The EMBO Journal vol.9 no.5 pp.1529–1534, 1990

Zfa is an expressed retroposon derived from an alternative transcript of the *Zfx* gene

Alan Ashworth, Barbara Skene[1], Sally Swift and Robin Lovell-Badge[1]

Chester Beatty Laboratories, Institute of Cancer Research, Fulham Road, London SW3 6JB and [1]National Institute of Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Communicated by R.A.Weiss

***ZFY*, a gene on the Y chromosome encoding a zinc finger protein, has been proposed as a candidate for the human testis determining gene. Sequences related to *ZFY*, called *ZFX*, are present on the X chromosome of a wide range of placental mammals. Unlike most mammals the mouse has four genes homologous to *ZFY*; two on the Y chromosome, *Zfy-1* and *Zfy-2*, an X-linked gene, *Zfx*, and an autosomal gene, *Zfa*. We show here that *Zfa* has arisen recently by retroposition of one of at least three alternatively spliced mRNAs transcribed from the *Zfx* gene. *Zfa* is an unusual retroposon in that it has retained an open reading frame and is expressed, although its function may be limited or altered by the presence of a potentially inactivating mutation in the third of its zinc fingers. This mutation must have occurred at the same time or soon after the retroposition event as it is also present in the *Zfa* gene of *Mus spretus*. Interestingly the third finger of the *M.musculus musculus Zfy-2* gene has also sustained a mutation suggesting that this gene family may be rapidly evolving in mice.**

Key words: Zfa/retroposon/transcript/*Zfx*

Introduction

Transcriptional regulatory proteins in eukaryotes often contain highly conserved motifs involved in DNA binding (Mitchell and Tjian, 1989). One of these, the zinc finger motif, is specified by a characteristic arrangement of cysteines and histidines which are thought to bind Zn^{2+} ions. The co-ordination of Zn^{2+} ions by these residues forms the finger structure which is thought to interact with the major groove of the double helix of DNA (Diakun *et al.*, 1986; Fairall *et al.*, 1986). Since the original description of the zinc finger motif in the *Xenopus* 5S gene transcription factor TFIIIA (Brown *et al.*, 1985; Miller *et al.*, 1985) several genes have been identified which contain various numbers of zinc fingers. These include transcription factors such as TFIIIA and SP1 (Kadonaga *et al.*, 1987), the yeast regulatory genes *ADR1* (Hartshorne *et al.*, 1986) and *SW15* (Nagai *et al.*, 1988), and the *Drosophila* developmental genes *Kruppel* (Rosenberg *et al.*, 1986), *Snail* (Boulay *et al.*, 1987) and *Glass* (Moses *et al.*, 1989). It is likely that these genes act by binding directly to target genes and affecting their transcription (Stanojevic *et al.*, 1989; Triesman and Desplan, 1989).

Sex in mammals is determined by a dominantly acting gene present on the Y chromosome which diverts the differentiation of the genital ridge from the 'default' ovarian pathway to that of the testis (McLaren, 1988). Recently, a gene, *ZFY*, encoding a zinc finger protein has been identified in the sex-determining region of the human Y chromosome (Page *et al.*, 1987) and suggested as a candidate for the human testis determining gene. However, the pattern of expression of this gene in mice (Koopman *et al.*, 1989) is inconsistent with the cell autonomous action proposed for the testis determining gene (Burgoyne *et al.*, 1988). In addition, genetic evidence has been presented that argues against *ZFY* being the primary sex determining gene in humans (Palmer *et al.*, 1989). Nevertheless, the presence of a Y-linked homologue of *ZFY* in all placental mammals analysed suggests that this gene has an important male-specific function. A gene related to *ZFY*, called *ZFX*, is present on the X chromosome of a wide range of placental mammals. Unlike most mammals the mouse has four genes homologous to *ZFY*; two on the Y chromosome, *Zfy-1* and *Zfy-2*, an X-linked gene, *Zfx*, and an autosomal gene, *Zfa*, located on chromosome 10 (Mardon *et al.*, 1989; Nagamine *et al.*, 1989). It is important to determine whether the presence of these additional genes is of any significance in sex determination or in other biological processes. *Zfy-1* and *Zfy-2* appear to have arisen from a recent gene duplication event (A.Ashworth, unpublished results) and both may encode functional proteins, although genetic studies have established that *Zfy-2* is not necessary for testis formation (Mardon *et al.*, 1989; Nagamine *et al.*, 1989). Here we describe the structure, origin and pattern of expression of *Zfa*.

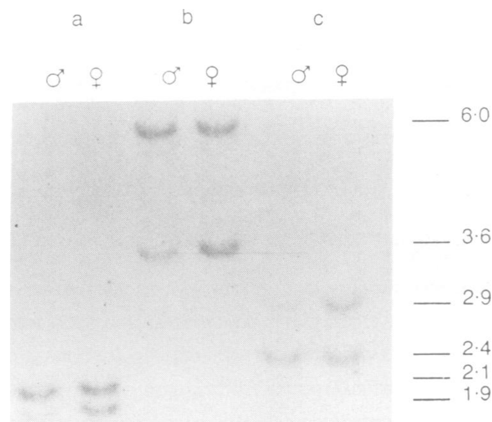

Fig. 1. Southern analysis of genomic DNA hybridized with *Zfa* cDNA. DNA from male or female CBA/Ca mice was digested with *Eco*RI (**a**), *Hin*dIII (**b**) and *Taq*I (**c**). Fragments derived from *Zfx* (1.9, 3.6 and 2.9 kb, respectively) show double intensity in female DNA on hybridization with p705.

© Oxford University Press

Results

Zfa is a retroposon derived from Zfx

We isolated several clones from an adult mouse testis cDNA library by virtue of their hybridization to a human *ZFY* probe (Ashworth *et al.*, 1989). Two of these, p703 and p705, were shown to be derived from the *Zfa* gene by restriction mapping. Hybridization of p705 to genomic DNA detected fragments of both the *Zfa* and *Zfx* genes (Figure 1). *Zfy-1* and *Zfy-2* specific fragments are only detected at reduced stringency of hybridization (not shown). This suggests that *Zfa* is more closely related to *Zfx* than to either of the *Zfy* genes. As *Zfa* is absent from most mammals including rats (Page *et al.*, 1987; A.Ashworth, unpublished observations) it seemed likely that *Zfa* was of recent evolutionary origin and derived from the *Zfx* gene. In order to investigate this

```
CCATGGTTGAAGTTCTTTGCATGCACAGTTCATGTGCTTTCACATTGTTCCTGAACTGCCTTTGTTAGAGGTGGCAGATGAAGCAGATGAAGAAGACTTCCTCCAGTTTGAAAGGATCAC  120
AAATGATGGATTTTGTCCCATTCAAAGCAGGAACAACTAAAATTGCCAGCAACAAGATTAGAGAATGGATGACTCTCCTGCTCTTCCCACCACTGCATTGCCAAGTTATGTGATTCAGTA  240
GAAACAAGTCAGATTAAACTGTGCACACATGAAGACATTTATTTATAATCAGGTGCAGACATGTACTATATGATATGATCCTGTTAAAACATATTTCCTACGGGGACCGGAGCCCGGGCG  360
CAGAGTCGAGTGTGCGCGCAGGGGCGTGTTCGCTCCCCATGGCGGCCATTGTCCTCGTCGCCATGATGAGCACTTGGATCCCAGACGCTTGGGGGCAGCGCCACCTTCCTGCCCAGCCCC  480
CCGCTGCCCCGTGACTGGCTGCTGTTTCCGCTGTGTTTCAGTTGAGCTGCCCCAGGCGGCCAGGGCTGCGTCTTGGCGCCGGAGCCATCTTATCTGTATGGCCGCGGTGGCAGTAGTGGA  600
GCCGGAGACATTTTCTTTGGACTTTATTTTCTCAGAGGTGGATAGACATGGCCGTTAGGGTGACCCTGGTGCGGTGCTTCTCTGCCAATCGTTGCTTTCACGACTCTATGAATAAAGAAA  720
GAAAAGTGATGGGAAAGTTCTCGCCACTCTTTTGACCAATAGGGGGGAAATGGCGTCCTACAAAAGCGGAAGACTAACATATCGGAGGGAAGGTTTTTGTGTTCCTTTCTGTGGAAACAA  840
GTTTCTTGGGAGGCTGCGTTTGAATTTTGGTGCGTGTAACTGTGCCCAACAACCTTGTAAGATGAATTGGGGAGCATCCCCTCCTTTTTGAAACAGGTGTTCTGGCTGATGAGAATTTAA  960
AGGCCATGGATGAAGATGGACTTGAATTACAACAACAAGCACCAAACTCATTTTTTGATGCAACAGGAGCTTGTTGCTCCACACATGGATGGTAGTCAAATTGTTGTGGAAGTTCAAGAA 1080
          M  K  M  D  L  N  Y  N  N  K  H  Q  T  H  F  L  M  Q  Q  E  L  V  A  P  H  M  D  G  S  Q  I  V  V  E  V  Q  E
ACGGTTTATGTGTCAGATGTTGTAGATTCCGACATAATTGTACATAACTATGTTCCTGATGACCCAGATTCAGTTGTAATCCAGGATGTCATTGAGGATGTTGTTATAGAAGATGTCCAG 1200
 T  V  Y  V  S  D  V  V  D  S  D  I  I  V  H  N  Y  V  P  D  D  P  D  S  V  V  I  Q  D  V  I  E  D  V  V  I  E  D  V  Q
TGTACAGACATCATGGACGAAGCAGATGTATCGGAAACAGTCATCATCCCTGAACAAGTGCTGAATTCAGATGTAACCGAAGAAGTTTCTTTAACACATTGCACAGTTCCAGATGATGTT 1320
 C  T  D  I  M  D  E  A  D  V  S  E  T  V  I  I  P  E  Q  V  L  N  S  D  V  T  E  E  V  S  L  T  H  C  T  V  P  D  D  V
TTAGCTTCTGATATTACTTCAGCCTCAATATCTATGCCGAAACGAGTCTTGACAAGTGAATCTATACATGTGTCTGACATTGGTCACGTTGAACATGTGGTTCATGATAGTGTAGTAGAA 1440
 L  A  S  D  I  T  S  A  S  I  S  M  P  K  R  V  L  T  S  E  S  I  H  V  S  D  I  G  H  V  E  H  V  V  H  D  S  V  V  E
GCAGAAATCATCACAGATCCTCTGGCCACTGATGTTGTCTCAGAAGAAGTGTTGGTAGCAGACTGTGCCTCTGAAGCAGTCATAGATGCCAACAGGATCCCTGTGAACCAGCAGGATGAA 1560
 A  E  I  I  T  D  P  L  A  T  D  V  V  S  E  E  V  L  V  A  D  C  A  S  E  A  V  I  D  A  N  R  I  P  V  N  Q  Q  D  E
GAGAAAAACAACTGTGAGGACTACCTTATGATCTCTTGTGGAACTGTAGATATTGTCGAAAGTGAGCCTGAGAATGAGCTTGGAGTTGAACTACTTGATCCAAACAATAGTATTTGCGTT 1680
 E  K  N  N  C  E  D  Y  L  M  I  S  C  G  T  V  D  I  V  E  S  E  P  E  N  E  L  G  V  E  L  L  D  P  N  N  S  I  C  V
CCCAGGGAAAAGATGGTTTATATGGCTGTTAACAACTCTCAGCAAGAAGAAGAACTGAATGTCACTGAGATTGCTGATGAAATTTATATGGAAGTGATTGTTGGAGAGGAGGATGCTACA 1800
 P  R  E  K  M  V  Y  M  A  V  N  N  S  Q  Q  E  E  E  L  N  V  T  E  I  A  D  E  I  Y  M  E  V  I  V  G  E  E  D  A  T
GCCTCAGAGCATGAGCAACAAGTAGAGGACAACAAAATGAAAACCTTCATGCCGATTGCATGGGAAGCTGCTTATGGTAATAATTCTGATGGAATTGAAAACCGGAATGGCACTGCAAGT 1920
 A  S  E  H  E  Q  Q  V  E  D  N  K  M  K  T  F  M  P  I  A  W  E  A  A  Y  G  N  N  S  D  G  I  E  N  R  N  G  T  A  S
GCCCTCTTGCACATAGATGAGTCTGCTGGCCTCGGCAGGCTGGCTAAACAAAAACCAAAGAAAAGGAGACCTGATTCCAGGCAGTAC[CAAACAGCAATAATTATTGGCCCAGATGGACAT 2040
 A  L  L  H  I  D  E  S  A  G  L  G  R  L  A  K  Q  K  P  K  K  R  R  P  D  S  R  Q  Y [Q  T  A  I  I  I  G  P  D  G  H
CCTCTGACTGTCTACCCTTGCATGATTTGTGGCAAGAAGTTTAAATCGAGAGGTTTCTTGAAAAGACACATGAAAAACCATCCTGAACACCTTGCCAAGAAAAAGTACCGCTGTACTGAC 2160
 P  L  T  V  Y  P  C  M  I  C  G  K  K  F  K  S  R  G  F  L  K  R  H  M  K  N  H  P  E  H  L  A  K  K  K  Y  R  C  T  D
TGTGATTACACTACCAACAAGAAGATAAGTTTACACAATCACCTGGAGAGCCACAAGCTAACCAGCAAGGCAGAGAAAGCCATAGAATGCGATGAGTATGGAAAGCATTTCTCTCATGCT 2280
 C  D  Y  T  T  N  K  K  I  S  L  H  N  H  L  E  S  H  K  L  T  S  K  A  E  K  A  I  E  C  D  E  Y  G  K  H  F  S  H  A
GGCGCTTTGTTTACCCATAAAATGGTGCATAAGGAAAAAGGAGCCAACAAAATGCACAAGTGTAAGTTCTGTGAATATGAAACAGCTGAACAAGGCCTATTGAATCGCCATCTTTTGGCA 2400
 G  A  L  F  T  H  K  M  V  H  K  E  K  G  A  N  K  M  H  K  C  K  F  C  E  Y  E  T  A  E  Q  G  L  L  N  R  H  L  L  A
GTCCACAGCAAGAATTTTCCTCACATTTGTGTAGAGTGTGGCAAAGGTTTCTGCCATCCATCAGAGCTCAAAAAGCACATGCGAATCCATACTGGAGAAAAGCCCTATGAATGCCAGTAC 2520
 V  H  S  K  N  F  P  H  I  C  V  E  C  G  K  G  F  C  H  P  S  E  L  K  K  H  M  R  I  H  T  G  E  K  P  Y  E  C  Q  Y
TGCGAATACAGGTCCACAGATTCTTCTAACTTGAAAACCCATGTAAAAACTAAGCATAGTAAAGAGATGCCATTCAAGTGTGACATTTGTCTTCTGACTTTCTCAGATACCAAAGAGGTG 2640
 C  E  Y  R  S  T  D  S  S  N  L  K  T  H  V  K  T  K  H  S  K  E  M  P  F  K  C  D  I  C  L  L  T  F  S  D  T  K  E  V
CAACAACATGCTCTTGTCCATCAAGAGAGCAAAACTCACCAGTGTTTGCATTGTGACCACAAGAGTTCAAACTCAAGTGATTTGAAACGACACATAATTTCAGTTCATACGAAGGACTAT 2760
 Q  Q  H  A  L  V  H  Q  E  S  K  T  H  Q  C  L  H  C  D  H  K  S  S  N  S  S  D  L  K  R  H  I  I  S  V  H  T  K  D  Y
CCTCATAAGTGTGACATGTGTGATAAAGGCTTTCATAGGCCTTCAGAACTCAAGAAACATGTGGCTGCCCACAAGGGTAAAAAAACGCACCAGTGTAGACATTGTGACTTTAAGATTGCA 2880
 P  H  K  C  D  M  C  D  K  G  F  H  R  P  S  E  L  K  K  H  V  A  A  H  K  G  K  K  T  H  Q  C  R  H  C  D  F  K  I  A
GATCCGTTTGTACTAAGTCGCCATATTCTCTCGGTTCATACAAAGGACCTTCCATTTAGGTGTAAGAGATGTAGAACGAGATTTCGGCAACAAAGTGAGCTTAAAAAGCATATGAAGACC 3000
 D  P  F  V  L  S  R  H  I  L  S  V  H  T  K  D  L  P  F  R  C  K  R  C  R  T  R  F  R  Q  Q  S  E  L  K  K  H  M  K  T
CACAGTGGCCAGAAAGTGTATCAGTGTGAGTACTGTGACTATAGCACTACAGATGCCTCAGGCTTTAAGCGGCACGTTATCTCTATTCATACGAAAGACTATCCTCACTGCTGTGAGCAC 3120
 H  S  G  Q  K  V  Y  Q  C  E  Y  C  D  Y  S  T  T  D  A  S  G  F  K  R  H  V  I  S  I  H  T  K  D  Y  P  H  C  C  E  H
TGCAAGAAAGGATTCCGAAGACCCTCGGAAAAGAACCAGCACATAATGCGACATCATAAAGAAGTTGGCCTGCCCTAACAGTACTCTTCATAGCTGTTTGTAGAGAACTGGCCTTGAAAC 3240
 C  K  K  G  F  R  R  P  S  E  K  N  Q  H  I  M  R  H  H  K  E  V  G  L  P
AGAAAATTCATTTAGAAGCCAATCAGTCTTGTTCACATACAATACTGTATATTGATTTATGCTGTGTACAAATAGGATTATTGCTTCTCGTTGACTGTTTTTACGTTTTGTTCAATGGTG 3360
TGTTCTGAATTCTATTCCGTTTGTTTAATACATGGGAAAAGGCAGCAACACGTTAGTTGCTTTTACTAAGTAATCCCTGATTCTATACTGAAGTTTTCTATCTTAGAAGTTTTATATTCA 3480
TTTTAAAATATTTAACTTGCTTACCTTGATGGACCAGCTAACTCATGGTAACTCTGCATTGGTAACTGAGAGGATTCATGTGGACTTTTTTGTCTCACAAATTTCTCACAATTAAATCGT 3600
CAAGAGACATCTAGTGGAATGGGAGATTTTGTATGGTGAGGTCTCATGCTCGTAACATGGAAGTCATTTACTTGTCTTGCTTAGTAATGTAGACTGTATGACAATACAAGTTTTCCTATG]* 3720
AACTCTTGTGTCCCTGGCATCGCTGTGTAGAAGAGCAGTAGCTGGGTTTGTGTTTACCTTTCTTTCTGATTTTTTTTTTCCTTTTTAGTGGAGAAAATATACTTTCTTTTCTTTAGACTC 3840
TTTCTTTACATCTTCTGTCAGTGTAGTAAACTTTTAGAAAATCTTGTTGAAATTGTTTGCTTGATCCTGTTG 3910
```

Fig. 2. DNA sequence of the *Zfa* gene. The complete coding sequence of the *Zfa* ORF is shown as well as 969 nucleotides 5′ to the first in frame methionine and 592 nucleotides 3′ to the termination codon. The region of overlap with the p705 cDNA clone is indicated by square brackets. The complete sequence of p705 was also obtained and was identical to the sequence presented here. The site of polyadenylation of *Zfa* as determined by the presence of poly(A) in two independent cDNA clones is indicated by an asterisk. The mutation in the third finger is underlined. The direct repeat found at the 3′ end of *Zfa* is double underlined.

we isolated a genomic clone of *Zfa*. DNA sequence analysis (Figure 2) revealed that the gene had an uninterrupted coding region largely co-linear with *Zfy-1* and *Zfy-2* cDNAs and highly homologous to exons of the *Zfx* gene (Figure 3a). As the *Zfx* gene has a minimum of six introns (B.Skene and A.Ashworth, unpublished observations), and all of these appear to be absent from *Zfa*, we concluded that *Zfa* arose as the result of the reverse transcription and integration of a processed mRNA coding for *Zfx*. In agreement with this conclusion is the presence of a poly(A) sequence at the 3′ end of the *Zfa* gene at the point of divergence between *Zfx* and *Zfa*. A direct repeat of 10 bp has been identified at this point and at the 5′ end of *Zfa* close to the CpG island (Figure 3b).

The open reading frame (ORF) of *Zfa* encodes a protein of 85 kd, similar in overall structure to that of *Zfy-1*

Fig. 3. (**a**) Comparison of *Zfx* and *Zfa* gene sequences. The upper sequence is *Zfx* and the lower *Zfa*. Numbers indicate residues in *Zfa* as in Figure 2. Asterisks indicate differences between the two sequences and dashes deletions in *Zfa*. Consensus intron donor and acceptor splices sites are underlined. Only intron sequences adjacent to splice junctions are shown. (**b**) Point of divergence of *Zfa* and *Zfx* genes. The upper sequence is *Zfx* and the lower *Zfa*. Asterisks indicate differences between the two sequences. The poly(A) sequence is underlined and the direct repeat [also present at nucleotides 329–338 (Figure 2)] double underlined.

(Ashworth *et al.*, 1989), with a putative DNA-binding zinc finger domain and an acidic region which may constitute a transcription activation domain (Ptashne, 1988). Thus, *Zfa* may also encode a positively acting transcription factor. Upstream of the coding region are recognition sites for several rare-cutting restriction enzymes which are frequently found in 'CpG islands' overlapping the transcription start sites of genes (Bird, 1986). This region may be part of the promoter of *Zfa*.

Zfa is expressed in a tissue-specific manner

The isolation of cDNA clones from an adult mouse testis library demonstrated conclusively that *Zfa* is expressed. Northern hybridization was used to analyse this further. A fragment of the *Zfa* gene detects several transcripts in both male and female tissues (Figure 4). The 6 and 8 kb species present in all tissues are thought to derive from *Zfx*, since similar *ZFX* transcripts appear to be ubiquitously expressed in humans (Schneider-Gadicke *et al.*, 1989a). The site of polyadenylation defined by *Zfa* cDNA clones and the location of the putative transcription start site mentioned above suggest that the 3.5 kb adult testis-specific transcript is derived from the *Zfa* gene. Analysis of RNA by RNase protection and by the reverse transcriptase–polymerase chain reaction (RT–PCR) method (Saiki *et al.*, 1988) confirm that the *Zfa* gene is expressed only in testis.

Zfa is derived from one of the alternative splices of the *Zfx* gene

Primers deduced from the sequence of the *Zfa* gene were used to study the expression of *Zfa* and *Zfx* by RT–PCR. This analysis showed that at least three distinct mRNAs derive from the *Zfx* gene, whereas we found only one transcript derived from *Zfa*. Comparison with genomic DNA (see Figure 3a) shows that the multiple *Zfx* transcripts arise by alternative splicing. All of the transcripts are spliced so that they retain an ORF. One of these RNA species contains a 150 bp exon not present in *Zfa*. This transcript (I in Figure 5) is largely colinear with *Zfy-1/2* (Ashworth *et al.*, 1989; Mardon and Page, 1989). A second transcript of *Zfx* (II in Figure 5) lacks the exon (nucleotides 1877–2014, Figures 2 and 3a) encoding the putative nuclear localization signal. The third (III in Figure 5) is a transcript of *Zfx* that is co-linear with *Zfa* over this region and we assume that this is similar to the transcript from which the *Zfa* gene arose.

Zfa and *Zfy-2* contain mutations in the third zinc finger

Sequence comparison shows that the zinc finger domains of *Zfa* and *Zfx* are virtually identical. However, there is an amino acid substitution in the third finger of *Zfa* where one of the cysteines predicted to co-ordinate the Zn^{2+} ion has been altered to a tyrosine (Figure 6). Studies of other proteins with multiple zinc fingers such as *Kruppel* (Redemann *et al.*, 1988), TFIIIA (Vrana *et al.*, 1988) and *ADR1* (Blumberg *et al.*, 1987) suggest that such a mutation could alter the specificity or the affinity of the protein for DNA. *Zfa* is present in the wild mouse species *Mus spretus* which is thought to have diverged from *M.musculus* 3–5 million years ago (Ferris *et al.*, 1983). The polymerase chain reaction was used to demonstrate that the *spretus* gene also carries the mutation. Comparison of the nucleotide sequence of the *Zfa* gene with that of *Zfx* indicates that *Zfa* originated ~5 million

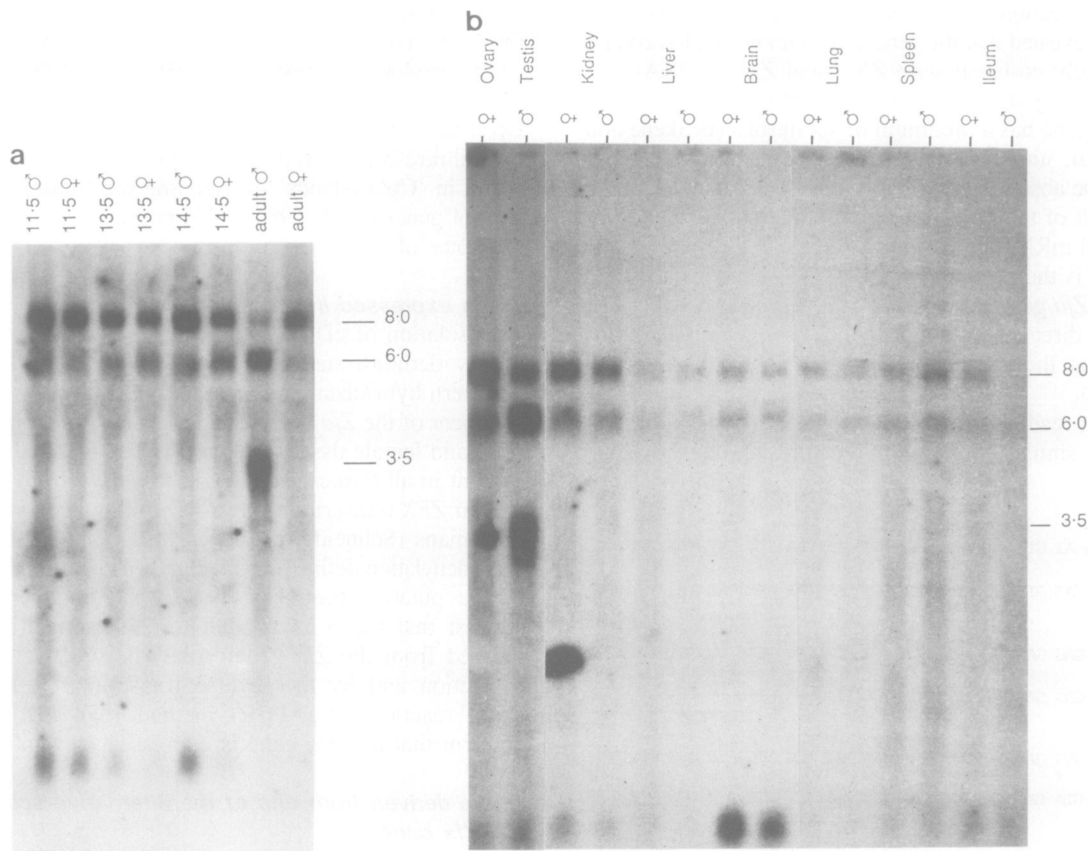

Fig. 4. (**a**) Northern analysis of *Zfx* and *Zfa* gene expression during gonadal development. (**b**) Northern analysis of *Zfa* and *Zfx* gene expression in adult male and female tissues. Hybridization was with a probe from the 5′ end of the *Zfa* gene. The 6 and 8 kb transcripts have been assigned to the *Zfx* gene, and the 3.5 kb transcript to the *Zfa* gene (see text). An additional 700 bp transcript of unknown origin is detected in some tissues. A probe derived from the 5′ end of *Zfy-2* detects a transcript of similar size but only in adult testis (Mardon and Page, 1989). The 3 kb adult testis-specific *Zfy* transcript (Mardon and Page, 1989) is not readily detected by this *Zfa* probe and rehybridization of these filters with a *Zfy-1* probe indicates that the 3 kb *Zfy* and 3.5 kb *Zfa* transcripts are distinguishable by size.

years ago (not shown). It is, therefore, possible that the mutation occurred at the same time as the retroposition event that created *Zfa*. Indeed, this may have been a prerequisite for its persistence as an additional active copy of the *Zfx* gene may be deleterious. However, the retention of an ORF in *Zfa* indicates selective pressure for the maintenance of some function of this protein. Figure 6 also shows the sequence of part of the finger domain of the *Zfy-2* gene from the *M.musculus musculus* Y chromosome which is present in most strains of laboratory mice (Bishop *et al.*, 1986). This gene also has a mutation in the third finger comprising of an in-frame deletion of six amino acids. This mutation is more recent than that found in *Zfa* as it is not found in the *M.musculus domesticus Zfy-2* gene (Mardon and Page, 1989). Both *Zfa* and *Zfy-2* expression appears to be restricted to the testis (Figure 4; Mardon and Page, 1989).

Discussion

The presence of additional *Zfy*-related genes in mice has raised questions as to their origin and possible functions. Here, we describe the structure and origin of an autosomal member of this gene family. Our results suggest that this gene, *Zfa*, may have a function in the adult testis.

Isolation of the *Zfa* gene and comparison with the *Zfx* gene show that *Zfa* was derived by retroposition of a *Zfx* transcript.

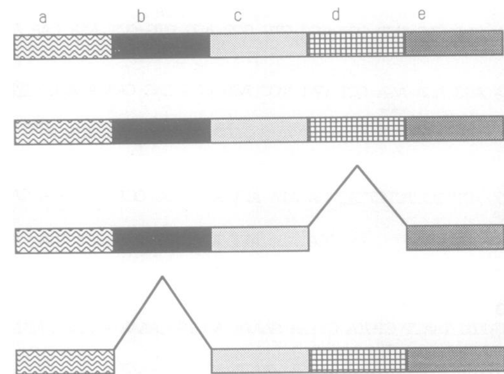

Fig. 5. Alternative splicing of *Zfx*. Boxes indicate regions of identity between different *Zfx* transcripts. Shown for comparison are the *Zfy-1* and *Zfy-2* transcripts which are co-linear with the *Zfx* I mRNA. These RNAs contain an additional 150 bp (box b—inserted after nucleotide 1597) compared to *Zfa*. A second transcript, *Zfx* II, lacks the exon coding for the putative nuclear localization signal (box d—equivalent to nucleotides 1877–2014 in *Zfa*). A third transcript, *Zfx* III, is co-linear with *Zfa* over this region. Sequencing of genomic DNA (B.Skene and A.Ashworth, unpublished observations) confirms that introns are present in the *Zfx* gene after nucleotides equivalent to positions 1597, 1876 and 2014. The sequence of the *Zfx* splice junctions at positions 1876 and 2014 is shown in Figure 3. It is likely that the *Zfx* gene is split by additional introns than those indicated here.

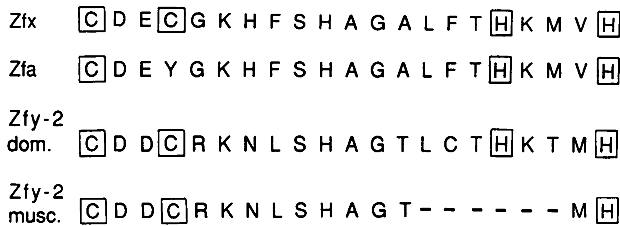

Fig. 6. Third zinc finger sequences of *Zfy*-related genes. The deduced amino acid sequences for the third finger of several *Zfy*-related genes is shown. The *Zfx* sequence is from 129/Sv mice. The *Zfa* sequence was determined from 129/Sv, C57BL/6J and *M.spretus* mice and is identical over this region. *Zfy2-mus* and *Zfy2-dom* sequences are from *M.musculus musculus* and *M.musculus domesticus*, respectively. The cysteines and histidines responsible for co-ordinating the Zn^{2+} ion are boxed.

As well as the absence of introns, the *Zfa* gene has the vestiges of a poly(A) tail and is flanked by direct repeats, the hallmarks of retroposons (Rogers, 1985). Unlike most retroposons, *Zfa* codes for a protein and is transcriptionally active. However, although the *Zfx* gene from which it is derived is ubiquitously expressed, *Zfa* expression appears to be restricted to adult testis. Among the relatively small number of known transcriptionally active retroposons two are derived from genes on the X chromosome [PGK-2 (McCarrey, 1987) and pyruvate dehydrogenase E1α subunit (G.Brown, pers. commun.)]. Interestingly both these retroposons are expressed exclusively in the adult testis. In this regard it may be significant that the X chromosome is thought to be inactivated early in spermatogenesis (Monesi, 1971). In the case of PGK-2 the promoter region responsible for the testis-specific pattern of expression is very close to the start of transcription and is within the region retroposed (Robinson *et al.*, 1989).

An alternative explanation for the difference in the expression pattern of the two genes is that multiple promoters exist for *Zfx* and that *Zfa* was derived from a transcript including sequences responsible for expression in the testis. This is the case in the human Hα44 tubulin gene. This gene is ubiquitously expressed but use of a different promoter in the testis results in the recruitment of a novel 5′ exon (Dobner *et al.*, 1987).

As described above, *Zfx* appears to be ubiquitously expressed. The use of the PCR technique, however, has revealed unexpected complexity in the expression of *Zfx*. Three different *Zfx* transcripts, presumably derived by alternative splicing, were detected. One of these is the alternative use of a 150 bp exon within the region of the protein which is rich in acidic residues. The omission of this exon might alter the potential of this putative transcription factor to activate transcription. This exon was absent from the transcript that gave rise to *Zfa*. One of the other splicing products results in the absence of the exon thought to encode a nuclear localization signal (Ashworth *et al.*, 1989; Mardon and Page, 1989). Thus, the protein encoded by this mRNA might be cytoplasmically located and be inactive or have a role other than in the inactivation of transcription. Although the differential use of exons may be important in the regulation of *Zfx* activity we have yet to discover if any of these splicing patterns are stage or cell-type specific. The control of cellular processes at the level of splicing has been demonstrated in a number of systems including the expression of the immunoglobulin genes (Rogers *et al.*, 1980) and sex determination in *Drosophila* (Nagoshi *et al.*, 1988). After this paper was submitted for publication Schneider-Gadicke *et al.* (1989b) reported that human *ZFX* also produces alternatively spliced transcripts. These, however, are distinct from the transcripts described here.

The coding regions of the *Zfa* and *Zfx* genes are highly homologous. Consistent with their proposed functions (Ptashne, 1988), the acidic region of the protein is more diverged than the potential DNA-binding zinc finger domain. Despite the very high level of sequence conservation in the finger domain of *Zfa* and *Zfx*, *Zfa* has sustained a potentially significant mutation that might compromise the binding of Zn^{2+} by the third zinc finger. However, as this mutation has been present in the *Zfa* gene for at least 3 million years, it seems not to have abolished any function of *Zfa* as an ORF has been retained. It appears, therefore, that *Zfa* has retained or acquired some function in the adult testis of the mouse. The presence of a similar mutation in the *M.musculus musculus Zfy-2* gene suggests that the third finger of these genes may be dispensable for their function in the testis.

In the mouse, *Zfx* is ubiquitously expressed whereas *Zfa* and the *Zfy* genes appear to be testis specific. In contrast the two human genes, *ZFY* and *ZFX* are expressed in all cell types (Schneider-Gadicke *et al.*, 1989a). From this change in expression pattern we conclude that the *ZFY*-related genes have dual functions; one necessary in all cell types and one specific to testis. If the testis-specific function is not subject to the same sequence constraints as the ubiquitous function, this may explain the presence of mutations in the zinc fingers of *Zfa* and *Zfy-2*. It could also account for the divergence of the murine *Zfy* genes compared to the *ZFY* genes of other mammals which are highly conserved (Page *et al.*, 1987).

Materials and methods

cDNA and genomic clones

cDNAs p705 and p703 were isolated from an adult mouse testis cDNA library in λZap (from K.Willison) using a human *ZFY* probe (from N.Affara; Ashworth *et al.*, 1989). The *Zfa* gene and fragments of the *Zfx* gene were isolated from a 129/Sv mouse genomic library (from Lisa Stubbs, ICRF, London) using a human *ZFY* cDNA clone, MF-1 (Sinclair *et al.*, 1988).

DNA sequencing

Unidirectional deletions were created in the *Zfa* gene as described (Henikoff *et al.*, 1984). Plasmids were sequenced directly using Sequenase™ (USB) according to the manufacturer's instructions. The sequence of the third finger of *Zfy-2* from the *M.musculus musculus* Y chromosome was obtained by sequencing plasmid pDP1171 (a gift of D.Page) with a synthetic oligonucleotide. The existence of a deletion in this region was suggested by the results of Nagamine *et al.* (1989). The *Zfx* sequence was obtained by sequencing fragments of the murine *Zfx* gene subcloned into Bluescript plasmids. The rate of divergence of the *Zfa* gene was estimated by comparing a portion of the *Zfa* gene 3′ flanking region which was retroposed but is now not transcribed with the corresponding region of *Zfx*. A 5% sequence divergence was observed over a region of ~300 bp. A rate of divergence of 1% per million years was assumed (Wu and Li, 1985).

Southern and Northern blotting

DNA (5 μg) was digested with restriction enzymes electrophoresed on a 0.8% agarose gel, transferred to Genescreen (Dupont) and hybridized with the ^{32}P-labelled insert of p705 as described by Church and Gilbert (1984). RNA was isolated from mouse (Parkes) genital ridges and foetal gonads at 11.5–14.5 days post-coitum, and from adult gonads and other tissues at ~8 weeks of age by the method of Auffray and Rougeon (1980). Embryos were sexed at 11.5 days by staining for sex chromatin in amnion cells (Monk and McLaren, 1981) and at later stages by gonad morphology. Total RNA (10 μg) was electrophoresed in formaldehyde–agarose gels and transferred to Hybond (Amersham). Hybridization was performed according to the

manufacturer's instructions with a ^{32}P-labelled probe derived from nucleotides 437–1268 of the *Zfa* gene and washed in 0.2 × SSC at 50°C.

PCR analysis

The *Zfa* gene was amplified from total genomic DNA using PCR and primers corresponding to nucleotides 1469–1488 and 2349–2368 of the *Zfa* sequence. PCR products were subcloned into the plasmid Bluescript (Stratagene) and sequenced. At least three independent clones were sequenced. C57BL/6J DNA was obtained from the Jackson Laboratory and *M.spretus* DNA from K.Willison. RNA isolated from 14.5 day old mouse embryos, and brain, liver and testis of adult mice was reverse transcribed into cDNA (Kawasaki *et al.*, 1988). *Zfx*/*Zfa* cDNA was amplified by PCR (Saiki *et al.*, 1988) with primers corresponding to nucleotides 1469–1488 and 2050–2069, and subcloned into the plasmid vector Bluescript. DNA sequencing was used to determine the structure of the transcripts.

Acknowledgements

This work was supported by the Cancer Research Campaign and the Medical Research Council. Part of this work was carried out in the laboratory of K.Harbers, supported by a Royal Society European Exchange Fellowship to B.S. We thank K.Willison for a cDNA library and *Mus spretus* DNA, L.Stubbs for a genomic library, N.Affara and P.Goodfellow for human *ZFY* probes, D.Page for pDP1171, and C.J.Marshall, A.Lloyd, R.Haffner, K.Willison and A.McLaren for helpful discussions. The assistance of J.Collignon, J.Gubbay, P.Koopman, A.Muensterberg, M.Riddell and N.Vivian with genital ridge dissection is appreciated.

References

Ashworth,A., Swift,S. and Affara,N. (1989) *Nucleic Acids Res.*, **17**, 2864.
Auffray,C. and Rougeon,F. (1980) *Eur. J. Biochem.*, **107**, 303–324.
Bird,A.P. (1986) *Nature*, **321**, 209–213.
Bishop,C.E., Boursot,P., Baron,B., Bonhomme,F. and Hatat,D. (1986) *Nature*, **315**, 70–72.
Blumberg,H., Eisen,A., Sledziewski,A., Bader,D. and Young,E.T. (1987) *Nature*, **328**, 443–445.
Boulay,J.L., Dennefeld,C. and Alberga,A. (1987) *Nature*, **330**, 395–398.
Brown,R.S., Sander,C. and Argos,P. (1985) *FEBS Lett.*, **186**, 271–274.
Burgoyne,P.S., Buehr,M., Koopman,P., Rossant,J. and McLaren,A. (1988) *Development*, **102**, 443–450.
Church,G. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
Diakun,G.P., Fairall,L. and Klug,A. (1986) *Nature*, **324**, 698–699.
Dobner,P.R., Kislawkis,E., Wentworth,B.M. and Villa-Komaroff,L. (1987) *Nucleic Acids Res.*, **15**, 199–218.
Fairall,L., Rhodes,D. and Klug,A. (1986) *J. Mol. Biol.*, **192**, 577–591.
Ferris,S.D., Sage,R.D., Prager,E.M, Ritte,U. and Wilson,A.C. (1983) *Genetics*, **105**, 681–721.
Hartshorne,T.A., Blumberg,H. and Young,E.T. (1986) *Nature*, **320**, 283–287.
Henikoff,S. (1984) *Gene*, **28**, 351–359.
Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R. (1987) *Cell*, **51**, 1079–1090.
Kawasaki,E.S., Clark,S.S., Coyne,M.Y., Smith,S.D., Champlin,R., Witte,O.N. and McCormick,F.P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5698–5702.
Koopman,P., Gubbay,J., Collignon,J. and Lovell-Badge,R. (1989) *Nature*, **342**, 940–942.
Mardon,G. and Page,D.C. (1989) *Cell*, **56**, 765–770.
Mardon,G., Mosher,R., Disteche,C.M., Nishioka,Y., McLaren,A. and Page,D.C. (1989) *Science*, **243**, 78–80.
McCarrey,J.R. (1987) *Gene*, **61**, 291–298.
McLaren,A. (1988) *Trends Genet.*, **4**, 153–157.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Monesi,V. (1971) *J. Reprod. Fertil.*, **13**, 1–14.
Monk,M. and McLaren,A. (1981) *J. Embryol. Exp. Morphol.*, **63**, 75–84.
Moses,K., Ellis,M.C. and Rubin,G.M. (1989) *Nature*, **340**, 531–536.
Nagai,K., Nakeseko,Y., Nasmyth,K. and Rhodes,D. (1988) *Nature*, **332**, 284–286.
Nagamine,C.M., Chan,K., Kozak,C.A. and Lau,Y.-F. (1989) *Science*, **243**, 80–83.
Nagoshi,R.N., McKeown,M., Burtis,K.C., Belote,J.M. and Baker,B.S. (1988) *Cell*, **53**, 229–236.
Page,D.C., Mosher,R., Simpson,E.M., Fisher,E.M.C., Mardon,G., Pollack,J., McGillivray,B., Chapelle,A. de la and Brown,L.G. (1987) *Cell*, **51**, 1091–1104.
Palmer,M.S., Sinclair,A.H., Berta,P., Ellis,N.A., Goodfellow,P.N., Abbas,N.E. and Fellous,M. (1989) *Nature*, **342**, 937–939.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Redemann,N., Gaul,U. and Jackle,H. (1988) *Nature*, **332**, 90–92.
Robinson,M.O., McCarrey,J.R. and Simon,M.I. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8437–8441.
Rogers,J., Early,P., Carter,C., Calame,K., Bond,M., Hood,L. and Wall,R. (1980) *Cell*, **20**, 303–312.
Rogers,J.H. (1985) *Int. Rev. Cytol.*, **93**, 187–279.
Saiki,R.K., Gelfand,D.H., Stofel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
Schneider-Gadicke,A., Beer-Romero,P., Brown,L.G., Nussbaum,R. and Page,D.C. (1989a) *Cell*, **57**, 1247–1258.
Schneider-Gadicke,A., Beer-Romero,P., Brown,L.G., Mardon,G., Luoh,S.W. and Page,D.C. (1989b) *Nature*, **342**, 708–711.
Sinclair,A.H., Foster,J.W., Spencer,J.A., Page,D.C., Palmer,M., Goodfellow,P.N. and Graves,J.A.M. (1988) *Nature*, **336**, 780–783.
Stanojevic,D., Hoey,T. and Levine,M. (1989) *Nature*, **341**, 331–335.
Treisman,J. and Desplan,C. (1989) *Nature*, **341**, 335–337.
Vrana,K.E., Churchill,M.E.A., Tullius,T.D. and Brown,D.D. (1988) *Mol. Cell. Biol.*, **8**, 1684–1696.
Wu,C.I. and Li,W.-H. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1741–1745.

Received on January 2, 1990; revised on February 5, 1990