Comparison of ZFY and ZFX gene structure and analysis of alternative 3′ untranslated regions of ZFY

M.North, C.Sargent, J.O'Brien, K.Taylor[1], J.Wolfe[1], N.A.Affara* and M.A.Ferguson-Smith
University of Cambridge Department of Pathology, Tennis Court Road, Cambridge CB2 1QP and
[1]Department of Genetics, The Galton Laboratory, University College London, London, UK

Received March 14, 1991; Revised and Accepted April 18, 1991

ABSTRACT

We have cloned and sequenced transcripts from the X- and Y-linked zinc-finger genes ZFX and ZFY respectively and discuss a possible mechanism of post-transcriptional control by which these genes can be widely expressed but translated in only specific tissues. We report the identification of a novel 3′UTR (untranslated region) present in ZFY which is highly conserved among primates and contains a series of motifs implicated as mRNA instability determinants. These sequences can be substantially removed by polyadenylation directed from consensus (AATAAA) and non-consensus (AATATAAA) sequences in adult testis. The DNA-binding domains of the ZFY and ZFX proteins are compared using present models for zinc-finger/DNA interactions. Additionally, the genomic organisation of the ZFY coding sequence is presented as compared to that of ZFX.

INTRODUCTION

Early hopes that ZFY was the testis determining factor (TDF) have receded in the light of strong evidence to the contrary. First, the homologous sequences in tested marsupial genomes are autosomal despite a primary sex-determining role for the Y in these mammals (1); second, ZFY is absent from the genomes of a distinct class of XX male and all XX true hermaphrodites thus far examined (2); third,the mouse homologue (Zfy) is not expressed in somatic cells of the developing gonad but is confined to germ cell lineages (3); finally, three ZFY-negative XX males and an XX true hermaphrodite have been shown to contain a 35Kb segment of the Y immediately proximal to the pseudoautosomal boundary which does not include ZFY (4) thus indicating that TDF must lie in this more distal part of the Y. Recently it has been shown that this region of the Y contains a plausible new candidate for TDF (referred to as SRY, the sex determining region of the Y) which contains a putative DNA binding domain similar to that identified in the RNA polymerase I transcription factor hUBF (5). Furthermore, its pattern of expression in the differentiating mouse gonad is entirely consistent with a sex-determining role since its expression occurs in the somatic components of the embryonic gonad and correlates with the differentiation of Sertoli cells (6).

Although testicular tissue can develop in the absence of ZFY, it nevertheless continues to be a gene of some developmental interest. In the mouse, Zfy-1 transcripts first appear at about 10.5 days post coitum, coinciding with the migration of germ cells into the differentiating gonad (3). The transcripts are absent in the testes of embryonic mice homozygous for the W^e mutation which lack germ cells (3). Thus the gene may have a role in early germ cell development. In adult mouse testis, its expression is mainly restricted to germ cells, persisting even in post-meiotic spermatids (7). In this paper, we have examined and compared the sequences of the human ZFY and ZFX transcripts and provide evidence that somatic ZFY transcripts contain a long 3′ untranslated region (UTR) which is substantially removed in germ cells and may be an important element of post-transcriptional control.

MATERIALS AND METHODS

General methods

Restriction enzyme digests, Southern analysis, purification and subcloning of DNA fragments, bacterial transformations and bacteriophage manipulations were carried out by standard methods (8). Oligonucleotides were synthesised on a Pharmacia LKB Gene Assembler Plus, treated for 24 hours with 30% ammonia solution at 55°C and purified by precipitation. DNA sequencing with denatured double-stranded template was carried out by extension from a primer in the presence of dideoxynucleoside triphosphate chain-terminating inhibitors (9). The sequencing reactions were carried out using a Pharmacia DNA sequencing kit, T7 DNA polymerase and alpha ^{35}S dATP.Sequence was analysed using the Genetics Computer Group (GCG) software and database searches carried out with the FASTA software. Low stringency Southern hybridisation was carried out for 48 hours at 37°C in standard 50% formamide buffer as described previously (10) with washing at 40°C in 1×SSC;0.1% (w/v) SDS.

Polymerase chain reaction

Amplification reactions were carried out in 25 μl volumes containing 10 mM Tris.HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% (w/v) gelatin and 200 μM of each dNTP.

* To whom correspondence should be addressed

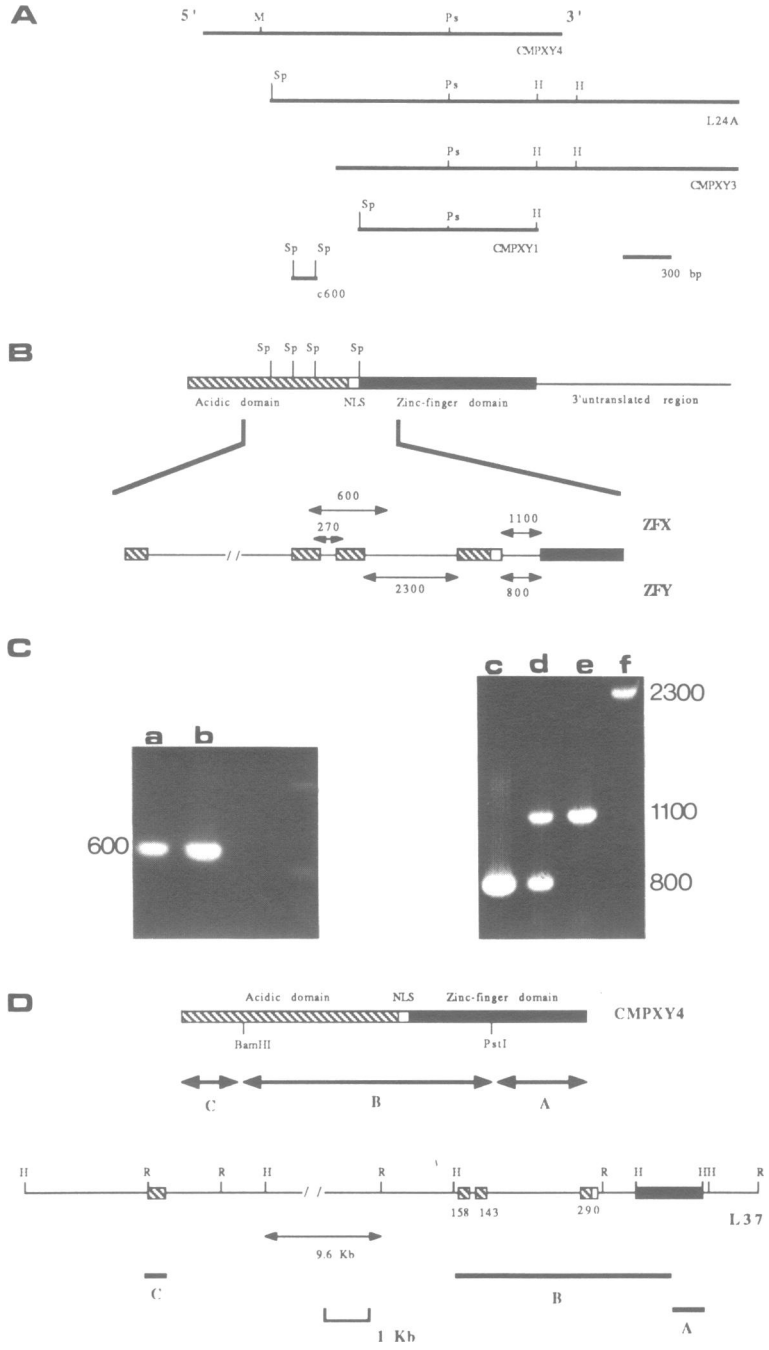

Figure 1. Section A: Alignment and orientation of the cDNA clones CMPXY4, L24A, CMPXY3 and the genomic clones CMPXY1 and c600. HindIII (H), PstI (Ps) and BamHI (M) restriction sites are shown, as are splice junctions (Sp). Clones L24A and CMPXY4 contain 5′ ends of autosomal origin which are not included in this diagram. Section B: The splice junctions positioned from sequence comparison and analysis of the clones described above (and aligned with the clones in section A) allowed some of the intron/exon structure of ZFY and ZFX to be determined. The sizes of some of the introns (calculated from the PCR experiments listed below) are shown. NLS = Nuclear Localisation Signal. Section C: PCR primers were designed to flank introns and successful amplification products are illustrated schematically in section B. Lanes a and b are male and female genomic DNA (respectively) amplified with the primers 5′ TAGCATAATAATCTCTAAAA 3′ (derived from the second most 5′ exon) and 5′ AACACGATGGTTCTTCTGAA 3′ (derived from the 3′ flanking intron sequence of the third most 5′ exon which is contained in clone c600) leading to the synthesis of an X-specific 600 bp product. From the known length of exon sequence in this product, the×intron size can be calculated as 270 bp. Analysis of cosmid L37 suggests that on the Y these two exons are separated by an intron of similar size. Lanes c, d and e are L37 cosmid DNA, male genomic DNA and female genomic DNA (respectively) amplified with the PCR primers 5′ TCTTCCCACAAATCATGCAAGG 3′ (from the 3′ end of the exon containing the NLS) and 5′GAGAAGACCTGATTCCAGGCA 3′ (from the 5′ end of the zinc finger exon) leading to the synthesis of an 800 bp Y-specific product and a 1100 bp X-specific product. Lane f is cosmid L37 DNA amplified with the PCR primers 5′ GAGGGCACTTGCAGTGCCATT 3′ (from the 3′ end of the c600 exon) and 5′ GTGTTCCCAGGGAAAAGATGG 3′ (from the 5′ end of the NLS exon) leading to the synthesis of a 2300 bp Y-specific product. The size of the equivalent ZFX intron has not been determined as no PCR product was obtained on genomic DNA suggesting it may larger than that of the Y. All PCR reactions were carried out under the following cycling conditions: 94°C, 60 seconds; 50°C, 120 seconds; 72°C, 300 seconds. Section D: The partial genomic organisation of ZFY in cosmid L37 with the positions of HindIII (H) and EcoRI (R) sites shown as well as those exons detected by various subfragments (A,B and C) of CMPXY4. Exon sizes, where known, are marked and the genomic organisation is consistent with PCR data.

Primers were present at 0.5 μM each and Taq polymerase (Cambio) at 0.02 units/μl. To identify differentially polyadenylated transcripts bacteriophage DNA was prepared from a testis cDNA library high titre stock and a sample of this preparation equivalent to 5×10^9 bacteriophage used as template. Genomic DNA or cosmid DNA to be used as PCR template was prepared as described previously (8). Amplification was performed with a Perkin Elmer Cetus DNA Thermal Cycler, programmed to use the fastest available transitions between each temperature. Cycle conditions used were 94°C, 60 seconds; 55°C, 60 seconds; 72°C, 180 seconds; 35 cycles. Amplification products were analysed by electrophoresis on a 1% agarose gel and detected by staining with ethidium bromide. Primer sequences and conditions of amplification are listed below.

RNaseA protection experiments

Antisense run-off transcripts were prepared using a RNA transcription kit (Stratagene), alpha ^{32}P rUTP (Amersham) and a linearised cDNA subclone in pBS. The labelled riboprobe was precipitated, gel purified and 500,000 cpm resuspended in 20

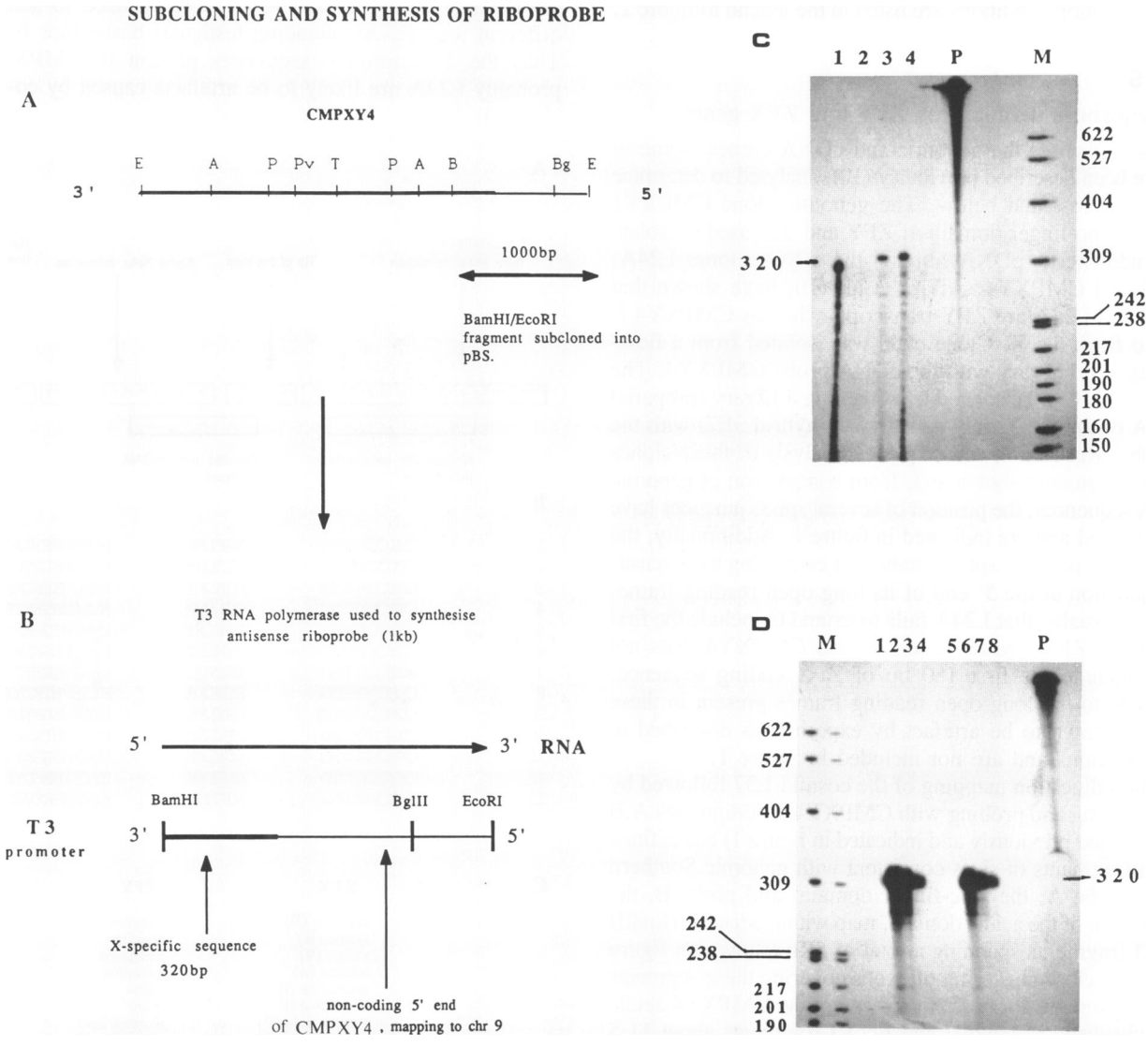

Figure 2. Section A: The coding sequence of CMPXY4 is indicated by a heavy line. The 5′ BamHI/EcoRI subfragment containing X-specific and autosomally-derived sequences was subcloned into pBS for the preparation of riboprobes. Section B: The orientation of the insert relative to the flanking T3 and T7 RNA polymerase promoters was determined by asymmetric restriction digestion and radiolabelled riboprobe synthesised complementary to the coding strand in CMPXY4 using T3 RNA polymerase. Samples of total RNA were incubated overnight with labelled riboprobe and subsequently treated with RNaseA and RNaseT1. Aliquots were then fractionated on 4% polyacrylamide gels with molecular weight markers prepared by digesting pBR322 with MspI and 'end-labelling' using alpha-^{32}P-dCTP. The RNA samples used and the conditions of nuclease digestion are listed below. In all cases, lanes containing molecular weight markers are labelled M and lanes containing undigested probe are labelled P. Section C: Samples of total RNA from a 48,XXXX lymphoblastoid cell line hybridised with the labelled riboprobe and treated with RNase for 90 minutes at 37°C. The autoradiogram is shown after exposure to the gel for 24 hours. Lane 1: 100 μg of 4× total RNA; lane 2: 10 μg tRNA (negative control); lane 3: 10 μg of 4× total RNA; lane 4: 30 μg of 4× total RNA. A slight abnormality in migration of the protected fragment is observed in lane 1 but was not reproduced in subsequent experiments (see section D). Section D: Samples of total RNA from the 47, XYY lymphoblastoid cell line and the adult testis hybridised with riboprobe and treated with RNase for 20 minutes at 37°C. The autoradiogram is shown after exposure to the gel for 48 hours. Lane 1: 10 μg tRNA (negative control); lane 2: 150 μg of XYY total RNA; lane 3: 50 μg of XYY total RNA; lane 4: 10 μg of XYY total RNA; lane 5: 10 μg tRNA (negative control); lane 6: 150 μg testis total RNA; lane 7: 50 μg testis total RNA; lane 8: 10 μg testis total RNA.

μl of 50% (v/v) deionised formamide, 500 mM NaCl, 1 mM EDTA, 40 mM Pipes pH 6.4. 10−150 μg of RNA was resuspended in this buffer and heated to 85°C for 5 minutes followed by overnight incubation at 45°C. After annealing , 300 μl of 10 mM Tris.HCl pH7.5, 5 mM EDTA, 300 mM NaCl containing 40 μg/ml RNaseA and 1 μg/ml RNaseT1 was added to the sample and the solution incubated at 37°C for up to 90 minutes. Samples were subsequently precipitated and fractionated on a polyacrylamide gel with end-labelled, MspI-digested pBR322 used as molecular weight marker. RNA used in the protection assay was derived from the following sources: testis, 46,XXXX, lymphoblastoid cell line and 46,XYY lymphoblastoid cell line. Details of digestion conditions are listed in the legend to figure 2.

RESULTS

Cloned sequences defining the ZFY and ZFX genes

Figure 1 summarises the genomic and cDNA clones (some of which have been described previously (10)) analysed to determine the sequence presented below. The genomic clone CMPXY1 contains the zinc-finger domain of ZFY and was used to isolate (from an adult testis cDNA library) the cDNA clones L24A, CMPXY3 and CMPXY4. Affara et al. (10) have shown that CMPXY3 and L24A are ZFY transcripts whereas CMPXY4 is transcribed from the X. Clone c600 was isolated from a flow-sorted X-specific library with the cDNA probe CMPXY4. The 40Kb cosmid L37 was isolated by screening a library (prepared from DNA of the the Y-only somatic cell hybrid 3E7) with the L24A probe. Restriction and sequence analysis of these clones permits the alignment shown and, from comparison of genomic and cDNA sequences, the position of several splice junctions have been ascertained and are indicated in figure 1. Additionally, the clone L24A is a partially spliced transcript containing a consensus acceptor junction at the 5′ end of its long open reading frame. Sequencing revealed that L24A fails to extend to include the first 630 bp of the ZFY coding sequence and CMPXY4 does not extend to include the first 180 bp of ZFX coding sequence. Sequences 5′ to the long open reading frames present in these cDNAs are shown to be artefact by experiments described in following sections and are not included in figure 1.

Restriction digestion mapping of the cosmid L37 followed by Southern blotting and probing with CMPXY4 subfragments A,B and C (described previously and indicated in figure 1) has defined restriction fragments of sizes consistent with genomic Southern data (10). Probe A, the zinc-finger domain, and probe B, the more 3′ portion of the acidic domain, map within adjacent HindIII and EcoRI fragments spanning a total of 5Kb (shown on figure 1), suggesting close clustering of exons covering these segments of the coding region. Probe C, the 5′ end of the CMPXY4 acidic domain hybridised to HindIII and EcoRI fragments about 11.5 Kb away from probe B, indicating the presence of a large intron of this size. Double digest results show that probe C hybridises to what may be a single exon within a 600 bp region defined by the indicated EcoRI site and a EcoRV site (not shown).

As shown in figure 1, PCR primers were designed to flank prospective introns and successful amplifications have allowed precise sizing of some in ZFX and ZFY. These are detailed in the legend to figure 1. Primers designed to amplify across the large 11.5 Kb intron present in L37 failed (not surprisingly) to give a PCR product. It is interesting to note that a large intron about 25 Kb in length is present in an analogous position in the ZFX gene (13). From figure 1, it can be seen that over the region examined, the points of interruption of the coding sequence of ZFY and ZFX are the same, but the size of some introns is different.

As reported previously, the cDNA clone CMPXY4 contains a 5′ end that detects fragments of autosomal origin (chromosome 9)(10). The possibility that this represents transplicing is excluded by the experiment in figure 2 where an RNase protection assay using an antisense riboprobe of 1Kb covering 320bp of the X specific sequence and all of the autosomal region of CMPXY4 fails to form a double-stranded RNA hybrid 5′ to the X-specific domain. Hence only a 320 bp fragment (the length of the X-specific region of the riboprobe) is protected against RNase digestion when the probe is hybridized in solution to different total RNAs including testicular tissue (see figure 2). Thus the 5′ autosomal sequences present in CMPXY4 and probably L24A are likely to be artefacts caused by co-cloning

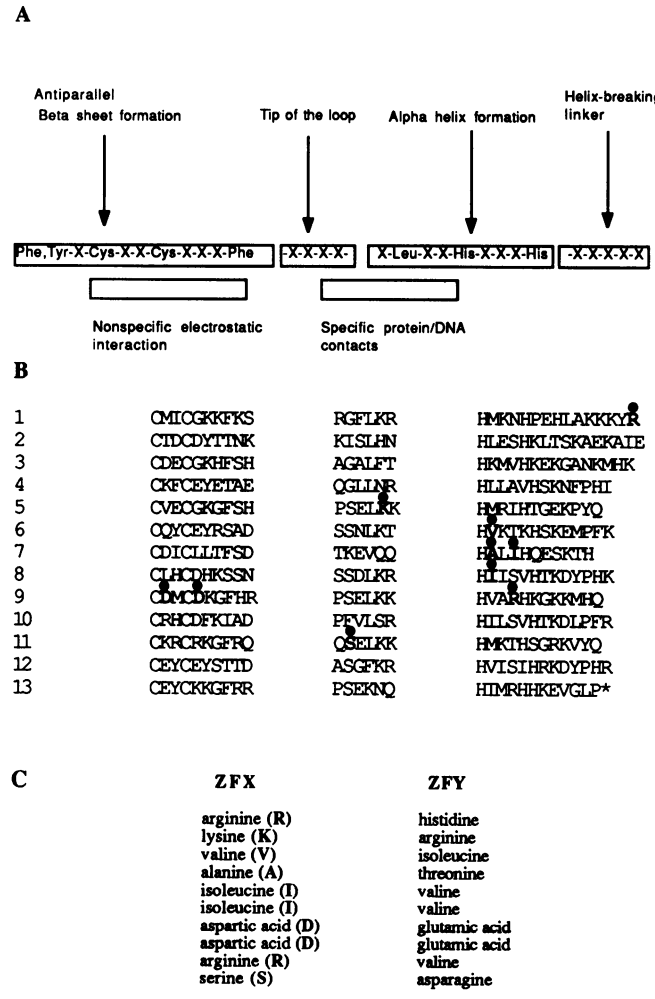

Figure 3. Section A: The postulated secondary structure motifs and functional domains present within the consensus C_2H_2 zinc finger, following Berg (16). Domains expected to be involved in sequence-specific interactions and in non-specific electrostatic interactions are boxed. Section B: The zinc-finger domain of ZFX is listed with fingers numbered 1 to 13 from N-terminus to C-terminus. Positions of amino acid differences between ZFX and ZFY are highlighted in bold and with a dot. Over the amino acid domain postulated to be involved in sequence-specific interaction only two amino acid differences are identified and both are substantially conservative substitutions (lysine to arginine and serine to asparagine). Section C: A table of amino acid differences between ZFY and ZFX over the zinc-finger domain, proceding from N-terminus to C-terminus.

events which have been reported by other workers during isolation of ZFY and ZFX cDNAs (12).

Sequence analysis of ZFY and ZFX

The coding regions of ZFY and ZFX defined by sequence analysis of the clones described above are in close agreement with the findings of other groups (11,12,13) and will not be presented in detail. The schematic in figure 1 indicates the location of the different domains of these genes in relation to the clones forming the basis of our sequence analysis. Briefly these are: 1) An acidic domain with a net negative charge and a propensity for alpha helical formation, shown to be capable of transcriptional

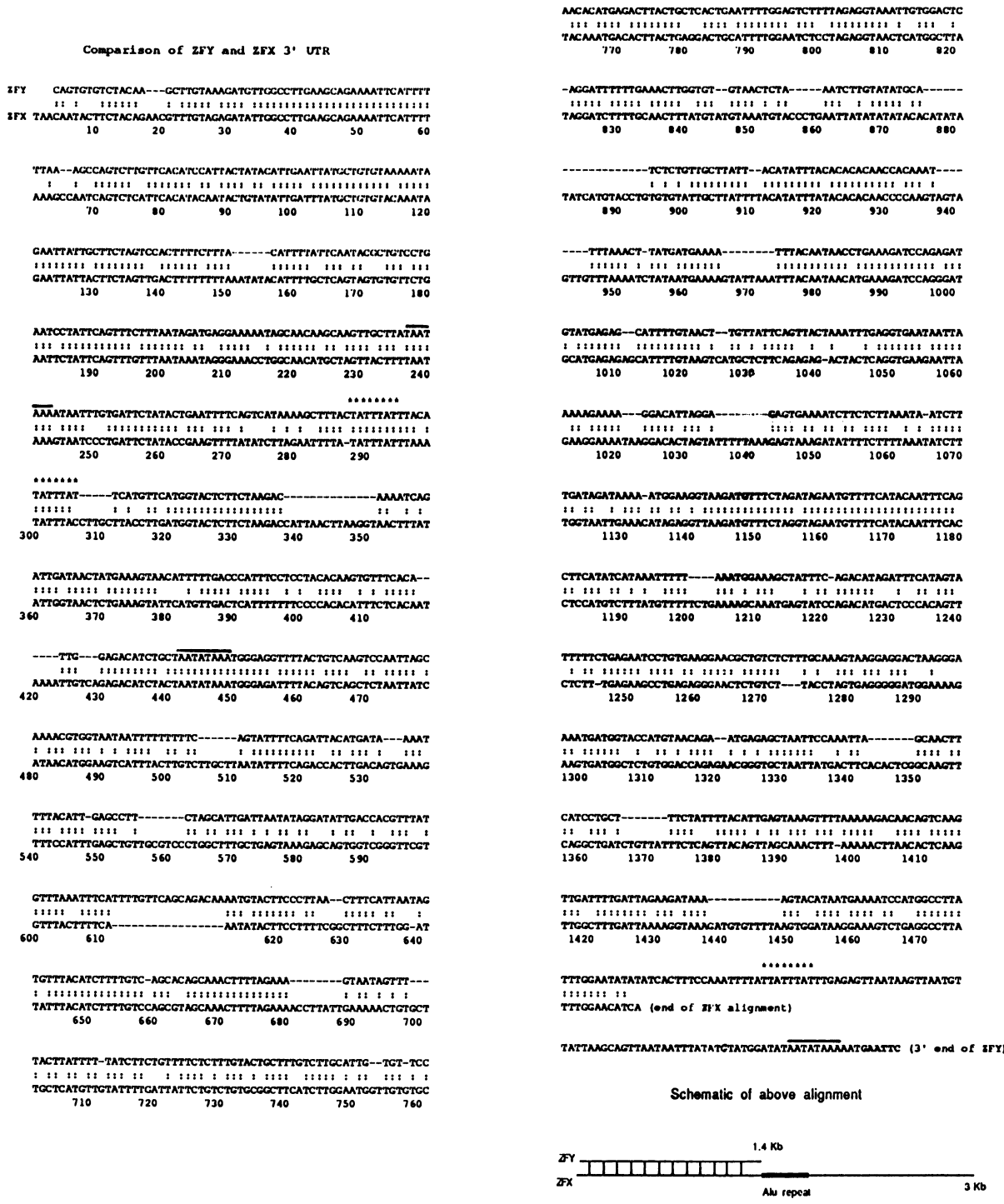

Figure 4. An optimal global alignment between the 1434 bp 3′ untranslated region of ZFY and the 3 Kb 3′ untranslated region of ZFX. Over the region of alignment, the two sequences are about 64% identical. Apart from this homology and a homology to the Zfx 3′ untranslated region, no further significant matches have been identified in database searches with the ZFY 3′ untranslated region. The domain is rich in A and T residues (71%) and A/T-rich clusters conforming closely to the consensus TTATTTAT (postulated by Kruys et al. to be involved in reduction of message stability and possibly efficiency of translation (19)) are starred. The AATATAAA sequences we implicate in directing polyadenylation are high-lighted by horizontal bars as is the AATAAA motif described by Lau et al. (12).

activation (14). 2) A cluster of positively charged amino acid residues immediately N-terminal to the zinc finger domain which may act as a nuclear localisation signal for these proteins. 3) The zinc finger DNA binding domain. Unlike the findings of others, however, we report that some ZFY transcript contain a 1.4Kb 3′ untranslated region (3′UTR) possessing sequences which may have a function in post-transcriptional control (described below).

Overall there is 95% homology between the coding regions of ZFY and ZFX. In the zinc-finger domain we find 60 positions of non-identity with no insertions, deletions or rearrangments. Further, of these 60 changes 50 do not alter the amino acid residue present. Of the remaining 10 amino acid substitutions (listed in figure 3C) 8 are very likely to be conservative. Figure 3B shows the 13 zinc fingers found in ZFX with the amino acid residues differing from ZFY high-lighted in bold and marked above with a dot.

Modelling studies, substantially confirmed by NMR experiments (15), with the consensus C_2H_2 zinc finger indicate that different regions of the motif can fold into distinct secondary structures and these are delineated in figure 3A. According to a model proposed by Berg, sequence-specific interactions with DNA only involve the amino acid residues at the tip of the loop and the start of the alpha helix and is consistent with recent studies in which sequence-specific interactions were altered by site-directed mutagenesis of residues in this domain (16,17). Only two of the amino acid differences between ZFX and ZFY (both conservative changes) occur in the region likely to be involved in sequence-specific interactions and may not differentiate significantly the nucleic acid binding specificities of the ZFX and ZFY proteins.

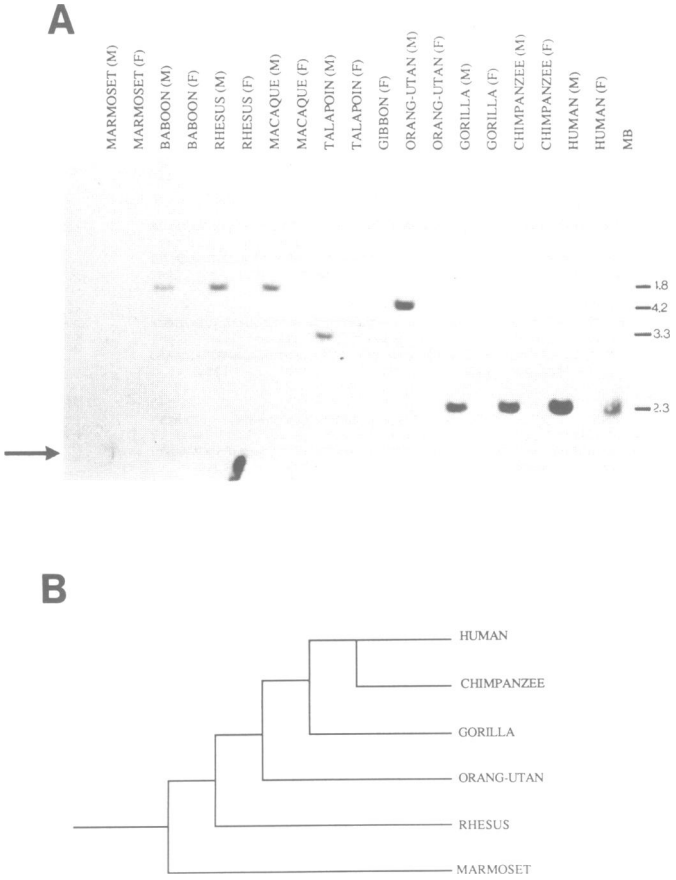

Figure 5. Section A shows a 'zoo-blot' prepared from TaqI-digested male and female primate DNA, hybridised for 48 hours at 37°C with the 3′ untranslated region of ZFY and subsequently washed in 1×SSC; 0.1% (w/v) SDS at 40°C and exposed for 18 hours. Male lanes display a hybridising fragment (only very faintly visible in the marmoset after such a short exposure, hence the position of the band is arrowed). The primates are listed from left to right in expected order of evolutionary distance from the human (as can be seen from the 'family tree' displayed in section B and taken from Koop et al., 1986 (25)). Patient MB is male.

The 3′ untranslated region of ZFY

The two ZFY transcripts (L24 and CMPXY3) show an extensive 3′UTR which harbours a number of interesting features. Figure 4 shows the alignment of the ZFY 1.4 Kb 3′UTR against that of ZFX published by Schneider-Gadicke et al.(13), which contains an Alu sequence and is approximately 1.6 Kb greater in length . Despite being non-coding, these sequences share 64% homology (insufficient to permit cross-hybridisation) over the first 1.4 Kb, which may reflect an important functional role for segments of this region of the mRNA molecule. Figure 5 indicates that the ZFY 3′ UTR is conserved in primates since it identifies a homologous sequence (under low stringency hybridisation and wash conditions) only in males. No hybridisation has been observed in non primate species such as mouse (not shown). Amongst the primates, the degree of hybridisation decreases as the evolutionary distance from human increases, so that in male marmoset only a faint band can be detected. This contrasts markedly with ZFX, where under more stringent conditions the 3′ UTR cross hybridises strongly with primate and mouse DNA (where it shares 83% homology with the 3′ UTR of Zfx) (14), indicating that it is conserved to a greater degree.

ZFY contains two potential non-consensus (AATATAAA) and several canonical (AATAAA) polyadenylation signals. Lau et al.

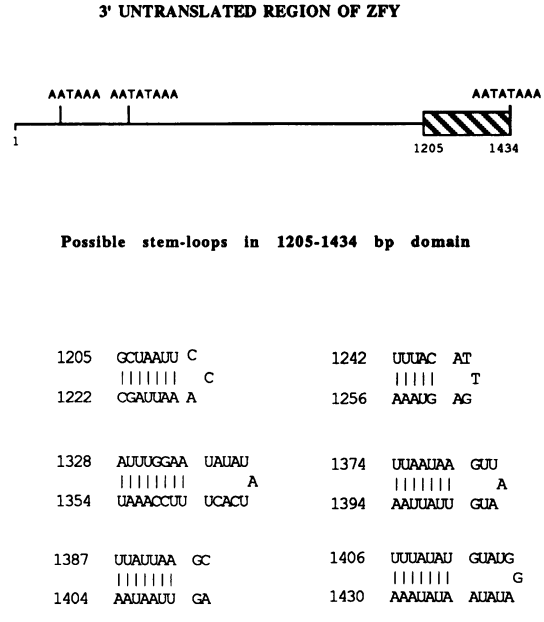

Figure 6. The Genetics Computer Group (GCG) 'stemloop' program calculates the most stable stem-loops present in a given sequence by scoring matches over a defined scanning window with G-T, A-T and G-C base pairing worth 1,2 and 3 respectively. Mismatches are scored as zero. A cluster of relatively stable stem-loops is located in the 3′ most 229 bp of the ZFY 3′ UTR and these structures are shown schematically above.

(12) indicate that the canonical AATAAA site (high-lighted with a horizontal bar in figure 4) is used to direct polyadenylation of ZFY transcripts found in adult testis. We show below that the other two non-consensus sites (also high-lighted with a horizontal bar in figure 4) are also used to terminate ZFY transcripts expressed in adult testis.

Two further features of the ZFY 3'UTR are of interest. First, the region is rich in A and T residues (37% T, 34% A, 14% C and 15% G) which often occur as clusters. Such structures (including T-rich stretches and sequences with close consensus to the octanucleotide TTATTTAT—starred in figure 4) have been implicated in reduction of stability and possibly inhibition of translation of a variety of transcripts such as c-fos, c-myc, lymphokine mRNAs and c-abl in post-meiotic germ cells (18,19,20,21,22). Second, sequence near the 3' end of the ZFY 3'UTR (see schematic in figure 6) is potentially capable of forming a series of stem-loop structures, of which the most energetically stable are shown. The 3' UTR of ZFX can form similar structures but these do not occur at a homologous position and share no homology with those formed in ZFY. The function of these motifs is not known.

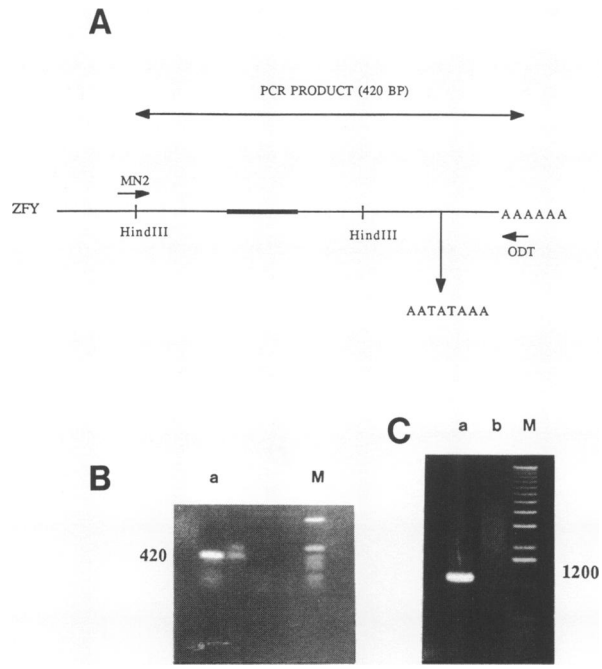

Figure 7. Section A is a schematic illustration of the use of primers MN2 (5' AACAGTGTGTCTACAAGCTT 3') and ODT (5' AATGAGCTC$(T)_{18}$ 3') to amplify a novel ZFY transcript with polyadenylation directed by the non-consensus motif AATATAAA some 400 bp 3' to the end of the long open reading frame. Cycling conditions were 94°C, 60 seconds; 55°C, 60 seconds; 72°C, 180 seconds (see methods). Section B (lane a) shows the result of 35 cycles of PCR amplification with primers MN2 and ODT and 5 μl of library template (equivalent to approximately 5×10^9 bacteriophage). The 420 bp PCR product produced could be digested with HindIII to give fragments of approximate size 270 bp and 150 bp and sequencing the larger HindIII fragment confirmed that the PCR product had been amplified from the ZFY 3' UTR. Section C (lane a) is a positive control and shows the result of 25 cycles of PCR amplification with AMD1 (5' GAGCTCGCAATAATTATTGGCCCTGAT 3') and AMD2 (5' GAGCTCCATTATGTGCTGGTTCTTTT 3') from 5 μl of the same library template. These primers allow amplification of the ZFY and ZFX zinc-finger domains (at 1.2 Kb). Section C (lane b) is a negative control in which the library template was substituted for DNA prepared from a high titre stock of bacteriophage lambda strain NM1149 and the PCR reaction with AMD1 and AMD2 repeated. Cycling conditions for AMD1/AMD2 was the same as listed for MN2/ODT.

Differential polyadenylation of ZFY transcripts

We have not found a poly(A) tract associated with the 3'end of the ZFY transcripts and thus, in theory, it is possible that a large amount of 3'UTR remains to be determined. This is unlikely in that (a) two independent clones (L24A and CMPXY3) end at the same point 3' to the octanucleotide AATATAAA, and that (b) in mouse Zfy-1 and Zfy-2 transcripts, the poly(A) tract has been found 13 bp downstream from an identical non-consensus AATATAAA motif, thus indicating that it can function to direct polyadenylation (14)

We examined whether the second (more 5') AATATAAA sequence, or further non-consensus motifs, directed polyadenylation in adult testis by PCR amplification from a testis cDNA library preparation with the oligonucleotides MN2 (corresponding to ZFY sequence at the 5' end of the 3' UTR) and ODT (complementary to poly(A) tracts). Following amplification under the conditions described in figure 7, a 420 bp fragment was synthesised, allowing a poly(A) tract to be mapped approximately 20 bp 3' to the AATATAAA shown in figure 8 (and high-lighted in figure 4). The sequence of the PCR product (not shown) corresponded to that of the ZFY 3'UTR and it is therefore likely that both AATATAAA non-consensus motifs described above direct ZFY polyadenylation in adult testis.

DISCUSSION

In this paper, we have been able to demonstrate that (a) despite the high degree of conservation between ZFX and ZFY coding regions, the size of their respective introns has altered but the points of interruption are conserved implying strongly that they have arisen from a common ancestral gene; (b) that there is differential polyadenylation of ZFY transcripts directed by non-consensus poly A addition sites; (c) that it is probable that the resulting differing 3' UTR regions confer differential stability since sequences similar to those found in the 3' UTR of large ZFY transcripts are associated with increased turnover in other mRNAs; (d) that the 3' UTR of ZFY is less conserved than that of ZFX and appears to be conserved only amongst the primates. In view of the observation that no long 3' UTRs have been identified in any of the reported mouse Zfy-1 and Zfy-2 cDNA clones, this latter point may indicate that the ZFY 3' UTR arose by duplication and transposition of the ZFX 3' UTR after divergence of mouse and primate lineages.

The long 3'UTR of ZFY contains sequences which have been implicated not only as playing a role in increased mRNA turnover, but also possibly in inhibition of translation (18,19,21,22). The small 3Kb ZFY transcript found specifically in testis has been correlated with germ cell numbers (12), implying that its expression is limited to this lineage. Thus removal in germ cells of a substantial portion of ZFY 3' UTR sequences from the 5.7 Kb transcript which, like Zfx in the mouse and ZFX in humans is found expressed in many cell types, may be an important post-transcriptional control in the expression of the ZFY protein product. Consequently, ZFY may have a function specifically in germ cells, a possibility which is further underscored by the post-meiotic expression of Zfy during mouse spermatogenesis (7). Clearly, it will be important to examine the fate of ZFY 3'UTR sequences during spermatogenesis in order to determine at what stages of germ cell differentiation the different polyadenylation signals are utilized.

It is also possible that the 5' UTR contributes also to translational control since both the human ZFX gene and the

mouse Zfy-1 gene have several AUG codons upstream of the start of the long open reading frame (13,3). It has been shown that such upstream initiation codons can result in translational repression (23,24). Thus both the 3′ and 5′ untranslated regions may influence mRNA activity.

In spite of their highly conserved putative DNA-binding domains, it is still unclear whether ZFY and ZFX are functionally equivalent both because of the potentially complex post-transcriptional controls which may be present and because of the observation that Zfx and Zfy-2 stimulate transcription to quite different extents in spite of having broadly similiar acidic domains (14). Thus, although expression of Zfy-1 and Zfy-2 and the removal of most of the ZFY 3′ UTR appears to be substantially restricted to maturing germ cells in the adult, it remains to be determined what role these genes play in this cell type. Further, it is not clear what function the expression of Zfy-1 and Zfy-2 serves in the developing mouse embryo.

ACKNOWLEDGEMENTS

We are grateful to Marie Ferguson-Smith and Marion Perryman for the culture of cell lines used in this study. We would like to acknowledge the assistance of Zanna Ulinowski in the construction of the 3E7 cosmid library. We thank Collin Bishop and Nacer Abbas for isolating clone L24 with our probe CMPXY1 and subsequently making it available to us. This work was supported by grants from the Medical Research Council. MN was supported by the Department of Education for Northern Ireland.

REFERENCES

1. Sinclair, A., Foster, J., Spencer, J., Page, D., Palmer, M., Goodfellow, P.and Graves, J. (1988) Nature **336**, 781–783.
2. Ferguson-Smith, M.A. and Affara, N.A (1988) Phil. Trans. R. Soc. Lond. **B 322**, 133–144.
3. Koopman, P., Gubbay. J., Collignon, J. and Lovell-Badge, R. (1989) Nature **342**, 940–942.
4. Palmer, M., Sinclair, A., Berta, P., Ellis, N., Goodfellow, P., Abbas, N., Fellous, M. (1989) Nature **342**, 937–939.
5. Sinclair, A., Berta, P., Palmer, M., Hawkins, J., Griffiths, B., Smith, M., Foster, J., Frischauf, A-M., Lovell-Badge, R. and Goodfellow, P. (1990) Nature **346**, 240–244.
6. Koopman, P., Munsterberg, A., Capel, B., Vivian, N. and Lovell-Badge, R. (1990) Nature **348**, 450–452.
7. Nagamine, C., Chan, K., Hake, L. and Lau, Y. (1990) Genes and Development **4**, 63–74.
8. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual (Cold Spring Harbour laboratory, Cold Spring Harbour, NY).
9. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463–5467.
10. Affara, N., Chambers, D., O'Brien, J., Habeebu, S., Kalaitsidaki, M., Bishop, C. and Ferguson-Smith, M.A. (1989) Nucl. Acids Res. **17**, 2987–2999.
11. Palmer, M., Berta, P., Sinclair, A., Pym, B. and Goodfellow, P. (1990) Proc. Natl. Acad. Sci. USA **87**, 1681–1685.
12. Lau, C. and Chan, K. (1989) Am. J. Hum. Genet. **45**, 942–952.
13. Schneider-Gadicke, A., Beer-Romero, P., Brown, L., Mardon, G., Luoh, S. and Page, D. (1989) Nature **342**, 708–711.
14. Mardon, G., Luoh, S-W., Simpson, E., Gill, G., Brown, L. and Page, D. (1990) Mol. Cell Biol. **10**, 681–688.
15. Lee, M., Gippert, G., Soman, K., Case, D. and Wright, P. (1989) Science **245**, 635–637.
16. Berg, J. (1988) Proc. Natl. Acad. Sci. USA **85**, 99–102.
17. Nardelli, J., Gibson, T., Vesque, C. and Charnay, P. (1991) Nature **349**, 175–178.
18. Kruys, V., Marinx, O., Shaw, G., Deschamps, J.and Huez, G. (1989) Science **245**, 852–855.
19. Kruys, V., Wathelet, M., Poupart, P., Contreras, R., Fliers, W., Content, J. and Huez, G. (1987) Proc. Natl. Acad. Sci. USA **84**, 6030–6034.
20. Shaw, G. and Kamen, R. (1986) Cell **46**, 659–667.
21. Wilson, T. and Treisman, R. (1988) Nature **336**, 396–399.
22. Oppi, C., Shore, S. and Reddy, E. (1987) Proc. Natl. Acad. Sci. USA **84**, 8200–8204.
23. Mueller, P. and Hinnebusch, A. (1986). Cell **45**, 201–207.
24. Liu, C., Simonsen, C. and Levinson, A. (1984) Nature **309**, 82–85.
25. Koop, B., Goodman, M. Xu, P., Chan, K. and Slighton, J. (1986) Nature **319**, 234–237.