Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1681–1685, March 1990
Genetics

Comparison of human *ZFY* and *ZFX* transcripts

(sex determination/"zinc fingers"/X chromosome/Y chromosome/polymerase chain reaction)

MARK S. PALMER, PHILIPPE BERTA*, ANDREW H. SINCLAIR, BARBARA PYM, AND PETER N. GOODFELLOW

Human Molecular Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, United Kingdom

Communicated by M. F. Lyon, November 16, 1989

ABSTRACT **ZFY is a candidate for the primary sex-determining gene (TDF, testis-determining factor) on the human Y chromosome. We have isolated cDNA clones of ZFY and its homologue on the X chromosome, ZFX. The transcripts of these genes are very similar to each other and encode predicted proteins of equal size. The conceptual amino acid sequence of both proteins contains an acidic domain, similar to the activation domain of transcription factors, and a potential nucleic acid-binding domain of 13 "zinc fingers." We have used the polymerase chain reaction to demonstrate the expression of ZFY and ZFX in a wide range of adult and fetal human tissues and to show that ZFX is expressed from the inactive X chromosome present in human–mouse hybrids.**

The primary signal for male sex determination in mammals is a testis-determining factor (1) encoded by *TDF*, a gene on the Y chromosome. Genetic analysis of sex-reversed individuals has been used to refine the chromosomal location of *TDF* to an interval of 140 kilobases (kb) on the short arm of the human Y chromosome (2–4). Sequences within this interval were found to be conserved on the Y chromosome in a variety of eutherian mammals and also cross-hybridized to sequences on the X chromosome. The nucleic acid sequence of a Y chromosome genomic fragment encoded an open reading frame whose predicted amino acid sequence was a tandem repeat of 13 "zinc-finger" domains. The Y chromosome gene encoding this zinc-finger protein was called *ZFY*, and the homologous sequence on the normal X chromosome was called *ZFX*. The zinc-finger motif has been associated with nucleic acid-binding proteins, and if *ZFY* is *TDF* this is consistent with the proposed regulatory function of *TDF*. Support for the equivalence of *ZFY* and *TDF* was obtained by studying the sex-determining region of mice (5–7). However, the finding of homologous genes in marsupials solely on autosomes, and not on the sex chromosomes, suggests that *ZFY* might not be the primary signal for sex determination (8). Alternatively, eutherians and metatherians may have evolved different mechanisms for sex determination.

We have isolated cDNA clones derived from transcripts of both *ZFY* and *ZFX* and we report their sequences here.† We used the nucleotide sequences to design primers for the polymerase chain reaction (PCR), which enabled us to look at their expression in different tissues and cell types.

MATERIALS AND METHODS

cDNA Libraries. Adult testis cDNA and HeLa cell cDNA libraries, both in bacteriophage λgt11, were obtained from Clontech. A cDNA library (pCD2Bassing) derived from a human foreskin fibroblast line was kindly provided by H. Okayama in the vector pCD2 (9).

Cell Lines. These are described in Figs. 2 and 6.

Southern Blotting. Endonuclease-digested DNA was electrophoresed in 0.8% agarose gels and transferred and fixed to Hybond-N+ filters (Amersham) (10).

Hybridization. (*i*) *Oligonucleotide probes*. Oligonucleotide 1818 was labeled with ^{32}P at its 5′ end with phage T4 polynucleotide kinase. Library filters were hybridized in 5× SSC/20 mM sodium phosphate, pH 7.0/10× Denhardt's solution/10% dextran sulfate/7% SDS containing oligonucleotide (10^6 cpm/ml) and heterologous DNA (100 μg/ml). (*ii*) *Plasmid probes*. Plasmid inserts were radiolabeled by the random primer method (11). Filters were hybridized in 5× SSPE/5× Denhardt's solution/0.5% SDS containing 0.5 × 10^6 cpm/ml. Final wash conditions were 0.2× SSC at 65°C. (SSC is 150 mM NaCl/15 mM sodium citrate, pH 7.0; Denhardt's solution is 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin; SSPE is 180 mM NaCl/10 mM sodium phosphate, pH 8.0/1 mM EDTA.)

Sequencing of cDNA Clones. cDNAs derived from λgt11 libraries or by PCR were subcloned into pBluescript vectors (Stratagene). Plasmids were sequenced as double-stranded DNA by using synthetic oligonucleotide primers and Sequenase (United States Biochemical).

Preparation of RNA. Total cellular RNA was extracted from cell lines and human tissues by the guanidinium thiocyanate method (12), and poly(A)$^+$ mRNA was isolated by oligo(dT)-cellulose column chromatography.

Northern Hybridization. mRNA (5 μg) was electrophoresed in a 0.8% agarose gel with Mops buffer and 2.2 M formaldehyde and then transferred to Hybond-N. The filter was hybridized at 46°C in 5× SSPE/50% (vol/vol) formamide/5× Denhardt's solution/0.5% SDS with probe at 0.5 × 10^6 cpm/ml. The filter was washed 15 min in 2× SSPE/0.1% SDS, 30 min in 1× SSPE/0.1% SDS at 42°C, and 15 min in 0.1× SSPE/0.1% SDS at room temperature.

PCR Amplification of cDNA. (*i*) *Specific first-strand cDNA synthesis*. RNA (5 μg) was reverse-transcribed (13), using avian myeloblastosis virus reverse transcriptase and synthetic oligonucleotide primers. (*ii*) *PCR*. This was done essentially as described (14) for 35 cycles, with annealing at 60°C.

Analysis of PCR-Amplified cDNA. Amplified products were extracted in chloroform and then precipitated with spermidine (15). Washed precipitates were digested with *Bam*HI and run in a 1% agarose gel containing ethidium bromide.

Oligodeoxynucleotides. The sequence of oligonucleotide 1818 (5′-AACAAGATGCATAAATGCAAATTCTGCGAA-TATGAGAC-3′) is from a region of the published genomic sequence of the *ZFY* zinc-finger domain (4) predicted to have minimum codon degeneracy; the sequence corresponds to nucleotides 1560–1597 of the *ZFY* transcript reported here.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PCR, polymerase chain reaction.

*Permanent address: Centre de Recherches de Biochimie Macromoleculaire, Centre National de la Recherche Scientifique LP 8402/Institut National de la Santé et de la Recherche Médicale U.249, rte de Mende, 34033 Montpellier Cedex, France.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M30607 and M30608).

Oligonucleotide 5242 (5'-CCCGAATTCGCTGTGACTGATGAGAATTAAAGGC-3') is identical with nucleotides 233–257 of *ZFX* plus an *Eco*RI linker at the 5' end. Oligonucleotide 3039 (5'-CCTCGACTTAAACTTCTTCC-3') is complementary to nucleotides 1306–1325 of *ZFY* and 1550–1569 of *ZFX*. Oligonucleotide 3043 has the same sequence as oligonucleotide 3039 but has an *Eco*RI linker at the 5' end. Oligonucleotide 5245 (5'-CATGATAGTGTAGTGGAAGCAGAAA-3') is identical with nucleotides 486–510 of *ZFY* and 730–754 of *ZFX*. Oligonucleotide 5244 (5'-CCTCTCCTACGATCACTTCCATATA-3') is complementary to nucleotides 981–1005 of *ZFY* and 1225–1249 of *ZFX*.

RESULTS

Identification of Partial Transcripts of *ZFY* and *ZFX*. An adult testis cDNA library was screened with oligonucleotide 1818. A positive phage plaque was isolated and found to contain a 2.3-kb *Eco*RI insert. This was subcloned into pBluescript and called pMF-1 (Fig. 1). pMF-1 includes at its 3' end a sequence that is identical (except for one base difference) with the nucleotide sequence of the genomic zinc-finger domain entered into the EMBL data base (accession no. J03134) from nucleotide 73 of the genomic sequence through to its 3' end.

pMF-1 was used to screen a HeLa cDNA library (HeLa cells carry no Y chromosome). The 1.3-kb *Eco*RI insert of a positive phage was subcloned into plasmid (pPB, Fig. 1). pPB must contain a transcript from *ZFX* or a related transcript from an autosome. pPB lacks the zinc-finger domain but is homologous to the 5' end of pMF-1. It also extends further 5' than pMF-1 by about 400 bp. There is one continuous open reading frame in pPB with no evidence of a 5' untranslated region. Comparison of these sequences suggested that the first 52 bp at the 5' end of pMF-1 are inverted and probably arose as an artifact during the construction of the library.

Isolated cDNAs Hybridize to Restriction Fragments That Distinguish the X and Y Chromosomes. When DNA from males is probed with pMF-1, three bands are seen that are not present in female DNA. These bands are also present in DNA from hamster–human hybrid cells containing the human Y chromosome and are thus Y-specific. The remaining bands, which are present in both males and females, are also present in the hybrid cell line containing the human X chromosome and are therefore X-specific (Fig. 2). Mapping panels screened with pPB show stronger hybridization of the X-related bands than is observed with pMF-1; because pPB does not contain the zinc-finger domain, these data confirm that the 1.8-kb X band and the 3.8-kb Y band correspond to finger-encoding *Eco*RI fragments.

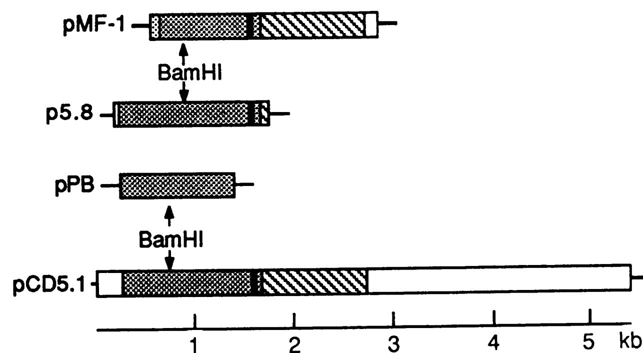

FIG. 1. Plasmids containing cDNA. Inserts are aligned for comparison. Shading represents distinct coding regions: open, untranslated region; hatching, zinc-finger region; black, potential nuclear localization signal; dark stippling, acidic region; light stippling, 52 base pairs (bp) at the end of pMF-1 that are inverted.

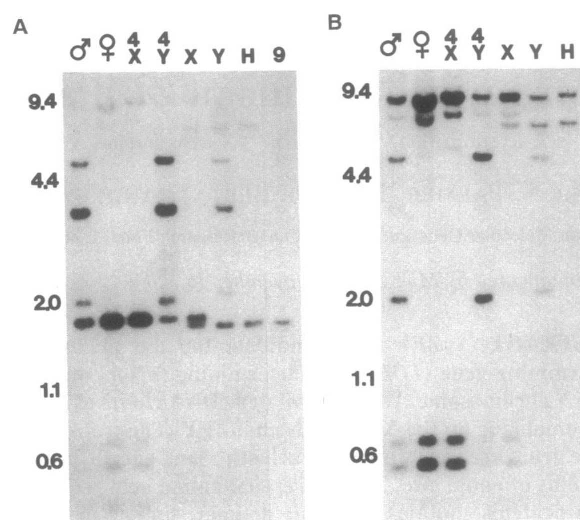

FIG. 2. Southern blot hybridizations. DNA was prepared from the following sources. Lanes ♂, male cell line PGF (16); lanes ♀, female cell line WT49 (17); lanes 4X, a 48XXXX cell line, GM1416B (Coriell Institute for Medical Research, Camden, NJ); lanes 4Y, a 49XYYYY cell line, Oxen (18); lanes X, a hamster–human hybrid cell line containing the human X chromosome, C12D (19); lanes Y, a hamster–human hybrid containing the human Y chromosome, 853 (20); lanes H, the hamster parent cell line, W3GH (19); lane 9, hamster–human hybrid containing human chromosome 9p, CF11-4 (21). Filters were screened with pMF-1 (*A*) or pPB (*B*). Numbers at left of autoradiographs are size markers (kb). Lanes do not all contain the same amount of DNA.

Affara *et al.* (22) isolated a transcript that contained the zinc-finger region but whose 5' sequence was shown, by *in situ* hybridization, to be derived from the short arm of chromosome 9 (9p). We used a cell line containing 9p to demonstrate that the zinc-finger sequences (and the 5' sequence present in pMF-1) do not hybridize to 9p (Fig. 2). This confirms the results of Müller and Schempp (23), who found by *in situ* hybridization that the zinc-finger domain hybridized only to Xp and Yp. We conclude that pMF-1 represents a partial transcript of *ZFY* and that pPB is derived from *ZFX*.

Identification of Transcripts Containing Long Open Reading Frames. A large cDNA containing a complete coding region of a *ZFX* transcript was identified by screening the pCD2Bassing library with pMF-1 as probe. The isolated clone, pCD5.1 (Fig. 1), contains an insert of 5.6 kb. This is similar to the size predicted for one of the RNA products seen by Northern analysis. The sequence of the first 3000 bp of pCD5.1 shows it to have a long open reading frame (2415 bp) preceded by a 5' untranslated region containing stop codons in all three reading frames. There are two adjacent stop codons 12 bp upstream from the translational initiation codon, and the initiation ATG is flanked by sequences showing a good consensus to those reported by Kozak (24). The sequence of pCD5.1 (Fig. 3 *Lower*) is identical with pPB along the entire length of their overlap and shows that pPB begins 4 bp before the proposed initiation codon.

As no clones containing the 5' end of *ZFY* were identified after screening several libraries, a PCR-based approach was used. The published sequences (5, 25) of transcripts of the mouse *Zfy-1* and *Zfy-2* genes show features similar to those of pCD5.1. They both have a stretch of nucleotides 5' to the initiation codon that encode two adjacent stop codons in frame with the first ATG. Due to the conservation of this region it was predicted that a similar sequence might exist in transcripts of *ZFY*. An oligonucleotide, 5242, identical to this region in pCD5.1 was synthesized for use in a PCR. RNA from a hamster–human hybrid, 853, containing the human Y

ZFY

GCTGTGACTGATGAGAATTAAAGGCC ATG GAT GAA GAT GAA TTT GAA TTG CAG CCA CAA GAG CCA AAC TCA TTT TTT GAT GGA ATA GGA GCT GAT GCT ACA CAC ATG GAT 110
GGT GAT CAG ATT GTT GTG GAA ATA CAA GAA GCA GTT TTT GTT TCT AAT ATT GTG GAT TCT GAC ATA ACT GTG CAT AAC TTT GTT CCT GAT GAC CCA GAC TCA GTT GTA 218
ATC CAA GAT GTT GTT GAA GAT GTT GTC ATA GAG GAG GAT GTT CAG TGC TCA GAT ATC TTA GAA GAG GCA GAT GTA TCT GAA AAT GTC ATC ATT CCT GAG CAA GTG CTG 326
GAC TCA GAT GTA ACT GAA GAA GTT TCT TTA CCA CAC TGC ACA GTC CCA GAT GAT GTT TTA GCT TCT GAC ATT ACT TCA ACC TCA ATG TCT ATG CCA GAA CAT GTT TTA 434
ACG AGT GAA TCC ATG CAT GTG TGT GAC ATT GGA CAT GTT GAA CAT ATG GTG CAT GAT AGT GTA GTG GAA GCA GAA ATC ATT ACT GAT CCT CTG ACG AGT GAC ATA GTT 542
TCA GAA GAA GTA TTG GTA GCA GAC TGT GCC CCT GAA GCA GTC ATA GAT GCC AGC GGG ATC TCA GTG GAC CAG CAA GAT AAT GAC AAA GCC AGC TGT GAG GAC TAC CTA 650
ATG ATT TCG TTG GAT GAT GCT GGC AAA ATA GAA CAT GAT GGT TCC ACT GGA GTG ACC ATC GAT GCA GAA TCA GAA ATG GAT CCT TGT AAA GTG GAT AGC ACT TGT CCT 758
GAA GTC ATC AAG GTG TAC ATT TTT AAA GCT GAC CCT GGA GAA GAT GAC TTA GGT GGA ACT GTA GAC ATT GTG GAG AGT GAA CCT GAA AAT GAT CAT GGA GTT GAA CTA 866
CTT GAT CAG AAC AGC AGT ATT CGT GTT CCC AGG GAA AAG ATG GTT TAT ATG ACT GTC AAT GAC TCT CAA CAA GAA GAT GAA GAT TTA AAT GTT GCT GAA ATT GCT GAT 974
GAA GTT TAT ATG GAA GTG ATC GTA GGA GAG GAG GAT GCT GCT GTT GCA GCA GCA GCA GCT GCT GTG CAT GAG CAG CAA ATT GAT GAG GAT GAA ATG AAA ACC TTC GTA 1082
CCA ATT GCA TGG GCA GCA GCT TAT GGT AAT AAT TCT GAT GGA ATT GAA AAC CGG AAT GGC ACT GCA AGT GCC CTC TTG CAC ATA GAT GAG TCT GCT GGC CTT GGC AGA 1190
CTG GCT AAA CAG AAA CCA AAG AAA AAG AGA AGA CCT GAT TCC AGG CAG TAC CAA ACA GCA ATA ATT ATT GGC CCT GAT GGT CAT CCT TTG ACT GTC TAT CCT TGC ATG 1298
ATT TGT GGG AAG AAG TTT AAG TCG AGG GGT TTT TTG AAA AGA CAC ATG AAA AAC CAT CCT GAA CAC CTT GCC AAG AAG AAG TAC CAC TGT ACT GAC TGT GAT TAC ACT 1406
ACC AAT AAG AAG ATA AGT TTA CAT AAC CAC CTG GAG AGC CAC AAG CTG ACC AGC AAG GCA GAG AAG GCC ATT GAA TGT GAT GAG TGT GGG AAG CAT TTT TCT CAT GCA 1514
GGG GCT TTG TTT ACT CAC AAA ATG GTG CAT AAG GAA AAA GGG GCC AAC AAA ATG CAC AAG TGT AAA TTC TGT GAA TAT GAG ACA GCT GAA CAG GGG TTA TTG AAT CGC 1622
CAC CTC TTG GCA GTC CAC AGC AAG AAC TTT CCT CAT ATT TGT GTG GAG TGT GGT AAA GGT TTC CGA TAC CCG TCG GAA CTG AGA AAG CAC ATG CGA ATC CAT ACC GGC 1730
GAG AAG CCA TAC CAA TGC CAG TAC TGT GAA TAT AGG TCT GCA GAC TCT TCT AAC TTG AAA ACA CAT ATA AAA ACA AAG CAT AGT AAA GAG ATG CCA TTC AAG TGT GAC 1838
ATT TGT CTT CTG ACT TTC TCA GAT ACC AAA GAA GTG CAG CAA CAT ACT CTT GTC CAC CAA GAA AGC AAA ACA CAT CAG TGT TTG CAT TGC GAC CAC AAG AGT TCA AAC 1946
TCA AGT GAT TTG AAA CGA CAT GTA ATT TCA GTT CAT ACG AAA GAC TAT CCT CAT AAG TGT GAG ATG TGC GAG AAA GGC TTT CAC AGG CCT TCA GAA CTT AAG AAA CAT 2054
GTG GCT GTC CAC AAA GGT AAA AAA ATG CAC CAA TGT AGA CAT TGT GAC TTT AAG ATT GCA GAC CCA TTT GTT CTA AGT CGC CAT ATT CTC TCA GTT CAC ACA AAG GAT 2162
CTT CCA TTT AGG TGT AAG AGA TGT AGA AAG GGA TTT AGG CAA CAA AAT GAG CTT AAA AAG CAT ATG AAG ACA CAC AGT GGC AGG AAA GTA TAT CAG TGT GAG TAC TGT 2270
GAG TAT AGC ACT ACA GAT GCC TCA GGC TTT AAA CGG CAC GTT ATT TCC ATT CAT ACA AAA GAC TAT CCT CAT CGG TGT GAG TAC TGC AAG AAA GGC TTC CGA AGA CCT 2378
TCA GAA AAG AAC CAG CAC ATA ATG AGA CAC CAT AAA GAA GTT GGT CTG CCC TAA CAGTGTGTCTACAAGCTTGTAAAGATGTTGGCCTTGAAGCAGAAAATTCATTTTTAAAGCCAGTCTTGTT 2503
CACATCCATTACTATACATTGAATTATGCTGTGTAAAAATAGAATTATTGCTTCTAGTCCACTTTTCTTTACATTTTATTCAATACGCTGTCCTGAATCCTATTCAGTTTCTTTAATAGATGAGTAAAAATAGCAACAAGCAA 2646
GTTGCTTATAATAAAATAATTTGTGATTCTCCCGAATTC 2685

ZFX

GTCCGTCCGGTTCTTGCTATATTGCCCCAGGCTGGTCTCGAACTCCTGGGCTCAAGCCGTTCTCCCGCCTCCCACTGCCCGCCTGTCACTGCCGTCTGTTCCCTGAGCTGTGCTTTACGCTGGGAAAGACATAGAAGAAC 140
TATTAAGAAGATAGAATTGTTTTGCTGCGCAGTACAGCAACAGTGGATGTTCAAGATTAAGATTAGAGTCAAGTTGTGTGATTAAGACAGGAGCTGTGACTGATGAGAATTAAAGGCC ATG GAT GAA GAT GGG CTT 276
GAA TTA CAA CAA GAG CCA AAC TCA TTT TTT GAT GCA ACA GGA GCT GAT GGT ACA CAC ATG GAT GGT GAT CAA ATT GTT GTG GAA GTA CAA GAA ACT GTT TTT GTT TCA 384
GAT GTT GTG GAT TCA GAC ATA ACT GTG CAT AAC TTT GTT CCT GAT GAC CCA GAT TCA GTT GTA ATC CAA GAT GTT ATT GAG GAC GTT GTT ATA GAA GAT GTT CAG TGC 492
CCA GAT ATC ATG GAA GAA GCA GAT GTG TCT GAA ACG GTC ATC ATT CCT GAG CAA GTG CTG GAC TCA GAT GTA ACT GAA GAA GTT TCT TTA GCA CAT TGC ACA GTC CCA 600
GAT GAT GTT TTA GCT TCT GAC ATT ACT TCA GCC TCA ATG TCT ATG CCA GAA CAC GTC TTG ACG GGT GAT TCT ATA CAT GTG TCT GAC GTT GGA CAT GTT GGA CAT GTT 708
GGA CAT GTT GAA CAT GTG GTT CAT GAT AGT GTA GTG GAA GCA GAA ATT GTC ACT GAT CCT CTG ACT ACC GAC GTA GTT TCA GAA GAA GTA TTG GTA GCA GAC TGT GCC 816
TCT GAA GCA GTC ATA GAT GCC AAT GGG ATC CCT GTG GAC CAG CAG GAT GAT GAC AAA GGC AAC TGT GAG GAC TAC CTT ATG ATT TCC TTG GAT GAT GCT GGC AAA ATA 924
GAA CAC GAT GGT TCT TCT GGA ATG ACC ATG GAC ACA GAG TCG GAA ATT GAT CCT TGT AAA GTG GAT GGC ACT TGC CCT GAG GTC ATC AAG GTG TAC ATT TTT AAA GCT 1032
GAC CCT GGA GAA GAT GAC TTA GGT GGA ACT GTA GAC ATT GTG GAG AGT GAG CCT GAG AAT GAT CAT GGA GTT GAA CTG CTT GAT CAG AAC AGC AGT ATT CGT GTT CCC 1140
AGG GAA AAG ATG GTT TAT ATG ACT GTC AAT GAC TCT CAG CCA GAA GAT GAA GAT TTA AAT GTT GCT GAA ATC GCT GAC GAA GTT TAT ATG GAA GTG ATC GTA GGA GAG 1248
GAG GAT GCT GCA GCA GCA CGG GCA GCC GCC GCC GTG CAC GAG CAG CAA ATG GAT GAC AAT GAA ATC AAA ACC TTC ATG CCG ATT GCA TGG GCA GCA GCT TAT GGT AAT 1356
AAT TCT GAT GGA ATT GAA AAC CGG AAT GGC ACT GCA AGT GCC CTC TTG CAC ATA GAT GAG TCT GCT GGC CTC GGC AGA CTG GCT AAA CAA AAA CCA AAG AAA AGG AGA 1464
AGA CCT GAT TCC AGG CAG TAC CAA ACA GCA ATA ATT ATT GGC CCT GAT GGA CAT CCT TTG ACT GTC TAT CCT TGC ATG ATT TGT GGG AAG AAG TTT AAG TCG AGA GGT 1572
TTT TTG AAA AGG CAC ATG AAA AAC CAT CCC GAA CAC CTT GCC AAG AAG AAA TAC CGC TGT ACT GAC TGT GAT TAC ACT ACC AAC AAG AAG ATA AGT TTA CAC AAC CAC 1680
CTG GAG AGC CAC AAG CTG ACC AGC AAG GCA GAG AAG GCC ATT GAA TGC GAT GAG TGT GGG AAG CAT TTC TCT CAT GCA GGG GCT TTG TTT ACT CAC AAA ATG GTG CAT 1788
AAG GAA AAA GGA GCC AAC AAA ATG CAC AAG TGT AAA TTC TGT GAA TAC GAG ACA GCT GAA CAA GGG TTA TTG AAT CGC CAC CTC TTG GCA GTC CAC AGC AAG AAC TTT 1896
CCT CAT ATT TGT GTG GAG TGT GGT AAG GGT TTT CGT CAC CCG TCA GAG CTC AAA AAG CAC ATG AGA ATC CAT ACT GGG GAG AAG CCG TAC CAA TGC CAG TAC TGC GAA 2004
TAT AGG TCT GCA GAC TCT TCT AAC TTG AAA ACG CAT GTC AAA ACT AAG CAT AGT AAA GAG ATG CCA TTC AAG TGT GAC ATT TGT CTT CTG ACT TTC TCG GAT ACC AAA 2112
GAG GTG CAG CAA CAT GCT CTT ATC CAC CAA GAA AGC AAA ACA CAC CAG TGT TTG CAT TGC GAC CAC AAG AGT TCG AAC TCA AGT GAT TTG AAA CGA CAC ATA ATT TCA 2220
GTT CAC ACG AAA GAC TAC CCC CAT AAG TGT GAC ATG TGT GAT AAA GGC TTT CAC AGG CCT TCA GAA CTC AAG AAA CAC GTG GCT GCC CAC AAG GGC AAA AAA ATG CAC 2328
CAG TGT AGA CAT TGT GAC TTT AAG ATT GCA GAT CCA TTT GTT CTA AGT CGC CAT ATT CTC TCA GTT CAC ACA AAG GAT CTT CCA TTT AGG TGC AAG AGA TGT AGA AAG 2436
GGA TTT AGG CAA CAG AGT GAG CTT AAA AAG CAT ATG AAG ACA CAC AGT GGC AGG AAA GTG TAT CAG TGT GAG TAC TGT GAG TAT AGC ACT ACA GAT GCC TCA GGC TTT 2544
AAA CGG CAC GTT ATT TCC ATT CAC ACG AAA GAC TAT CCT CAC CGG TGT GAG TAC TGC AAG AAA GGC TTC CGA AGA CCT TCA GAA AAG AAC CAG CAC ATA ATG CGA CAT 2652
CAT AAA GAA GTT GGC CTG CCC TAA CAATACTTCTACAGAACGTTTGTAGAGATATTGGCCTTGAAGCAGAAAATTCATTTTAAAGCCAATCAGTCTCATTCACATACAATACTGTATATTGATTTATGCTGTGTA 2787
CAAATAGAATTATTACTTCTAGTTGACTTTTTTTAAATATACATTTTGCTCAGTAGTGTGTTCTGAATTCTATTCAGTTGTTTAATAAATAGGGAAAACTGGCAACATGCTAGTTACTTTTAATAAAGTAATCCCTGATTC 2930
TATACCGAAGTTTTATATCTTAGAATTTTATATTTATTTAAATATTTACCTTGCTTACCTTGATGGTACT 3000

FIG. 3. Nucleotide sequences of *ZFY* and *ZFX* transcripts. Coding region is arranged in triplets. *ZFY* sequence is a composite of the cDNA clone pMF-1 and cDNA/PCR-derived material p5.8. Oligonucleotides used to generate p5.8 are underlined. The 5′ oligonucleotide sequence was determined from that present in pCD5.1 and so may not be identical to the sequence in *ZFY* transcripts. Potential transcriptional termination site is boxed. *ZFX* sequence is from the clone pCD5.1. Sequence 3′ to position 3000 has not been completed.

chromosome was reverse-transcribed using oligonucleotide 3039 as primer. The cDNA product was used as the template for PCR amplification using oligonucleotides 5242 and 3043. A band of 1300 bp was produced, corresponding to the size predicted for the 5′ end of *ZFY* by homology with *ZFX*. The amplified product was subcloned into pUC8.

The sequence identified, p5.8 (Fig. 1), was similar to the transcript of *ZFX* and was identical to pMF-1 in the overlapping region, except for the 52 bp of pMF-1 that are inverted. Direct sequencing of PCR-amplified material confirmed the fidelity of the new sequence in p5.8. The sequence of the *ZFY* transcript shown in Fig. 3 *Upper* is a composite from pMF-1 and p5.8.

The amino acid sequences that these *ZFY* and *ZFX* transcripts encode are remarkably similar to each other. Both contain an acidic domain (average pI 3.4) and a nucleic acid-binding domain of 13 zinc fingers. These two domains are separated by a short basic sequence, Pro-Lys-Lys-(Lys/Arg)-Arg-Arg-Pro (residues 400–406 in Fig. 4), that, by homology to sequences in the large tumor antigen of simian virus 40, has the characteristics of a nuclear localization signal (27). In the acidic domain there is 87% sequence identity between *ZFX* and *ZFY*, while in the zinc-finger domain this rises to 97%. These sequences are homologous to those reported for the genes *Zfy-1* and *Zfy-2* on the mouse Y chromosome (5, 25) and are compared in Fig. 4. The finger domain is more conserved between the human and mouse sequences: 79% compared to 70% for the acidic domain.

Adult Testis Expresses a Range of mRNA Transcript Sizes. Hybridization of pMF-1 to poly(A)$^+$ RNA from adult testis showed three bands of 3.0, 5.3, and 6.0 kb and a weaker band of 7.8 kb (Fig. 5). A broader region of hybridization around the 5.3-kb band was due either to partial degradation or to the presence of a range of transcripts of slightly different size. These sizes are consistent with previously published Northern blots (22, 28). Due to the low level of expression of *ZFY* and *ZFX*, the PCR was used on reverse-transcribed RNA to detect transcripts in a wide range of tissues.

Transcripts in Male and Female Tissues Can Be Distinguished by PCR. Comparison of the restriction maps of pMF-1 and pPB shows that *ZFY* and *ZFX* transcripts both contain a *Bam*HI site at their 5′ end, but that their relative positions are different. Material from this region amplified by PCR can be distinguished by *Bam*HI digestion and so the expression of X or Y transcripts in different tissues can be easily determined. To remove the possibility of artifacts arising due to genomic DNA contaminating cDNA samples, oligonucleotides were chosen that were known to span an intron. Oligonucleotides 5245 and 5244 flank the *Bam*HI site in the cDNAs, but we know from cosmid clones that we have

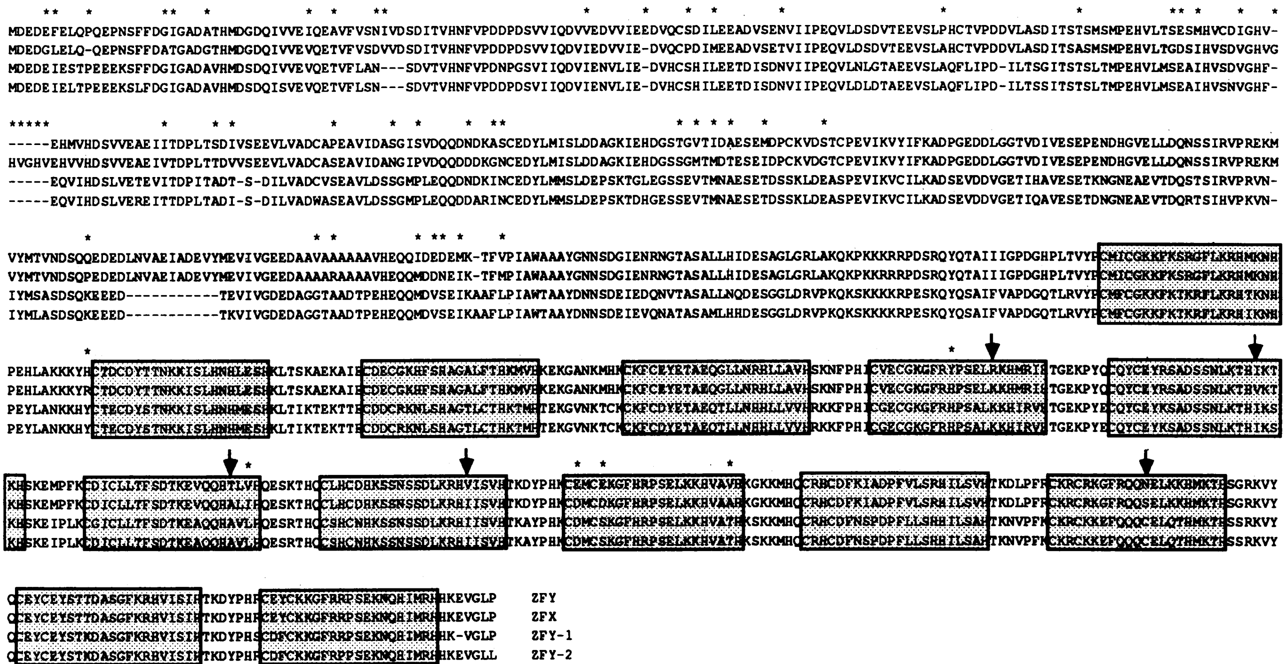

FIG. 4. Comparison of the amino acid sequences encoded by transcripts of human *ZFY* and *ZFX* and mouse *Zfy-1* and *Zfy-2*. Sequences are listed in that order. Residues not present in one or more sequences are indicated by dashes. Shaded boxes indicate the positions of the 13 zinc fingers. Residues that differ between the human *ZFY* and *ZFX* sequences are marked with asterisks. Arrows mark the 5 residues that differ in the fingers of the human *ZFY* and *ZFX* sequences, in positions suggested by Lee *et al.* (26) to be involved in specific base recognition.

isolated that they are separated by 20 kb in the genome. For those tissues tested, female adult and fetal tissues express *ZFX*, and male adult and fetal tissues express both *ZFX* and *ZFY* (Fig. 6). This situation is clearly distinct from that in the mouse, in which the expression of *Zfy-1* has been reported only in adult testis (7).

Evidence for Expression of *ZFX* from the Inactive X Chromosome. In normal female cells one X chromosome is inactivated so that the level of expression of X-linked genes is the same as that of males. If the products of *ZFY* and *ZFX* are interchangeable, then male cells would have twice as much product as female cells, unless *ZFX* escapes inactivation or another dosage-compensation mechanism exists. With the PCR approach, *ZFX* transcripts were found in somatic cell hybrids containing the inactive human X chromosome (Fig. 6).

Chelley *et al.* (33) reported that with PCR amplification of cDNA, the transcription of any gene can be detected in any cell. Although quantification through PCR is difficult, we have used only 5% of the RNA used by Chelley *et al.* and were able to detect transcripts on ethidium bromide-stained gels without having to use Southern hybridization.

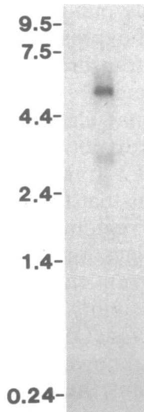

FIG. 5. Northern blot hybridization. mRNA from adult testis was probed with pMF-1. At left are positions and sizes (kb) of RNA size markers. The band at 7.8 kb is very faint on the photograph.

DISCUSSION

ZFY and *ZFX* can generate transcripts whose coding regions are the same size and specify remarkably similar protein products. The mRNAs for *ZFX* and/or *ZFY* that we find on Northern blots are 3.0, 5.3, 6.0, and 7.8 kb long. These are similar to the lengths reported by Schneider-Gädicke *et al.* (28), who found a 2.8-kb transcript particularly associated with adult testis. The sizes of the isolated transcripts are also consistent with the bands seen on Northern blots. pCD5.1 is 5.6 kb, while pMF-1 plus p5.8 add up to ≈3.0 kb. pMF-1 does not have a poly(A) tail but ends close to a potential polyadenylylation site. If this is the correct termination site and the *ZFY* transcript reported here is extended 5′ by the amount seen in pCD5.1, then a 3.0-kb transcript of *ZFY* can be accounted for. While the longer mRNA transcripts might encode different products, an alternative explanation is that they represent the use of different transcriptional termination sites. Clone pCD5.1 is a 5.6-kb cDNA with 3 kb of 3′ untranslated region. Within that part of the untranslated region that we have sequenced are three potential polyadenylylation sites (AATAAA). Long 3′ untranslated regions with multiple termination sites were reported for the human estrogen receptor (34) and glucocorticoid receptor (35), and in both cases polyadenylylated cDNAs using different termination sites were isolated. The different-size transcripts seen for *ZFY* and *ZFX* could all contain the same open reading frame and generate the same (X or Y) translated product but differ solely in the choice of termination site. We do have evidence, however, for differential splicing at the 5′ ends of both X and Y transcripts.

ZFY is expressed in several different male adult and fetal tissues, while *ZFX* is also expressed in all male and female, adult or fetal tissues tested. However, the cDNA-PCR approach is only semiquantitative and it is not possible to draw any firm conclusions about the precise levels of transcription.

We have also demonstrated the expression of *ZFX* from the inactive human X chromosome present in human–mouse hybrid cells, supporting the suggestion of Schneider-Gädicke

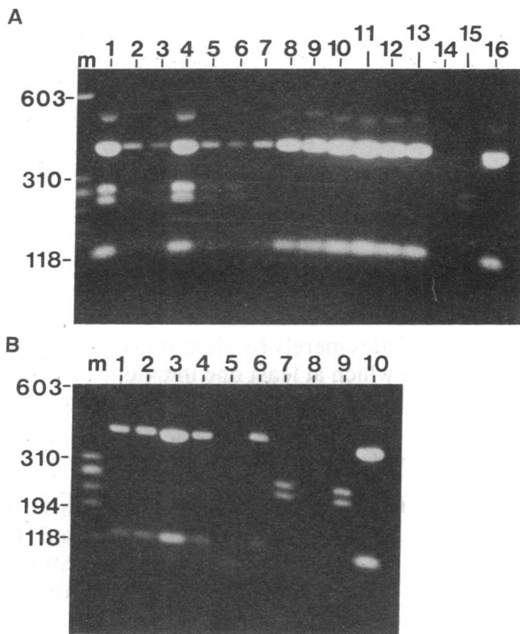

FIG. 6. Analysis of PCR-amplified cDNA products. Amplified products from both *ZFX* and *ZFY* transcripts give a 499-bp band. On digestion with *Bam*HI the *ZFY* product gives two bands of 231 and 268 bp. The *ZFX* product gives two bands of 100 and 399 bp. In both A and B, lane m shows *Hae* III-digested ϕX174 phage DNA (size markers; lengths in base pairs at left). (*A*) Tissue distribution. Lanes 1–6, male tissues: fetal brain, fetal gut, fetal testis, adult liver, adult lung, and adult testis, respectively. Lanes 7–13, female tissues: fetal liver, fetal lung, fetal gut, adult kidney, adult gut, adult muscle, and adult spleen, respectively. Lane 14, H_2O control. Lane 15, pMF-1, *ZFY* control. Lane 16, pPB, *ZFX* control. (*B*) Expression from inactive X chromosome. Lanes 1–4, human–mouse hybrid cell lines that contain the inactive human X chromosome as defined antigenically, and by karyotyping (29): 37-26-R-D, 70-62, 77-40b, and 11-10RA. Lane 5, mouse cell IRE3 (30). Lane 6, human–mouse hybrid containing active human X, HORL9X (31). Lane 7, human–mouse hybrid containing human Y, 3E7 (32). Lane 8, H_2O control. Lane 9, pMF-1, *ZFY* control. Lane 10, pPB, *ZFX* control.

et al. (28) that *ZFX* escapes inactivation on the X chromosome and is not dosage-compensated.

The close sequence relationship of *ZFY* and *ZFX*, the expression of *ZFX* from the inactive X chromosome, and the ubiquitous expression of transcripts of both genes have a number of functional implications for the biology of sex determination. The high percentage similarity of the zinc-finger domain (97% amino acid identity) suggests that both gene products would bind to the same or very similar DNA sequences. However, of the 11 differences between *ZFX* and *ZFY* that occur in the finger region, 5 fall on residues that Lee *et al.* (26) suggested might be involved in specific base recognition. One of these differences (amino acid 555 in *ZFY*, Fig. 3 *Upper*) does not occur in the sequence reported by Page *et al.* (4) and arises due to the single base difference (position 1689) between pMF-1 and the published sequence. Although the association of specific residues in zinc fingers with DNA sequences has not yet been demonstrated, the differences noted here might account for sex-specific DNA binding. The closely related acidic domains (87% amino acid identity) might also imply functional identity, although the activity of the proteins, presumably mediated through the acidic domain, may show subtle differences. Clearly the function of *ZFY* and *ZFX* cannot be resolved by sequence analysis alone. However, now that complete open reading frames have been identified for *ZFY* and *ZFX* transcripts, the potential role(s) of these genes in sex determination can be addressed through functional studies at the protein level.

We thank Dr. J. Trowsdale and Mr. V. Cunliffe for their advice and helpful comments. We are grateful to Dr. H. Okayama for providing us with a cDNA library, Drs. S. Povey and Y. Edwards for RNA samples, Drs. T. Mohandas and L. J. Shapiro for cell lines, and Mr. I. Goldsmith for synthesis of oligonucleotides.

1. Jacobs, P. A. & Strong, J. A. (1959) *Nature (London)* **183**, 302–303.
2. Vergnaud, G., Page, D. C., Simmler, M.-C., Brown, L., Rouyer, F., Noel, B., Botstein, D., de la Chapelle, A. & Weissenbach, J. (1986) *Am. J. Hum. Genet.* **38**, 109–124.
3. Guellaen, G., Casanova, M., Bishop, C., Geldwerth, D., Andre, G., Fellous, M. & Weissenbach, J. (1984) *Nature (London)* **307**, 172–173.
4. Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M. C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. & Brown, L. G. (1987) *Cell* **51**, 1091–1104.
5. Mardon, G., Mosher, R., Disteche, C. M., Nishioka, Y., McLaren, A. & Page, D. C. (1989) *Science* **243**, 78–80.
6. Nagamine, C. M., Chan, K., Kozak, C. A. & Lau, Y.-F. (1989) *Science* **243**, 80–83.
7. Mardon, G. & Page, D. C. (1989) *Cell* **56**, 765–770.
8. Sinclair, A. H., Foster, J. W., Spencer, J. A., Page, D. C., Palmer, M. S., Goodfellow, P. N. & Graves, J. A. M. (1988) *Nature (London)* **336**, 780–783.
9. Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. & Arai, K. (1987) *Methods Enzymol.* **154**, 3–28.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
11. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267.
12. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
13. Rappolee, D. A., Wang, A., Mark, D. & Werb, Z. (1989) *J. Cell Biochem.* **39**, 1–11.
14. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
15. Wallace, D. M. (1987) *Methods Enzymol.* **152**, 41–48.
16. Goodfellow, P. J., Nevanlinna, H. A., Gorman, P., Sheer, D., Lam, G. & Goodfellow, P. N. (1989) *Ann. Hum. Genet.* **53**, 15–22.
17. De Kretser, T. A., Crumpton, M. J., Bodmer, J. G. & Bodmer, W. F. (1982) *Eur. J. Immunol.* **12**, 600–606.
18. Bishop, C. E., Guellaen, G., Geldworth, D., Voss, R., Fellous, M. & Weissenbach, J. (1983) *Nature (London)* **303**, 831–832.
19. Westerveld, A., Visser, R. P. L. S., Meera Khan, P. & Bootsma, D. (1971) *Nature (London) New Biol.* **234**, 20–24.
20. Burk, R. D., Ma, P. & Smith, K. D. (1985) *Mol. Cell. Biol.* **5**, 576–581.
21. Hsu, L. C., Yoshida, A. & Mohandas, T. (1986) *Am. J. Hum. Genet.* **38**, 641–648.
22. Affara, N. A., Chambers, D., O'Brien, J., Habeebu, S. S. M., Kalaitsidaki, M., Bishop, C. E. & Ferguson-Smith, M. A. (1989) *Nucleic Acids Res.* **17**, 2987–2999.
23. Müller, G. & Schempp, W. (1989) *Hum. Genet.* **82**, 82–84.
24. Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8148.
25. Ashworth, A., Swift, S. & Affara, N. (1989) *Nucleic Acids Res.* **17**, 2864.
26. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635–637.
27. Boulikas, T. (1987) in *Int. Rev. Cytol. Suppl.* **17**, 493–684.
28. Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. & Page, D. C. (1989) *Cell* **57**, 1247–1258.
29. Goodfellow, P. N., Pym, B., Mohandas, T. & Shapiro, L. J. (1984) *Am. J. Hum. Genet.* **36**, 777–782.
30. Nabholz, M., Miggiano, V. & Bodmer, W. (1969) *Nature (London)* **223**, 358–363.
31. Goodfellow, P. N., Banting, G., Levy, R., Povey, S. & McMichael, A. J. (1980) *Somat. Cell Genet.* **6**, 777–783.
32. Marcus, M., Trantravali, R., Dev, J. G., Miller, D. A. & Miller, O. J. (1976) *Nature (London)* **262**, 63–65.
33. Chelley, J., Concordet, J.-P., Kaplan, J.-P. & Kahn, A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 2617–2621.
34. Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G. & Evans, R. M. (1985) *Nature (London)* **318**, 635–641.
35. Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P. & Chambon, P. (1986) *Nature (London)* **320**, 134–139.