cambridge BioScience, Cambridge, U.K. Recombinant bacteriophage containing sequences complementary to rat ESP I mRNA were then identified by screening approx. 750000 independent plaques (Benton & Davis, 1977) with a 700 bp fragment of rat ESP I cDNA, isolated from plasmid prE23, and labelled with ^{32}P to high specific radioactivity by random priming (Feinberg & Vogelstein, 1984). Five independent strong signals were obtained on the primary screen and each was plaque-purified through four successive rounds of screening. One, designated λrESP-I, was used in all further analyses.

DNA sequence analysis

Suitable restriction fragments of λrESP-I were gel-purified and subcloned into M13 mp18 and M13 mp19 vectors. DNA sequence analysis was carried out on a Du Pont Genesis 2000 automated sequencer with the use of fluorescently labelled dideoxynucleotides. Sequencing primers (20-mers) were synthesized on a Du Pont Coder 300 DNA synthesizer with the use of phosphoramidite chemistry, and were used without further purification. Sequencing was complete on both DNA strands and across all restriction sites.

Determination of transcriptional start site

The ESP I transcriptional start site was determined by primer extension as follows. A synthetic oligonucleotide primer (5′-CCCAGCAAGGCGAAGGGCAT-3′) was synthesized, complementary to nucleotide sequence (positions 57–76) encoding part of the signal peptide (amino acid residues −18 to −12) of the ESP I gene, and 5′-end-labelled with the use of polynucleotide kinase and [γ-^{32}P]ATP. A 15 ng portion of this labelled primer was added to 400 ng of rat epididymal polyadenylated RNA in 10 μl of 50 mM-Tris/HCl buffer, pH 8.3, containing 50 mM-NaCl and 9 mM-$MgCl_2$ and heated at 100 °C for 1 min to minimize secondary structure. It was then incubated at 65 °C for 10 min and finally allowed to cool slowly to 37 °C in a beaker of hot water, over a period of 2 h. The volume was increased to 25 μl by the addition of suitable stock solutions to achieve a final composition of 50 mM-Tris/HCl buffer, pH 8.3, 10 mM-$MgCl_2$, 5 mM-dithiothreitol, 25 μM-spermidine and 250 μM with respect to each of the four dNTPs. Avian-myeloblastosis-virus reverse transcriptase (10 units) was added and the mixture was incubated at 37 °C for 15 min. The reaction was then terminated by the addition of 150 μl of 25 mM-EDTA solution, and the extended primer-template was purified and concentrated by ethanol precipitation in the presence of 5 μg of *Escherichia coli* tRNA. Finally, the extended primer was redissolved in 10 μl of 80% (v/v) formamide containing 1 mM-EDTA, 10 mM-NaOH and 10 mg of Bromophenol Blue/ml and analysed on a urea/8% (w/v) polyacrylamide sequencing gel. A sequencing ladder corresponding to the same region of the gene, with the same primer, was electrophoresed alongside.

Castration, testosterone treatment, efferent-duct ligation and cryptorchidism of rats

Adult male Wistar rats were bilaterally castrated through a scrotal incision and the epididymides were replaced inside the tunica vaginalis. After 14 days animals were killed by cervical dislocation, and the epididymides were removed and trimmed free of fat and connective tissue. The tissues were snap-frozen on solid CO_2 and then stored in liquid N_2.

A second group of animals was castrated, and then treated 14 days later with 1 mg of testosterone propionate per day (10 mg/ml solution in sesame-seed oil) for 1, 2, 4 or 7 days. Epididymides were removed and frozen as described above.

In efferent-duct-ligated rats the efferent ducts were ligated close to the extra-testicular rete testis with sterile braided silk to interrupt the connection between testis and epididymis, before replacement of the two organs in the scrotum. Animals were killed 2, 14 and 28 days after efferent-duct ligation, and the epididymides were removed, dissected into different regions (caput plus corpus, and cauda) according to the scheme of Hinton *et al.* (1979), and frozen as above.

In cryptorchidized rats the testis and epididymis were displaced into the abdominal cavity and retained therein by ligatures on the inguinal canal region. Rats were killed 14 days later and epididymides were collected and frozen as above.

RNA isolation and Northern-blot analysis

Frozen rat epididymal tissue (0.5–1 g), was pulverized in liquid N_2 and then homogenized in 15 ml of 10 mM-Tris/HCl buffer, pH 7.6, containing 1 mM-EDTA, 2% (w/v) SDS and 0.5 mg of proteinase K/ml. After incubation at 45 °C for 1 h the homogenate was made 0.2 M with respect to NaCl and extracted with an equal volume of phenol/chloroform (1:1, v/v). Total nucleic acid was precipitated from the aqueous phase with 2.5 vol. of ethanol at −20 °C and re-precipitated at least twice to remove residual phenol. The nucleic acid was redissolved in 10 mM-Tris/HCl buffer, pH 7.4, containing 10 mM-NaCl and 10 mM-magnesium acetate, and incubated with 2 μg of RNAase-free DNAase/ml at room temperature for 20 min. After addition of SDS to a final concentration of 2% (w/v) and proteinase K to 1 mg/ml, the incubation was continued at 37 °C for a further 20 min. The solution was then re-extracted with phenol/chloroform and precipitated with ethanol several times, as before. Finally, the RNA was redissolved in double-distilled water, and the concentration was determined from its absorbance at 258 nm.

Total RNA (15 μg) was fractionated by electrophoresis through a 1.1% (w/v) agarose gel containing formaldehyde, blotted on to a Hybond-N nylon membrane, prehybridized and hybridized as described previously (Walker *et al.*, 1990), with random-primed (Feinberg & Vogelstein, 1984) ^{32}P-labelled prE23 cDNA insert as a probe. All Northern blots were reprobed with a cloned mouse actin cDNA (Humphries *et al.*, 1981) to confirm equivalent track loadings and the integrity of the RNA preparations.

Amplification of ESP I sequences by PCR

Reaction mixtures contained 1 μg of purified rat genomic DNA (or a few nanograms of rat total epididymal cDNA) in 50 μl of 10 mM-Tris/HCl buffer, pH 8.3, containing 50 mM-KCl, 1.5 mM-$MgCl_2$, 200 μM with respect to each dNTP, 1 μM ESP I-specific primers, 0.01% (w/v) gelatin and 1 unit of Ampli*Taq* DNA polymerase. Reactions were subjected to 30 amplification

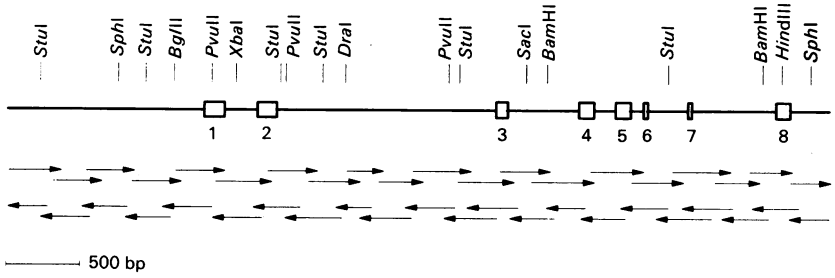

Fig. 1. Genomic organization and sequencing strategy for the rat ESP I gene

Arrows indicate the direction and extent of individual sequencing runs.

cycles (each 94 °C for 90 s, 60 °C for 90 s and 72 °C for 120 s), then fractionated on a low-melting-point agarose gel. The required PCR products were excised and purified, phosphorylated with T4 polynucleotide kinase, ligated into *Sma*I-linearized dephosphorylated pUC18 vector DNA, and used to transform competent *E. coli* TG2 *recA*$^-$ cells. Recombinant plasmid DNA was prepared and subjected to automated DNA sequence analysis as above.

RESULTS AND DISCUSSION

Structure of the rat ESP I gene

A rat genomic library in bacteriophage λ Charon 4A (750000 independent plaques) was screened with a radioactively labelled ESP I cDNA probe [derived from recombinant plasmid prE23 (Walker *et al.*, 1990; Moore *et al.*, 1990)]. Five strong signals were obtained on the primary screen and each was plaque-purified through four successive rounds of screening. Subsequent restriction analysis of two of these independently isolated recombinants revealed them both to have indistinguishable restriction maps (results not shown), suggesting that they were derived from a single initial recombinant DNA molecule. All subsequent work was carried out on one of these clones, designated λrESP-I.

Preliminary restriction analysis of λrESP-I DNA with a series of hexanucleotide-specific endonucleases revealed that this recombinant bacteriophage contained about 12–13 kb of rat genomic sequence. When this was mapped by Southern-blot analysis with cDNA probes, the ESP I exons were found to be located within two *Sac*I fragments (4300 bp and 4800 bp) and two *Bam*HI fragments (1450 bp and 4000 bp). These four fragments were subsequently subcloned into bacteriophage M13 and subjected to DNA sequence analysis on the Du Pont Genesis 2000 automated sequencer with the use of synthetic DNA primers (Fig. 1). Comparison of the resulting DNA sequence (Fig. 2) with that of rat ESP I cDNA (Brooks *et al.*, 1986; Moore *et al.*, 1990) revealed the presence of eight exons and showed no discrepancies in those regions common to genomic and cloned cDNA sequences, except for the previously reported putative cloning artifact at the 5′-end of the Brooks sequence (see Moore *et al.*, 1990).

Analysis of the nucleotide sequence of the rat ESP I gene for conserved sequences of potential regulatory or functional significance revealed a number of features in common with other mammalian genes. We have established, by primer extension of a synthetic oligonucleotide annealed to rat epididymal polyadenylated RNA (Fig. 3), that transcription (+1) was initiated with, or very close to, a uracil residue, 45 bases from the putative initiating methionine codon in exon 1 (see Fig. 2). This was preceded by an A+T-rich region, TATAAA (centred at position −27), typical of that found within the promoter region of most eukaryotic genes. This in turn was preceded by the sequence CCAAT (centred at position −73), a feature (the so-called CAAT box) found associated with many highly expressed genes. At a more distant region from the site of transcriptional initiation can be found the sequence GGATTACAGTGTTCT (positions −477 to −463), which shows a high degree of similarity to the consensus sequence of a number of steroid-hormone-responsive elements (reviewed by Beato, 1989). Such elements are typically 15 bp pseudo-palindromes, comprising 6 bp 'arms' separated by a 3 bp 'spacer', which are recognized by the DNA-binding domain of the appropriate receptor molecule.

At the 3′-end, cleavage and polyadenylation of the ESP I transcript occurs following a C residue (position 4013), 17 bp downstream from a polyadenylation/cleavage signal (AATAAA; positions 3992–3997).

Identification of a rare splice site in the rat ESP I gene

In common with almost all other eukaryotic genes, the dinucleotides GT and AG occur at the 5′-end and 3′-end respectively of all but one of the introns of the rat ESP I gene. However, intron 3 has a rare GC splice donor site at its 5′-end (position 2080). Although this dinucleotide has previously been identified within the mouse αA-crystallin gene (King & Piatigorsky, 1983), and found to be accurately spliced (Lear *et al.*, 1990), its rarity is such that its authenticity required further confirmation in the case of the rat ESP I gene, particularly when only a single independent genomic clone was available for sequence analysis.

Firstly, genomic analysis by Southern blotting suggested that the rat ESP I gene was present as a single-copy gene with no evidence of pseudogenes (results not shown). It therefore seemed unlikely that the sequenced genomic clone represented a non-functional gene.

Secondly, several independent experiments were carried out in which the exon 3/intron 3 boundary region was amplified by PCR, with the use of rat genomic DNA (independently prepared from several different animals as well as from established rat cell lines) and primers from within exon 3 and intron 3 (or exon 3 and exon 4). The resulting PCR fragments were then cloned into plasmid pUC18, and the DNA sequences were determined. In all cases the 5′ splice site of intron 3 was found to be GC.

Thirdly, Northern-blot analysis of rat epididymal RNA pooled from several animals and probed with rat ESP I cDNA (see Fig. 4*a* for example) normally results in the identification of two differently sized transcripts: a major species of 850 nucleotide residues and a larger minor species of about 2000 nucleotide residues. Although we do not yet know the origin of the larger species, another epididymal secretory protein (acidic epididymal glycoprotein) has been shown to exhibit allelic polymorphism,

```
   -1350                                    -1300                                    -1250
GAAATAGGAACGGCGAAGTGACTGCTCACATTTTGTCTGTACACAGGAAGCAGAGAGCCACAGGGACTTTCTAAGATACAGATCCTTACTCCTGGTTATACAGCCCCCCACAAAACTTTT

                                  -1200                                    -1150
GGTGCCAGAGGTCCCAGAGCTTACTTTATTGGACTTCAGGGATCACTTCCAAATCACAAATTCTAGCATGAGAGTGTCAGCATCAGGACAGCTCAGTTCTGAACAGACCCAACATTCGTT

            -1100                                    -1050                                    -1000
TGAGGTTTTCAGAGAGGCCTGGCTGCTACTCTGGGGCTGGCTGAGGTGGGAAGGAGTCTGGCATAGCCCTAGTGGTCGCCTGACTTGATCTGGGCTACCTCAGAACAGCCGGTTGACTGG

                                  -950                                     -900
TTTTAAGAAGAGCTGGGAACACTGGCTTCTGTCAGGAAGAAAGAGGGTGCATCGGCCCACAAGCTGAGCTAGAAGATCCAGTGGGCCGTCACCTCTGTCTGCATTAGATGTCCCTTTCTA

                         -850                                     -800
CATTCTTATAGCCCTGAATCATTTTTATAGTCCCTAACTGTTTGCCATGCCTTCTCCCCGGGTCCTCTCTCTCTCCTTATAGTTTGAACACTGGCTGTAGTTGACAATACTCCCCTAT

    -750                                     -700                                     -650
CCTTGTCTGTGCCCTGAAGGATCGGGACCTGAGCTGCCCCTAGGGTGTGATATCACAAGAGGCTGTCTCTTTGCTGAGGACTCCAGGTGGCCTGGCCCTACACACTGAGCTGGAAGCTGG

                                  -600                                     -550
AAATTGTGGTCTCTGCCCATAAAGACAAATTTGTGGGCAAGATAATGCCAGGGCAACAGTGCTGGCTTCAGCCCGGGCATGCTGGCTGCTCATCTGGTGACCCTTGGCTGACTGCCATGG

            -500                                     -450                                     -400
CTTCTTCAAAGTACTCTGAAAGACAGTGGTGGGGTTGGAGGGATTACAGTGTTCTAACCCACTCCTAAGAGTCTCAGTCTGCTGGACATAGAGCTGAACAGATTCAGTCATGCATGGCTC

                                  -350                                     -300
CAGTGCTCTTAGACCTGCCCTGCCATCTCCAGGCCTCTGATGCCAGGGTCACCTGCTACGGCTGCCACAGTTGTGCCCCTATCAGGTCTTTAGCTGCTCTGTGCCTATGACCAAAGCCTG

                         -250                                     -200
CCAAGGGGCCTTGGTGGAACCGGCTCTGTCACCCACATGTATCTGAAAAGAATCAGCTCAATCACTGAGTGGATGGACAGATGCATACATGAGTAGACACCCAGAGGCAAGATCTGTGTA

    -150                                     -100                                     -50
CCCTGGAGGCGTTCCCCCGAGGCCCTGCTATCTAGCAGGTGGGTGATCCACTTGGCTGGCTTACAGTGTGCTTCCAACCCACCCAATGAGAACCTTTATGCCTATTCCAAGTATCCACGG

                                  +1<- exon 1                                  50
GGAAGGATATAAAGTCCAGAGGCTCTGAGCCAAAGGCTGTGGCTTCTGGTTGGGGAGCCAGTGTGTCCTATTGCTATCCAT ATG GAG AAT ATC ATG CCC TTC GCC TTG CT
                                                                                Met Glu Asn Ile Met Pro Phe Ala Leu Le
                                                                                -22     -20
                                  100                                          exon 1 ->        150
G GGA CTG TGT GTG GGG CTG GCA GCT GGC ACA GAG GGT GCA GTG GTG AAG GAC TTC GAC ATA AGC AAG GTAGTCACAGTTACAGTGGGCCAGGGTCAG
u Gly Leu Cys Val Gly Leu Ala Ala Gly Thr Glu Gly Ala Val Val Lys Asp Phe Asp Ile Ser Lys
         -10                                  -1  +1                                  10
                         200                                      250
TGCTTTTTATGGGGATTGGGCCCTGCTGAGATTCAATCTTCAATGCTCTACCGTGTGGGGAACCCTTTGCTCCTAGACCTTTCTAGACTGTTGCATCCCTTGGAAAGGAGGGGGAGAGAA

       300                                      350          <- exon 2
TGATTCCCGGGTCCCAGGTCCTGAATGATGGCACACCCGGGCAGTTCCTTTATCTACTCCCTACCCCCAG TTT TTA GGC TTC TGG TAT GAG ATT GCC TTT GCC TCC A
                                                             Phe Leu Gly Phe Trp Tyr Glu Ile Ala Phe Ala Ser L
                                                                                                          20
 400                                                 450
AG ATG GGT ACA CCT GGC TTG GCA CAC AAG GAG GAG AAG ATG GGA GCC ATG GTG GTC GAG CTG AAA GAG AAC CTT CTG GCT CTG ACC ACC A
ys Met Gly Thr Pro Gly Leu Ala His Lys Glu Glu Lys Met Gly Ala Met Val Val Glu Leu Lys Glu Asn Leu Leu Ala Leu Thr Thr T
                         30                                       40                                       50
    exon 2 -> 500                                    550                                              600
CC TAC TAC AG GTGAGTTAGGGCTCTGACTACAGCCATAGAGGGAAGTTGGAGCAGTCGGTACAGTGCTGTAGGCCTTCCCCAGAGATGCCTCTACCCATGTCTGTGCTGCAGCTTC
hr Tyr Tyr Se

                                  650                                      700
AGCTGCAGAAGCGGACGGGGAGGCAACTGTTTCAGATGAGACCAGACATCTTTGGCATCGTGAGTGTCCCATTGTCTATCAAGTCAGTTTAAGGAGTGACCAACTCTATGATCATGTTAG

                         750                                      800
GTGTTTAACCTCAGACTGTCACCCCGGACTCTGAGCAAAGCAAGTGACTAACCAGACAGCCTTAAGGTCCAGTTTCCTAGCTCTAGGGTCATGAGTCATGCCAGCCTCTGACTGACAGAT

    850                                      900                                      950
GTCAAGGCCTGAGAAGTGAGGCAAGGTTGACCTGAGTCATCTTGGAACTCAATCCAGCAGTGTGAAAGTTCGAGTGTTATTGATTGACCAGGAGACAACTTCGAGTTCCACAAGTTGAAG

                                  1000                                     1050
GCACATCCTTCCACAAAGACATGTGTGCATCCCTGTTCTCACAGTGTCCCTTTAAAGTGTGTACTTCTATCCCAGAGTGGCCCTCGGATGTGTGTACTTGTTTCCCTCAGCCATGCATGT

            1100                                     1150                                     1200
CCCTTATCCCCAGGGAACGTGTCCTCAGATGTACATAACTGTGGTTCTCAGAGCGTGTACCCTCAGATATGCATCCTGTCACCTTAATGTGTGTCCCACAAGATGGATGTGCTTGTACCA

                                  1250                                     1300
GAGTACGTCCCTGACTCTCAGAGTACGGGTCCTTCAGACCCGAACATGTCTCTCCACCTCAGTACTGTGTCCTTCAGCCGTGTGTTACTCAGAGACTATGGACCTCACATATGTCTCTTT

                         1350                                     1400
GCTCCTCTGCTCCTAAATATGTATGTCCCTCTACCACCAAGTGTGCTCCTCAGGAACTCTCCCCCCAGATGTGTGTCTCTTCGCTTCAACAAGTCTGTCTCTCAGATGTGTGTCTCCATG

   1450                                     1500                                     1550
TTCCCGGATGGCTTATCCATAGCAGTGTGTCCCTCGGTGGCACTGAGTGGGTCCCTTAGTGCGTTTGTTCCTTCAGACATGAGTGTGACCCCGTTTCTTAGACTATCCCTCAGACAGGTG

                                  1600                                     1650
CGCAAACCGATTCTCAGCACTGGTTAGTAGGTCTTCAGGCGGGATCACTGGGGACATATTTCTACAACCACTTCCTCTTGTTACTCCAGCTCCATCTCCCTGGATAGCCACTGGGTAGGG

            1700                                     1750                                     1800
GTCCTGTCCTCTGGGCTCACAGGCTCAGAAGACATTAGGCCAGCTGTATGTGGCTCGTCTCTGGACTCCTGCGTCACACACCCTGCTCCCTCGGGGCATGTATTATAAGGCCTGGTAGGC
```

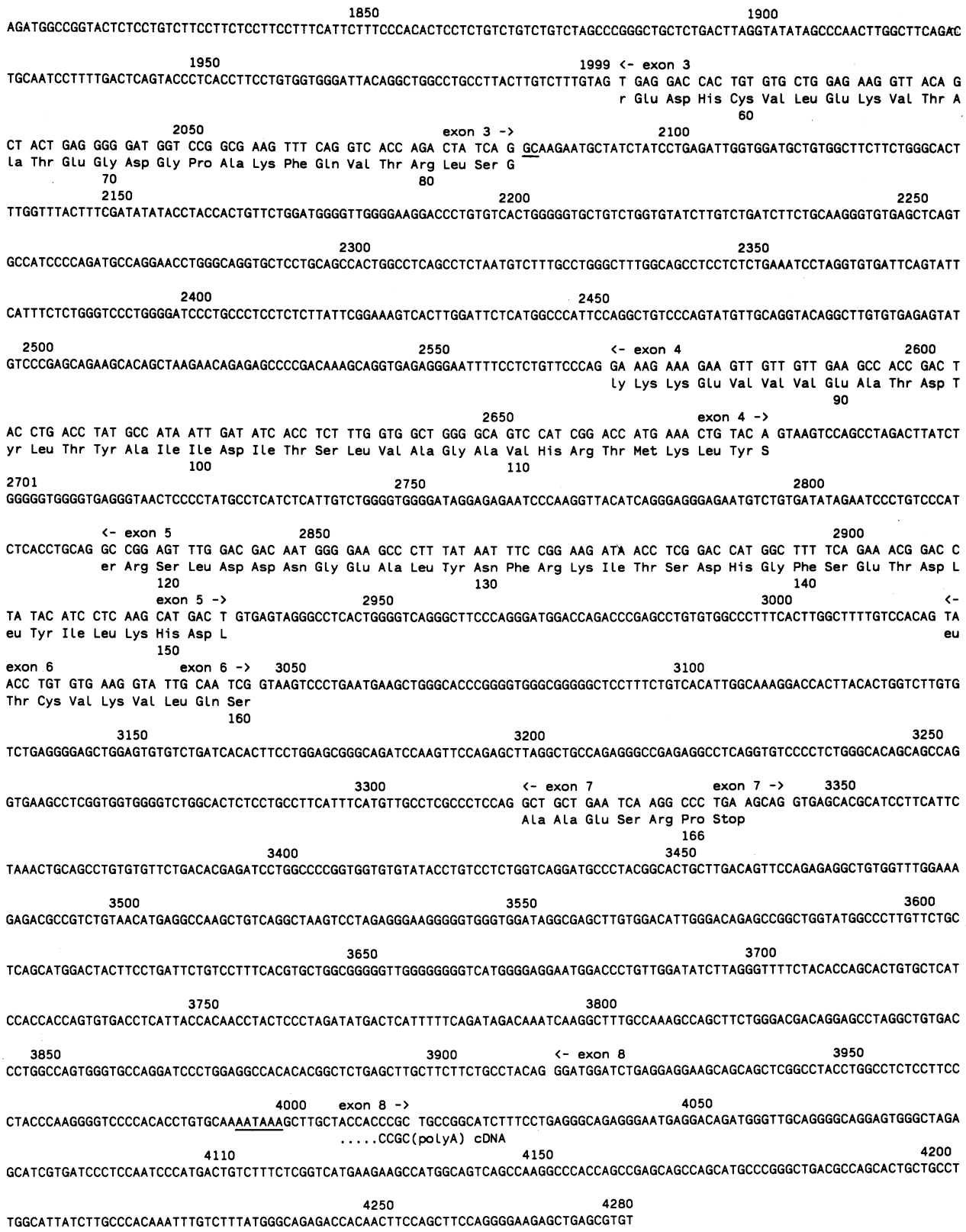

Fig. 2. Complete DNA sequence of the rat ESP I gene and immediate flanking sequences

The sequence is numbered from the transcriptional start site (+1; see Fig. 3). Underlined sequences indicate a putative hormone-receptor-binding site (positions −477 to −463), a putative 'CAAT box' (centred at position −73), 'TATA box' (centred at position −27), a rare GC splice donor site (position 2080) and a polyadenylation/3′ cleavage signal (AATAAA; positions 3992 to 3997).

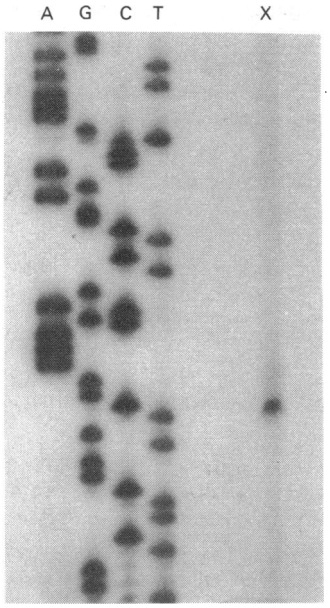

Fig. 3. Mapping of the transcriptional start site of rat ESP I RNA by primer extension

Primer extension of rat ESP I RNA transcripts was performed as described in the Materials and methods section, and the products (X) were resolved by electrophoresis on a urea/6% polyacrylamide sequencing gel. A sequencing ladder (A, G, C and T) was run in parallel with the use of the same oligonucleotide primer.

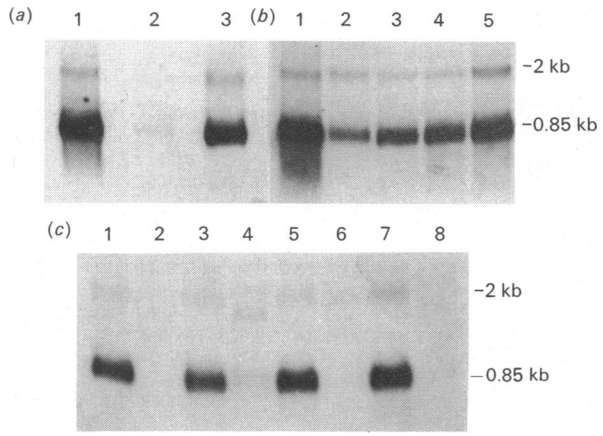

Fig. 4. Effects of castration, testosterone treatment, efferent-duct ligation and cryptorchidism on ESP I steady-state transcript concentrations in the rat epididymis

Epididymal RNA samples (15 μg) were electrophoresed on an agarose gel under denaturing conditions, blotted and probed with a rat ESP I cDNA insert, as described in the Materials and methods section. (*a*) Epididymal RNA isolated from normal (lane 1), 7-day-castrated (lane 2) and 14-day-cryptorchidized (lane 3) rats. (*b*) Epididymal RNA isolated from normal rats (lane 1) or from rats castrated for 14 days then treated with testosterone for 1 day (lane 2), 2 days (lane 3), 4 days (land 4) or 7 days (lane 5). (*c*) Epididymal caput plus corpus RNA (lanes 1, 3, 5 and 7) or cauda RNA (lanes 2, 4, 6 and 8) isolated from normal rats (lanes 1 and 2) or efferent-duct-ligated rats 2 days (lanes 2 and 4), 14 days (lanes 4 and 6) and 28 days (lanes 7 and 8) after ligation.

resulting in differently sized transcripts in different animals (Charest *et al.*, 1988). There was therefore the possibility that the sequenced rat ESP I cDNAs (Brooks *et al.*, 1986; Moore *et al.*, 1990) and the ESP I genomic sequence reported here represented different allelic forms. Indeed, the ESP I genomic sequence does contain an in-phase putative GT splice site within intron 3 (nucleotides 2107/2108) that could result in a slightly larger transcript and translation product. To address this possibility, genomic DNA and total epididymal RNA were isolated from the same animal and the latter was used to direct the synthesis of total cDNA with the use of reverse transcriptase. Both cDNA and genomic DNA were then subjected to PCR amplification, and the resulting fragments were subcloned and sequenced. The sequence of the cDNA product confirmed the site of the exon 3/exon 4 boundary [identical with that reported previously (Brooks *et al.*, 1986; Moore *et al.*, 1990)], and the genomic PCR product from the same tissue sample confirmed that splicing must occur at a GC donor site.

Comparison of the unusual splice donor site within the rat ESP I gene with that of the mouse αA-crystallin gene (King & Piatigorsky, 1983) indicated that the nucleotide residues flanking the GC were also conserved in both genes, i.e. AG/GCAAG. Determination of the hierarchy for 5′ splice-site preference *in vitro* (Lear *et al.*, 1990) has confirmed that the sequence CAAG/GCAAGT (as found in the mouse αA-crystallin gene) is spliced, although at a somewhat reduced efficiency when compared with a consensus test sequence (CCAG/GTAAGT). Nevertheless the efficiency observed for the GC splice site *in vitro* was comparable with, or greater than, that of some naturally occurring GT splice sites (e.g. TCTC/GTAAGT, ACAG/GTATAT, GAAG/GTAAGT). The evidence available would therefore indicate that, despite the rare splice site, the cloned rat ESP I gene would be properly processed, although definitive proof must await transcriptional analysis in transfected cell lines. Such studies would also permit an analysis of the relative efficiency of splicing at this unusual site.

Androgen regulation of ESP I gene expression

It has previously been established that castration leads to a marked decrease in the steady-state concentrations of ESP I transcripts in the epididymis (Brooks *et al.*, 1986; Walker *et al.*, 1990). A more detailed analysis (results not shown) has indicated that ESP I transcript concentrations gradually decrease over a period of a few days following castration to a minimum level (less than 5% of precastration values) after 7–14 days (see Fig. 4*a*). If testosterone is then administered 14 days after castration, steady-state ESP I RNA concentrations gradually increase to pre-castration values over a period of 7 days, as demonstrated by Northern-blot analysis (Fig. 4*b*). Although this result confirms our previous preliminary data (Walker *et al.*, 1990), it differs from the results obtained by Brooks (1987*a*), who reported a restoration of ESP I transcripts to only 30% of precastration steady-state concentrations. However, this difference could well be due to the much longer castration period (16 weeks) used by Brooks, which may have led to depletion of epididymal hormone receptor and/or depletion of ESP-I-expressing cell types.

Effect of efferent-duct ligation on ESP I transcript concentrations

The initial segment and proximal caput epididymidis are most active in absorption of testicular fluid and involute rapidly after efferent-duct ligation. Within 48 h of ligation there is a decrease in wet weight, blood flow, protein synthesis and content of several tissue-specific compounds (reviewed by Cooper, 1986). This effect has been attributed to factors (e.g. growth factors, mitogens) in testicular fluid, since supraphysiological concentrations of testosterone on their own do not maintain the morphology or secretory activity of the epithelial cells in efferent-duct-ligated rats (Fawcett & Hoffer, 1979; Feig *et al.*, 1980). In view of the high level of transcription of the ESP I gene in the

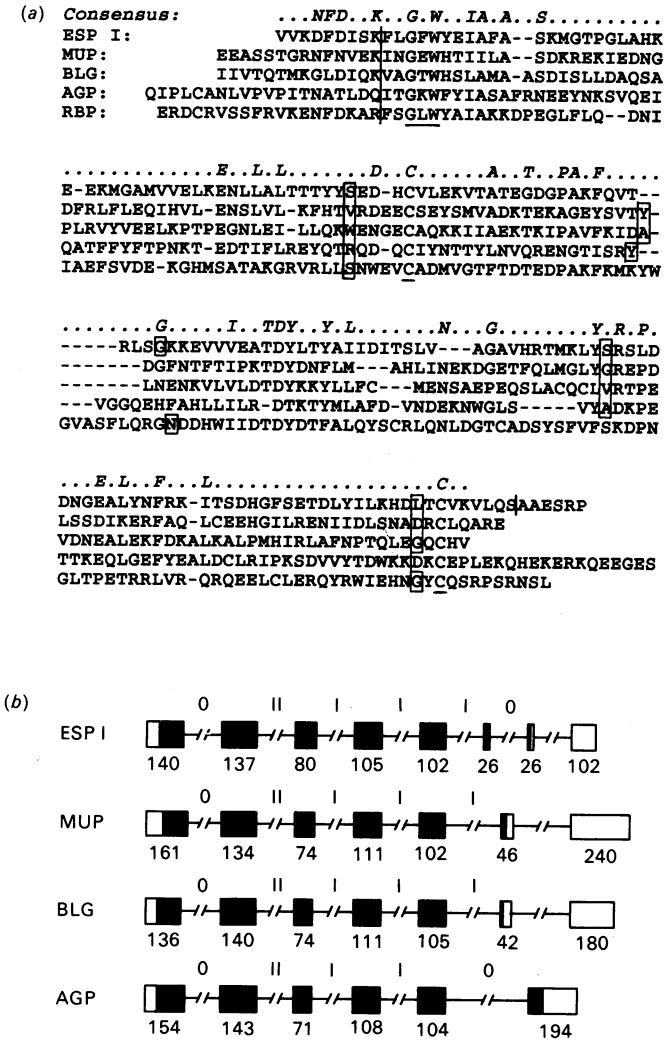

Fig. 5. Comparison of the amino acid sequence and exon/intron organization of ESP I with members of the lipocalin superfamily of secretory proteins

(*a*) The primary amino acid sequences of rat ESP I (Moore *et al.*, 1990), mouse major urinary protein (MUP; Clark *et al.*, 1984), sheep β-lactoglobulin (BLG; Ali & Clark, 1988), human α_1-acid glycoprotein (AGP; Dente *et al.*, 1987) and rat retinol-binding protein (RBP; Laurent *et al.*, 1985) are aligned. Dashes represent gaps inserted to maximize the alignment. Underlined regions indicate the conserved Gly-Xaa-Trp motif and conserved cysteine residues. The positions of introns that interrupt the coding sequences in the corresponding genes are indicated as either a vertical line (where an intron is inserted between codons) or as a box (where an intron is inserted within a codon). (*b*) Exon/intron organization of the above genes (except retinol-binding protein) indicating sizes of individual exons (bp). Exons are shown as boxes; coding regions are black. Intron phasing is represented as 0 (where an intron is inserted between codons), I (where an intron is inserted after the first base in a codon) or II (where an intron is inserted after the second base in a codon).

proximal caput region (Walker *et al.*, 1990), it was decided to examine the influence of efferent-duct ligation on its expression. In this experiment, caput plus corpus and cauda regions of the epididymis were analysed separately. However, Fig. 4(*c*) shows that absence of testicular fluid had little effect on the concentrations of caput ESP I transcripts as detected by Northern-blot analysis. No expression was observed in the cauda, as expected (Walker *et al.*, 1990). Clearly if efferent-duct ligation has an effect on the secretory activity of the proximal caput region, as has been suggested, either this does not include ESP I or else the effect is translational or post-translational.

Effect of body temperature on ESP I transcript concentrations

A further influence on epididymal function is elevated temperature. In scrotal mammals the testis and epididymis are 3–6 °C lower than core body temperature (Setchell, 1978). This has been demonstrated to be important for maintenance of spermatogenesis and sperm maturation/viability in the epididymis in a wide variety of species. At elevated temperature androgen secretion by Leydig cells, production of androgen-binding protein in Sertoli cells and the flow of testicular fluid are all significantly decreased, with important implications for the functional activity of the caput epididymidis (Setchell, 1978). Bedford and co-workers have presented evidence that storage of spermatozoa in the cauda epididymidis is very sensitive to elevated temperature and that this is due to a failure of the water- and ion-transport functions of the principal cells in the epithelium (Foldesly & Bedford, 1982; Wong *et al.*, 1982; Rasweiler & Bedford, 1982). The same workers also reported that the amounts of several androgen-dependent proteins (of molecular mass 18 kDa, 38 kDa and 48 kDa) in cauda epididymidal plasma were greatly diminished after artificial cryptorchidism, suggesting that protein synthesis was also temperature-sensitive (Esponda & Bedford, 1986). As shown in Fig. 4(*a*), the steady-state concentrations of ESP I transcripts were not adversely affected by 14 days of cryptorchidism, suggesting that either the expression of this gene is not adversely affected by the resulting increase in temperature or the effect, if any, is again translational or post-translational.

ESP I is a member of the lipocalin superfamily

Pervaiz & Brew (1985) reported a similarity between the primary amino acid sequences of human serum retinol-binding protein and dolphin and bovine β-lactoglobulins, a finding that was later supported by comparison of the three-dimensional structures of these proteins (Newcomer *et al.*, 1984; Papiz *et al.*, 1986). Since then the number of related secretory proteins in this family has grown to encompass about 20 different members, including bovine β-lactoglobulin, human and rat serum retinol-binding proteins, mouse major urinary protein, rat odorant-binding protein, crustacyanins A and C, human α_1-acid glycoprotein, human apolipoprotein D, human protein HC, human α_2-pregnancy endometrium protein and human complement C8-γ polypeptide, as well as rat epididymal secretory protein I (Ali & Clark, 1988; Godovac-Zimmermann, 1988).

Although statistical analyses indicate that these proteins are evolutionarily related, sequence identity between any pair is usually only about 15–20%. However, analysis of the three-dimensional structures of retinol-binding protein, β-lactoglobulin and two insect bilin-binding proteins has indicated that these proteins possess a central core made up of eight β-structure strands, forming a hollow cone, probably with a hydrophobic interior, to which small hydrophobic ligands may attach non-covalently (e.g. vitamin A, pheromones, biliverdin, porphyrin, cholesteryl esters, odorants etc., depending on the particular family member). In fact, all members of this family contain the sequence Gly-Xaa-Trp, known to interact with retinol in human serum retinol-binding protein. Because of this ability, or in some cases assumed ability, to bind small lipophilic molecules, they are often referred to as the lipocalin superfamily.

However, perhaps the most striking evidence that members of this superfamily of secretory proteins are evolutionarily related comes from comparisons of their genomic organizations. Gene structures are now available for a number of these proteins and it can be seen that the exon/intron arrangements are often very

similar, with several comparably sized analogous exons and similar, if not identical, sites of intron insertion and intron phasing (Fig. 5).

Whilst it has previously been suggested that rat ESP I may also be a member of the lipocalin family on the basis of primary amino acid sequence comparisons (Brooks, 1987*b*), comparison of the ESP I gene sequence with that of the mouse major urinary protein, sheep β-lactoglobulin and human α_1-acid glycoprotein genes (to which the ESP I gene is most closely related) clearly indicates considerable conservation of exon/intron organization (Fig. 5). In particular, exons 1–5 are of similar length in each case. However, exon 6 of the major urinary protein (46 bp) and β-lactoglobulin (42 bp) genes appears to be interrupted by the presence of an additional intron in the ESP I gene to produce two very small exons (each 26 bp). Nevertheless, the overall similarity in genomic organization of ESP I and other members of the lipocalin family clearly supports the hypothesis that they are evolutionarily related. As yet the function of ESP I is unknown. Clearly, the finding of a hydrophobic ligand might help to establish the function of this abundant androgen-regulated secretory protein in the male reproductive tract.

We thank the Medical Research Council for partial support of this work, the Science and Engineering Research Council for a postgraduate studentship for M.G. and the University of Bristol Molecular Recognition Centre (funded by the Science and Engineering Research Council) for automated DNA sequencing and synthesis facilities.

REFERENCES

Ali, S. & Clark, A. J. (1988) J. Mol. Biol. **199**, 415–426.
Beato, M. (1989) Cell **56**, 335–344
Benton, W. D. & Davis, R. W. (1977) Science **196**, 180–182
Brooks, D. E. (1981*a*) Physiol. Rev. **61**, 515–555
Brooks, D. E. (1981*b*) Biol. Reprod. **25**, 1099–1117
Brooks, D. E. (1987*a*) Mol. Cell. Endocrinol. **53**, 59–66
Brooks, D. E. (1987*b*) Biochem. Int. **14**, 235–240
Brooks, D. E. & Tiver, K. (1984) J. Reprod. Fertil. **71**, 249–257
Brooks, D. E., Means, A. R., Wright, E. J., E. J., Singh, S. P. & Tiver, K. K. (1986) J. Biol. Chem. **26**, 4956–4961
Charest, N. J., Joseph, D. R., Wilson, E. M. & French, F. S. (1988) Mol. Endocrinol. **2**, 999–1004
Clark, A. J., Clissold, P. M., Al Shawi, R., Beatie, P. & Bishop, J. (1984) EMBO J. **3**, 1045–1052
Cooper, T. G. (1986) The Epididymis, Sperm Maturation and Fertilization, pp. 33–51, Springer-Verlag, Berlin
Dente, L., Pizza, M. G., Metspalu, A. & Cortese, R. (1987) EMBO J. **6**, 2289–2296
Eddy, E. (1988) in The Physiology of Reproduction, vol. 1 (Knobil, E. & Neill, J., eds.), pp. 27–68, Raven Press, New York
Esponda, P. & Bedford, J. M. (1986) J. Reprod. Fertil. **78**, 505–514
Fawcett, D. W. & Hoffer, A. P. (1979) Biol. Reprod. **20**, 162–181
Feig, L. A., Bellue, A. R., Erickson, N. H. & Klagsbrun, M. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 4774–4778
Feinberg, A. P. & Vogelstein, B. (1984) Anal. Biochem. **137**, 266–267
Foldesly, R. G. & Bedford, J. M. (1982) Biol. Reprod. **26**, 673–682
Godovac-Zimmermann, J. (1988) Trends Biochem. Sci. **13**, 64–66
Hamilton, D. W. (1981) Biol. Reprod. **25**, 385–392
Hamilton, D. W., Wenstrom, J. C. & Moore, A. (1986) Adv. Exp. Med. Biol. **205**, 121–129
Hinton, B. T., Dott, H. M. & Setchell, B. P. (1979) J. Reprod. Fertil. **55**, 167–172
Holpert, M. & Cooper, T. G. (1990) J. Reprod. Fertil. **90**, 503–514
Humphries, S. E., Whittall, R., Minty, A., Buckingham, M. & Williamson, R. (1981) Nucleic Acids Res. **9**, 4895–4908
King, C. R. & Piatigorsky, J. (1983) Cell **32**, 707–712
Laurent, B. C., Nilsson, M. H. L., Bavik, B. O., Jones, T. A., Sundelin, J. & Peterson, P. A. (1985) J. Biol. Chem. **260**, 11476–11480
Lear, A. L., Eperon, L. P., Wheatley, I. M. & Eperon, I. C. (1990) J. Mol. Biol. **211**, 103–115
Moore, A., Hall, L. & Hamilton, D. W. (1990) Biol. Reprod. **43**, 497–506
Newcomer, M. E., Jones, T. A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L. & Peterson, P. A. (1984) EMBO J. **3**, 1451–1454
Papiz, M. Z., Sawer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E. & Kraulis, P. J. (1986) Nature (London) **324**, 383–385
Pervaiz, S. & Brew, K. (1985) Science **228**, 335–337
Rasweiler, J. J. & Bedford, J. M. (1982) Biol. Reprod. **26**, 691–705
Setchell, B. P. (1978) The Mammalian Testis, pp. 90–108, P. Elek, London
Smith, C. A., Hartman, T. D. & Moore, H. M. (1986) J. Reprod. Fertil. **78**, 337–345
Vernon, R. B., Muller, L. H. & Edd, E. M. (1987) J. Androl. **8**, 123–128
Voglmayer, J. K., Sawyer, Jr., R. F. & Dacheux, J. L. (1985) Biol. Reprod. **33**, 165–176
Walker, J. E., Jones, R., Moore, A., Hamilton, D. W. & Hall, L. (1990) Mol. Cell. Endocrinol. **74**, 61–68
Wong, P. Y. D., Au, C. L. & Bedford, J. M. (1982) Biol. Reprod. **26**, 683–689

Received 15 May 1991; accepted 14 June 1991