31. Menkes, B. & Ilies, A. *Revue rom. embryol. Cytol.* **2**, 161–172 (1965).
32. Kapan, R., Traska, G. & Fuchs, E. *J. cell. biol.* **105**, 427–440 (1987).
33. Kapan, R. & Fuchs, E. *J. cell. Biol.* **109**, 295–307 (1989).
34. Green, S., Issemann, I. & Scheer, E. *Nucleic Acids Res.* **16**, 369–372 (1988).
35. Dollé, P. & Duboule, D. *Embo J.* **8**, 1507–1515 (1989).

ACKNOWLEDGEMENTS. We thank Drs D. Duboule and G. Morriss-Kay for discussions and critical reading of the manuscript, C. Werlé, A. Landmann and J. M. Lafontaine for illustrations, and the secretarial staff. This work was supported by the INSERM, the CNRS, the Fondation pour la Recherche Médicale and the Association pour la Recherche sur le Cancer. M.P. was supported by a fellowship from the FRMF.

Mutations in the *p53* gene occur in diverse human tumour types

Janice M. Nigro*, Suzanne J. Baker*, Antonette C. Preisinger*, J. Milburn Jessup†, Richard Hostetter†, Karen Cleary†, Sandra H. Bigner‡, Nancy Davidson*, Stephen Baylin*, Peter Devilee§, Thomas Glover‖, Francis S. Collins‖, Ainsley Weston¶, Rama Modali¶, Curtis C. Harris¶ & Bert Vogelstein*#

* The Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, Maryland 21231, USA
† M. D. Anderson Hospital, Houston, Texas 77030, USA
‡ Department of Pathology, Duke University, Durham, North Carolina 27710, USA
§ Department of Human Genetics, University of Leiden, Holland
‖ Department of Human Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA
¶ Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, Maryland 20892, USA

THE *p53* gene has been a constant source of fascination since its discovery nearly a decade ago[1,2]. Originally considered to be an oncogene, several convergent lines of research have indicated that the wild-type gene product actually functions as a tumour suppressor gene[3–9]. For example, expression of the neoplastic phenotype is inhibited, rather than promoted, when rat cells are transfected with the murine wild-type *p53* gene together with mutant *p53* genes and/or other oncogenes[3,4]. Moreover, in human tumours, the short arm of chromosome 17 is often deleted (reviewed in ref. 10). In colorectal cancers, the smallest common region of deletion is centred at 17p13.1 (ref. 9); this region harbours the *p53* gene, and in two tumours examined in detail, the remaining (non-deleted) *p53* alleles were found to contain mutations[9]. This result was provocative because allelic deletion coupled with mutation of the remaining allele is a theoretical hallmark of tumour-suppressor genes[11]. In the present report, we have attempted to determine the generality of this observation; that is, whether tumours with allelic deletions of chromosome 17p contain mutant *p53* genes in the allele that is retained. Our results suggest that (1) most tumours with such allelic deletions contain *p53* point mutations resulting in amino-acid substitutions, (2) such mutations are not confined to tumours with allelic deletion, but also occur in at least some tumours that have retained both parental 17p alleles, and (3) *p53* gene mutations are clustered in four 'hot-spots' which exactly coincide with the four most highly conserved regions of the gene. These results suggest that *p53* mutations play a role in the development of many common human malignancies.

To search for mutations, two approaches were used, both based on the polymerase chain reaction[12] (PCR) (Fig. 1). For tumour cell lines and for xenografts passaged in athymic nude mice, complementary DNA was generated from messenger RNA using oligo(dT) as a primer. A 1,300 base-pair (bp) fragment including the entire *p53* coding region was generated from the cDNA using PCR, and this fragment was cloned and sequenced in its entirety. For primary tumours, sufficient RNA was often not available for the first approach, and PCR was used to generate a 2.9-kilobase (kb) fragment from tumour DNA. This was the longest fragment that we could reproducibly amplify from the *p53* locus, and included all of the exons found to contain mutations through the first approach.

Using these approaches, we analysed *p53* sequences of tumours derived from the breast, lung, brain, colon or mesenchyme. Tumours of these types have been previously shown to exhibit frequent deletions of chromosome 17p when studied by restriction-fragment length polymorphism (RFLP) methods[10]. To test for allelic deletions, tumour DNA samples were digested with *Hin*fI and, following Southern transfer, hybridized sequentially to two probes (p144D6 (ref. 13) and pYNZ22.1 (ref. 14)) detecting variable-number tandem-repeat ('VNTR' or 'minisatellite') sequences. DNA samples from normal tissues exhibited two alleles with at least one of these probes in 29 of 31 different individuals tested. Because of this high degree of polymorphism, allelic loss could be assessed with greater than 95% certainty in cell lines and xenografts, even when corresponding normal tissue was not available for comparison.

Nineteen tumours with allelic deletions of chromosome 17p were selected for sequence analysis. Thirteen of the tumours were found to contain a single missense mutation; two tumours each contained two missense mutations; one tumour contained a frame-shift mutation at codon 293; and no mutation was detected in three tumours (Table 1). The PCR reaction is known to be associated with a relatively high rate of base misincorporation[12], and we confirmed this observation by noting several sequence variants (13 out of 34,000 bp sequenced) in individual clones that were not reproducibly present in other PCR reactions from the same tumour sample. All of the mutations listed in Table 1 were confirmed by performing a second PCR reaction and re-sequencing the products *en masse* as described in the legend to Fig. 2.

Two observations indicated that the nucleotide substitutions described in Table 1 represented somatic mutations. First, none of these presumptive mutations have been observed in the sequences of human *p53* genes derived from normal cells, SV40-transformed fibroblasts, or lymphoblastoid cell lines (ref. 15, and references therein). Second, in six cases (tumours 2, 3, 9, 12, 13, 16), normal tissue from the patients whose tumours are described in Table 1 were available for study. To test for the presence of the presumptive mutations in the germline of these patients, a strategy was devised that used both PCR and cloning. Although direct sequencing of PCR products has been shown to be possible by several methods, we found that none of the published methods could be reproducibly applied to all parts of the *p53* coding region. To circumvent this difficulty, we cloned the PCR products into a phagemid vector and used the DNA pooled from 10^3 to 10^4 independent phage clones as a template for DNA sequencing (see legend to Fig. 2). This procedure resulted in sequence data quality as high as that produced using individual plasmid DNA clones as templates, and was used to demonstrate that in each of the six cases noted above, the mutations in the tumour DNA were not present in the germline of the patient (examples in Fig. 2).

The data described above indicated that most tumours with one *17p* allele contained a mutation of the *p53* gene in the remaining allele. To begin to assess the status of tumours that had not lost a 17p allele, we examined cDNA clones from three such tumours. In each case, two cDNA clones derived from PCR products, generated as described in the legend to Fig. 1, were sequenced. In one case (tumour 11), both clones contained a single point mutation at codon 134 (Table 1). In the second case (tumour 16), one clone contained a point mutation at codon 281 and one clone was wild-type. In the third case (tumour 17), both clones were wild-type. To assess the relative expression levels of the mutant alleles, the sequencing strategy described

To whom correspondence should be addressed.

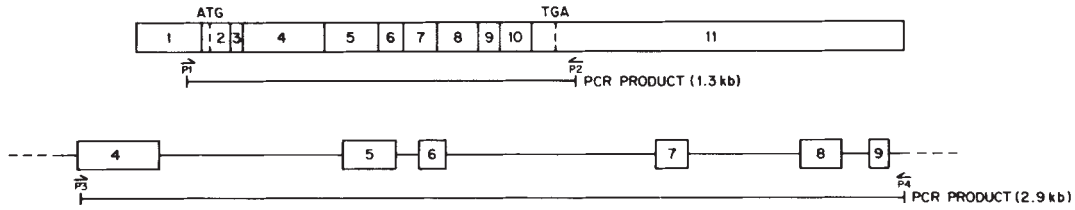

FIG. 1 Strategies for amplification of *p53* gene sequences. Messenger RNA was used to generate a cDNA template for a polymerase chain reaction (PCR) using primers P1 and P2 (top). The PCR product was 1.3 kb long and included the entire coding region. Alternatively, total genomic DNA was used in a PCR reaction using primers P3 and P4. The PCR product was 2.9 kb long and included exons 4–9 (bottom). The numbered boxes indicate exons and the vertical dotted lines indicate the start (ATG) and stop (TGA) codons respectively.

METHODS. RNA was purified using guanidium isothiocyanate and mRNA selected by binding to Messenger Affinity Paper (Amersham). Complementary DNA was synthesized from 500–750 ng of mRNA using oligo(dT) as a primer. The oligo(dT) primer was removed by isopropanol precipitation; 10 μg of transfer RNA and sodium perchlorate (to a final aqueous concentration of 0.5 M) were added to the reaction, and this was followed by addition of 1/2 volume of isopropanol[27]. The cDNA was pelleted by centrifugation for 15 min at room temperature and used in a 50-μl PCR reaction consisting of 35 cycles of 93 °C (1 min), 58 °C (1 min), and 70 °C (2 min). Genomic DNA (2 μg) was used in a 200-μl PCR reaction consisting of 30 cycles at 95 °C (1 min), 58 °C (1 min), and 70 °C (4 min). PCR reactions contained magnesium chloride at a final concentration of 2 mM. The primers used were P1, 5′-GGAATTCCACGACGGTGACACG-3′; P2, 5′-GGAATTCAAAATGGCAGGGGAGGG-3′; P3, 5′-GTAGGAATTCGTCCCAAGCAATGGATGAT-3′; P4, 5′-CATCGAATTCTGGAAACTTTCCACTTGAT-3′. All primers had extraneous nucleotides comprising *Eco*RI sites at their 5′ ends to facilitate cloning. The PCR products were digested with *Eco*RI, fractionated by electrophoresis, and following purification from agarose, ligated to *Eco*RI digested Bluescript vectors (Strategene). Individual clones were sequenced with primers derived from the p53 coding and intron sequences[15] using T7 polymerase.

in the legend to Fig. 2 was used using cDNA from tumour mRNA as a template. In tumour 11, only the mutant allele was expressed (data not shown); in tumour 16, the mutant and wild-type alleles were expressed at approximately equal levels (Fig. 2, panel 4).

The fact that three (tumours 5, 8 and 10) of the 19 tumours with single chromosome 17p alleles did not contain *p53* gene mutations has potentially important practical and conceptual implications. At least three explanations could account for this result. First, it is possible that mutations existed in these three tumours but were outside the region sequenced. Such mutations could affect the transcription of *p53* mRNA, its stability or translational capacity. Experiments to test this possibility are in progress, but it is notable that although tumour 5 expressed *p53* mRNA, no p53 protein could be detected on western blots (S. E. Kern and J.M.N., unpublished data). Alternatively, it is possible that a second tumour suppressor gene on chromosome 17p exists and was the target of allelic deletion in some tumours. Finally, the possibility that some allelic deletions represent 'nonselected losses', coincidently occurring with other independent mutations elsewhere in the genome, cannot be excluded.

Neurofibrosarcomas are tumours that predominantly occur in patients with neurofibromatosis type I. This syndrome can be inherited in an autosomal dominant fashion, and the gene responsible for it (*NF*) has been mapped to chromosome 17q near the centromere[16]. Interestingly, chromosome 17 sequences are lost from neurofibrosarcomas from these patients, but the region of deletion sometimes includes the short arm and does not include the region of 17q harbouring the *NF* gene (T. G. and B. Seizinger, unpublished data). The result shown in Table 1 (tumour 12) suggests that a target of allelic deletion, in at least one neurofibrosarcoma, was *p53*.

Altogether, 20 point mutations (19 missense, 1 frameshift) were identified in the present study. These are mapped in Fig. 4, together with the two human *p53* gene missense mutations previously described[9]. Several features are notable. Although the sample size is limited, the mutations tended to be clustered in four hotspots which accounted for 86% of the 21 missense mutations (five mutations in region A, codons 132–143; five mutations in region B, codons 174–179; three mutations in region C, codons 236–248; five mutations in region D, codons 272–281). There have been two missense mutations identified in murine tumour cells, both in the carcinogen-induced fibrosarcoma cell line Meth A: one allele contained a mutation in region A, and the other contained one mutation in region C and one mutation in region D[17,18]. Interestingly, the four hotspots for *in vivo* mutation coincided exactly with the four most highly conserved

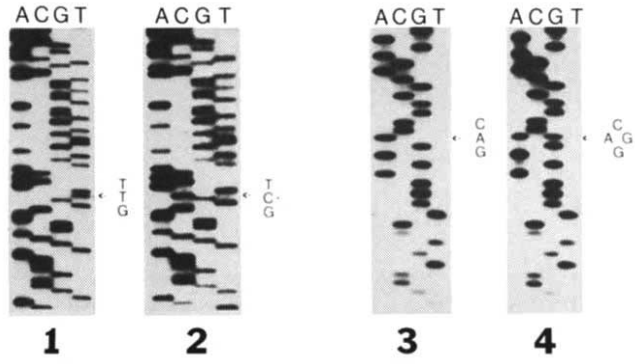

FIG. 2 Examples of sequencing reactions demonstrating *p53* gene mutations. The templates used for the sequencing reactions shown in panels 1–4 consisted of pools of greater than 10^3 clones generated from PCR products. Tumour 13 genomic DNA contained a mutation at codon 239 (antisense GCT, panel 2), instead of the wild-type sequence (GTT) found in the genomic DNA from normal lymphocytes from the same patient (panel 1). Panel 4 shows a sequencing reaction of pooled cDNA clones from tumour 16 showing that both wild-type codon 281 (GAC) and mutant codon 281 (GGC) were both expressed. Only the wild-type sequence (GAC) was found in pooled genomic DNA clones from normal lymphocytes of this patient (panel 3).

METHODS. PCR reactions were carried out as described in the legend to Fig. 1, and the reaction products digested with *Eco*RI. The entire reaction was ligated to 0.25 μg of λ ZAP phage vector arms (Stratagene) and packaged using 1/4 of a GIGA-Pack extract (Stratagene). *Escherichia coli* BB4 cells were then infected, and 10^3–10^4 phage clones plated on a 7-cm Petri dish. The λ ZAP vector contains the sequences for a phagemid into which the PCR inserts were cloned, and single-stranded DNA phage can be rescued from the λ phage clones using a helper phage[28]. An overnight culture of XL-1 Blue cells (Stratagene) was grown in 0.4% maltose and resuspended in 1.5 volumes of 10 mM magnesium sulphate. Phages were eluted from the 7-cm dish in 5 ml phage-dilution buffer (100 mM sodium chloride, 10 mM magnesium sulphate, 20 mM Tris, pH 7.5, 0.02% gelatin) for 2 h at room temperature with gentle agitation. Fifty microlitres of eluate was used to infect 200 μl of XL-1 Blue cells (Stratagene) in the presence of 1 μl helper phage R408 (10^{11} P.F.U. per ml). After 15 min at 37°, 5 ml of 2×YT broth was added and the culture shaken for 3 h at 37 °C, then heated to 70 °C for 20 min. Cell debris was pelleted at 3,000*g* for 5 min, and 10 μl supernatant, containing single-stranded DNA phage, was used to infect 200 μl of XL-1 Blue cells, prepared as described above. After 15 min at 37 °C, 100 μl of the mixture (containing over 10^4 clones determined by titration on XL-1 Blue cells) was inoculated into 50 ml L-broth and shaken overnight at 37 °C. Double-stranded DNA was isolated by alkaline lysis and sequenced as described in the legend to Fig. 1. The primer used for sequencing in panels 1 and 2 was 5′-GAGGCAAGCAGAGGCTGG-3′. The primer used for sequencing in panels 3 and 4 was 5′-TGGTAATCTACTGGGACG-3′.

TABLE 1 *p53* gene mutations in human tumours

Tumour	Tumour name	Tumour type*	Tumour cells tested†	Number of 17p alleles‡	Codon	Mutation Nucleotide	Amino acid
1	D263	Brain	B, X	1	175	CGC→CAC	Arg→His
2	D274	Brain	X	1	273	CGT→TGT	Arg→Cys
3	D303	Brain	B, X	1	216	GTG→ATG	Val→Met
4	D317	Brain	B, X	1	272	GTG→ATG	Val→Met
5	D247	Brain	C	1		None detected	
6	MDA 468	Breast	C	1	273	CGT→CAT	Arg→His
7	T47D	Breast	C	1	194	CTT→TTT	Leu→Phe
8	BT123	Breast	B	1		None detected	
9	1012	Lung	B	1	293	Deleted a G	Frameshift
10	5855	Lung	B	1		None detected	
11	H231	Lung	C	2	134	TTT→TTA	Phe→Leu
12	88-3/14	NFS	B, C	1	179	CAT→TAT	His→Tyr
13	C×4A	Colon	B, X	1	239	AAC→AGC	Asn→Ser
14	C×5A	Colon	X	1	248	CGG→TGG	Arg→Trp
15	C×6A	Colon	X	1	132	AAG→AAC	Lys→Asn
					133	ATG→TTG	Met→Leu
16	C×7A	Colon	B, X	2	281	GAC→GGC	Asp→Gly
17	C×19A	Colon	X	2		None detected	
18	C×20A	Colon	B, X	1	175	CGC→CAC	Arg→His
19	C×22A	Colon	X	1	175	CGC→CAC	Arg→His
20	C×26A	Colon	X	1	141	TGC→TAC	Cys→Tyr
21	SW480	Colon	C	1	273	CGT→CAT	Arg→His
					309	CCC→TCC	Pro→Ser
22	SW837	Colon	C	1	248	CGG→TGG	Arg→Trp

* The brain tumours were glioblastoma multiforme; the colon and breast tumours were adenocarcinomas, the NFS tumour was a neurofibrosarcoma developing in a patient with type-I neurofibromatosis; H231 was a small cell carcinoma of the lung, and the other two lung tumours were non-small-cell carcinomas.

† B, Tumour biopsy; C, cell line passaged *in vitro*; X, xenograft derived from biopsy, passaged in athymic nude mice. Whenever two sources of tumour cells are listed, both contained the indicated mutation.

‡ The number of alleles was determined by RFLP analysis, as described in the text.

regions of the *p53* gene, previously identified by Soussi *et al*[19]. Of the 41 amino acids contained within regions A–D, 93% are identical in the wild-type *p53* genes of amphibian, avian and mammalian species, compared to a conservation of only 51–57% over the entire *p53* coding sequence. The clustering of mutations and evolutionary conservation of regions A–D suggest that they play a particularly important role in mediating the normal function of the *p53* gene product.

Previous cytogenetic and RFLP studies have shown that allelic deletions of chromosome 17p occur in at least 60% of tumours of the colon, breast, lung, ovaries, cervix, adrenal cortex, bone and bladder, and in at least 30% of brain tumours[10,20–23]. These tumour types account for most of the neoplasms occurring in humans. Our data suggest that the great majority of tumours with 17p allelic deletions contain mutant *p53* genes, and similar mutations occur in at least some tumours without 17p allelic deletions.

These data lead to the following hypothesis. Mutations in the *p53* gene occur during the process of tumorigenesis, and through a dominant negative effect[3–5,9,24], cause tumour progression. The dominant negative effect may be mediated by binding of the mutant p53 product to the wild-type product, creating an inactive oligomeric complex[18,25]. Because wild-type products remain in the cell, however, a further loss of growth control can be exerted when the wild-type allele is deleted, leaving the cell with only a mutant allele. The first example of an intermediate step in this scheme is provided by tumour 16, which had not lost a chromosome 17p allele, but had developed a *p53* gene mutation and expressed both the wild-type and mutant alleles (Table 1 and Fig. 2). We imagine that tumour 16, had it not been surgically removed, would have eventually lost the wild-type *p53* gene through allelic deletion. Indeed, it has been shown that the loss of a chromosome 17p allele is significantly associated with tumour progression in the human host[26]. Although this model requires much further study before it can be verified, it is now supported by several lines of research, including the demonstration of mutant *p53* genes in most tumours with 17p allelic losses. □

Received 30 August; accepted 30 October 1989.

1. Lane, D. P. & Crawford, L. V. *Nature* **278,** 261–263 (1979).
2. Linzer, D. I. H. & Levine, A. J. *Cell* **17,** 43–52 (1979).
3. Finlay, C. A., Hinds, P. W. & Levine, A. J. *Cell* **57,** 1083–1093 (1989).
4. Eliyahu, D. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
5. Sturzbecher, H.-W., Addison, C. & Jenkins, J. R. *Molec. cell. Biol.* **8,** 3740–3747 (1988).
6. Munroe, D. G., Rovinski, B., Bernstein, A. & Benchimol, S. *Oncogene* **2,** 621–624 (1988).
7. Wolf, D. & Rotter, V. *Proc. natn. Acad. Sci. U.S.A.* **82,** 790–794 (1985).
8. Masuda, H., Miller, C., Koeffler, H. P., Battifora, H. & Cline, M. J. *Proc. natn. Acad. Sci. U.S.A.* **84,** 7716–7719 (1987).
9. Baker, S. J. *et al. Science* **244,** 217–221 (1989).
10. Cavenee, W. K., Hastie, N. D. & Stanbridge, E. J. eds *Current Communications in Molecular Biology: Recessive Oncogenes and Tumor Suppression* (Cold Spring Harbor Press, New York, 1989).

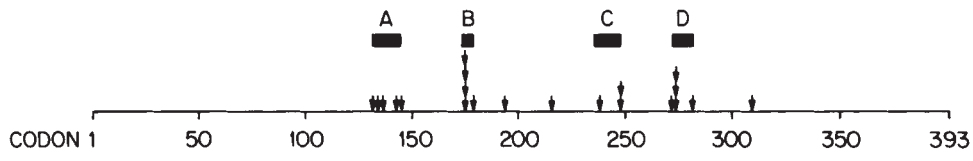

FIG. 3 Summary of *p53* point mutations in human cancer. Each of the missense mutations listed in Table 1 is indicated with an arrow. In addition, the two point mutations described previously[9] in human cancers (at codons 143 and 175) are also included. The four regions containing most (86%) of the mutations are indicated by the black bars marked A–D.

11. Knudson, A. *Cancer Res.* **45,** 1437–1443 (1985).
12. Saiki, R. K. *et al. Science* **239,** 487–491 (1988).
13. Kondoleon, S. *et al. Nucleic Acids Res.* **15,** 10605 (1987).
14. Nakamura, Y. *et al. Nucleic Acids Res.* **16,** 5707 (1988).
15. Buchman, V. L. *et al. Gene* **70,** 245–252 (1988).
16. Collins, F. S., Ponder, B. A. J., Seizinger, B. R. & Epstein, C. J. *Am. J. hum. Genet.* **44,** 1–5 (1989).
17. Finlay, C. A. *et al. Molec. cell. Biol.* **8,** 531–539 (1988).
18. Eliyahu, D. *et al. Oncogene* **3,** 313–321 (1988).
19. Soussi, T. *et al. Oncogene* **1,** 71–78 (1987).
20. Lee, J. H. *et al. Proc. Am. Ass. Cancer Res.* **30,** 442 (1989).
21. Atkin, N. B. & Baker, M. C. *Cancer Genet. Cytogenet.* **37,** 229–233 (1989).
22. Yano, T. *et al. J. natn. Cancer Inst.* **181,** 518–523 (1989).
23. Tsai, Y. C. *et al. Cancer Res.* (in the press).
24. Herskowitz, I. *Nature* **329,** 219–222 (1987).
25. Kraiss, S., Quaiser, A., Oren, M. & Montenach, M. *J. Virol.* **62,** 4737–4744 (1988).
26. Kern, S. E. *et al. J. Am. Med. Ass.* **261,** 3099–3103 (1989).
27. Haymerle, H. *et al. Nucleic Acids Res.* **14,** 8615–8624 (1986).
28. Short, J. M., Fernandez, J. M., Sorge, J. A. & Huse, W. D. *Nucleic Acids Res.* **16,** 7583–7600 (1988).

ACKNOWLEDGEMENTS. We thank Y. Nakamura, R. White and M. Litt for providing the RFLP probes and T. Gwiazda for preparation of the manuscript. The work was supported in part by The Clayton Fund, the McAshan Fund, The National Neurofibromatosis Foundation and the NIH.

Putative transcription activator with alternative isoforms encoded by human *ZFX* gene

Ansbert Schneider-Gädicke, Peggy Beer-Romero, Laura G. Brown, Graeme Mardon, Shiuh-Wen Luoh & David C. Page

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

THE *ZFY* gene in the sex-determining region of the human Y chromosome encodes a protein with 13 zinc fingers, and may determine whether an embryo develops as a male or female[1]. *ZFX*, a related gene on the human X chromosome, may also function in sex determination; it encodes a protein with a very similar zinc-finger domain and escapes X inactivation[1,2]. *ZFY* and *ZFX* diverged from a common ancestral gene before the radiation of placental mammals, and retain a similar genomic organization[2]. Analysis of complementary DNAs from the mouse Y-chromosomal homologues of *ZFY* indicates that these genes encode probable transcription activators[3,4]. Here, we report that *ZFX* encodes a protein composed of a highly acidic amino-terminal domain, a basic putative nuclear-localization signal, and a carboxy-terminal zinc-finger domain. This combination of features, also found in the *ZFY* gene product, is typical of transcription activators. Alternative splicing generates *ZFX* transcripts encoding isoforms of 575 and 804 amino acids. These ZFX protein isoforms differ in the length of their acidic domains and may be functionally distinct.

ZFX is transcribed in all human cells analysed[2]. We cloned *ZFX* cDNAs from a male lymphoblastoid cell line. Analysis by restriction mapping and hybridization with genomic probes revealed three distinct types of cDNAs. Whereas two types were represented by a single clone each (cDNAs 1 and 3), a third type was represented by three clones (cDNAs 2, 4 and 5).

We determined the nucleotide sequence of one cDNA of each type (Fig. 1). Comparison of cDNAs 1, 2, and 3 showed that alternative splicing and polyadenylation had produced structurally distinct 5′ untranslated, coding, and 3′ untranslated regions (Fig. 2*a*). Differential splicing involved an invariant donor site (nucleotide −378) and alternative acceptor sites (at nucleotides +647, −28 and −377, in cDNAs 1, 2 and 3, respectively). Each cDNA contains a single long open reading frame (ORF), and in each case the first ATG occurs in a sequence context favourable for initiation of translation[5]. Complementary DNAs 2 and 3, with identical ORFs, encode a protein of 804 amino acids (ZFX^{804}), whereas cDNA 1 encodes an isoform of 575 amino acids (ZFX^{575}). (Alternatively, with cDNA 1, translation initiation at a second ATG, whose context is highly favourable, would result in production of an isoform of 573 amino acids.) *In vitro* transcription and translation of cDNAs 3 and 1 yielded the predicted full-length and truncated proteins, respectively (data not shown). Both ZFX protein isoforms contain three domains—an N-terminal acidic portion (25% aspartic and glutamic acid), a small basic domain, and a C-terminal run of 13 zinc fingers, each with two cysteines and two histidines (Cys-Cys/His-His zinc fingers). The isoforms differ in that the acidic domain of ZFX^{575} is half that of ZFX^{804}.

Comparison of *ZFX* cDNAs with the genomic locus by restriction mapping and oligonucleotide hybridization gave an overview of the intron–exon organization (Fig. 2*a*). The C-terminal zinc-finger domain and 3′ untranslated region (UTR) are encoded by a single exon, whereas the N-terminal acidic domain is encoded by a minimum of four exons for ZFX^{575} (cDNA 1) and a minimum of six exons for ZFX^{804} (cDNAs 2 and 3). The cDNAs span 67 kilobases (kb) in the genome.

Of the three types of cDNAs, type 2, encoding ZFX^{804}, seems to be most representative of the 6.3- and 8-kb transcripts observed on northern blots[2]. First, the coding exon defined by oligonucleotides '150' and '637' is present in cDNAs 2 and 3 but not in cDNA 1 (Fig. 2*a*). Northern analysis using the corresponding genomic DNA fragment revealed that this exon is present in the 6.3- and 8-kb transcripts (not shown). Second, northern analysis indicated that the polyadenylation site used in cDNA 2 corresponds, at least roughly, to the 3′ end of the main *ZFX* transcripts (data not shown). Clone 2, which is 5.6 kb long, could represent a 6.3- or 8-kb transcript that is incomplete at the 5′ end.

Transcripts corresponding to cDNA 1, encoding ZFX^{575}, have not been detected by northern analysis. Using polymerase chain reaction (PCR) amplification, however, we confirmed the differential splicing predicted from cDNA analysis and crucial to the generation of ZFX protein isoforms (Fig. 2*b*). We designed splice-specific 5′ primers spanning splices of the invariant donor site at nucleotide −378 to alternative acceptor sites at nucleotides +647, −28 and −377, as in cDNAs 1, 2 and 3, respectively. The

FIG. 1 Composite nucleotide sequence of *ZFX* cDNAs 1, 2 and 3 (Fig. 2*a*) and predicted amino-acid sequences. Numbering of nucleotides and amino acids is with reference to the first in-frame ATG codon in cDNAs 2 and 3. Known splice sites are indicated by upward arrows. Alternative starts of poly(A) tails are indicated by downward arrows. Complementary DNA 1 begins at nucleotide −527, uses a splice donor site at −378 and an acceptor site at +647; a tail of 20 adenosines follows nucleotide 2,585. Complementary DNA 2 starts at nucleotide −426, uses the same donor, but a different acceptor at −28; a tail of 46 adenosine residues follows nucleotide 5,450. Complementary DNA 3 starts at nucleotide −426, uses the same donor, but yet another different acceptor at −377; a tail of 26 adenosines follows nucleotide 3,298. Complementary DNAs 2 and 3 contain the same ORF with the first in-frame ATG codon at nucleotide 1; the predicted protein of 804 amino acids is composed of an acidic N-terminal domain, a small basic domain (residues 391–406, boxed), and 13 zinc fingers (cysteines of the Cys-X-X-Cys repeats are circled). Complementary DNA 1 has a shorter ORF with the first in-frame ATG codon at nucleotide +688, encoding a protein of 575 residues. The 3′ UTR of cDNA 2 contains an *Alu* sequence[16] (nucleotides +3,959 to +4,250, boxed). An AATAAA polyadenylation signal[17] occurs 20 nucleotides 5′ of the poly(A) tail of cDNA 2. Similar sequences are present 5′ of the poly(A) tails in cDNAs 1 and 3.
METHODS. Complementary DNA libraries were prepared[3] using poly(A)$^+$ RNA prepared[2] from human male lymphoblastoid cell line WHT1659. Five million recombinant phages from unamplified libraries were screened using plasmids pDP1007, pDP1041 and pDP1006 (*ZFY* genomic fragments which cross-hybridize to *ZFX*)[1,2] as probes. Complementary DNA inserts of seven phages were subcloned into Bluescript vectors (Stratagene). The full DNA sequences of three inserts and partial sequences of the other four inserts were determined[2]. As judged by comparison with genomic sequences, five cDNAs originated from *ZFX* and two from *ZFY*.