Am. J. Hum. Genet. 52:192–203, 1993

A Novel Zinc Finger Gene on Human Chromosome 1qter That Is Alternatively Spliced in Human Tissues and Cell Lines

Mary Saleh,[1] Licia Selleri, and Glen A. Evans

Molecular Genetics Laboratory and Human Genome Center, The Salk Institute for Biological Studies, La Jolla, CA

Summary

DNA-binding proteins that share the conserved C_2-H_2 zinc finger motif have been shown to have important roles as transcriptional regulators of gene expression and have been implicated in several hereditary human diseases. In order to define potential candidate genes for inherited disorders characterized by aberrant gene expression, we utilized *Kruppel*-related sequences to isolate zinc finger–containing cDNAs. We isolated and characterized two novel zinc finger–encoding cDNAs from a human hepatoblastoma cell line, which demonstrate DNA sequence homology to a recently described human *Kruppel*-related gene HZF-3 and appear to be derived from a single gene by alternate mRNA splicing. This gene, denoted "HZF-16," gives rise to at least two gene products. One cDNA (i.e., HZF-16.2) has nine zinc finger domains, while alternative splicing of the message gives rise to a smaller product (i.e., HZF-16.1) that has four domains. Despite the internal splicing event, both the 5′- and 3′-untranslated sequences in both cDNAs are identical, as are the first three domains. In the HZF-16.1 cDNA, the fourth zinc finger domain is a fusion product of domains four and nine of HZF-16.2 and could potentially give rise to a new DNA-binding specificity. These alternatively spliced transcripts are differentially regulated in human tissues and transformed cell lines and show a different distribution of expression between human cell lines and normal human tissue. This novel gene was mapped to human chromosome 1q44 by chromosomal in situ suppression hybridization and thus represents a candidate gene for trisomy 1q syndrome and for several other disorders.

Introduction

One of the fundamental questions in developmental biology today concerns the understanding of the complex cascade of events that lead to the regulation of gene expression. An essential step in this pathway is the binding of regulatory proteins or transcription factors to specific DNA sequences. Many DNA-binding proteins that play a role in regulating gene transcription have been shown to share common structural motifs that are conserved even among highly divergent species. Thus, these proteins can be classified into several major classes based on the structure of the conserved motifs.

Received March 23, 1992; final revision received September 23, 1992.

Address for correspondence and reprints: Glen A. Evans, Molecular Genetics Laboratory, The Salk Institute for Biological Studies, P.O. Box 85800, San Diego, CA 92138.

[1]Present address: Prince Henry's Institute for Medical Research, Clayton, Victoria, Australia.

0002-9297/93/5201-0022$02.00

Proteins that share the helix-turn-helix motif represent one of the best-defined classes of DNA-binding proteins. This class of protein was originally described in prokaryotic repressors (Pabo and Sauer 1984) and later in eukaryotic proteins as the homeobox domain (Gehring 1987). A second class of sequence-specific DNA-binding protein has the conserved zinc finger motif that was initially identified in TFIIIA, a *Xenopus laevis* transcription factor required for initiation of 5S RNA gene transcription by RNA polymerase III (Brown et al. 1985; Miller et al. 1985). Although several different classes of zinc finger proteins have been described, a major class of these proteins has the conserved C_2-H_2 (cysteine-histidine) motif typified by TFIIIA. The basic structural unit of this class consists of tandemly repeated units of 28 amino acids with the consensus sequence $(Y/F)XCX_2C\text{-}X_3FX_5LX_2HX_3$-HTGEKP (X is any amino acid), where the two pairs of cysteine and histidine residues coordinate a zinc ion (Hanas et al. 1983; Miller et al. 1985; Diakun et al. 1986).

Kruppel is a member of the gap class of *Drosophila* segmentation genes that codes for a protein with tandemly repeated zinc finger domains that comply with the TFIIIA consensus sequence (Preiss et al. 1985; Rosenberg et al. 1986; Tautz et al. 1987). In addition, each repeated finger unit is separated by a conserved stretch of seven or eight amino acids (HTGEKPY[Y/F]X), termed the "H/C link" (Schuh et al. 1986), which extends from the last histidine of one finger unit to the first histidine of the next finger unit. The *Kruppel* gene protein plays an important role in *Drosophila* development and segmentation, as mutations in this gene lead to the *Drosophila* embryo lacking all thoracic and five abdominal segments (Preiss et al. 1985). Furthermore, it has been demonstrated that the *Kruppel* gene product acts as a transcriptional regulator in cultured *Drosophila* cells and can functionally interact with other zinc finger and homeobox gene proteins (Zuo et al. 1991) in effecting precise patterns of gene expression in the developing embryo (Ingham et al. 1986; Ingham and Martinez-Aria 1986; Harding and Levine 1988). On the basis of the hypothesis that conserved structural motifs imply conserved function, various studies have defined mammalian zinc finger genes that potentially play an important role in development and show structural homology to the *Kruppel* consensus sequence (Chowdhury et al. 1987; Chavrier et al. 1988; Bellefroid et al. 1989; Cuncliffe et al. 1990; Huebner et al. 1991). Recent findings have also shown that zinc finger genes may be directly involved in the etiology of human malignancies, such as the t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia (McGuire et al. 1989) and the 11p13 deletion in Wilms tumor (Call et al. 1990; Gessler et al. 1990).

In order to isolate and characterize novel genes containing zinc finger domains that might play an important role in mammalian development, we have isolated, characterized, and mapped new zinc finger genes by virtue of sequence homology to other zinc finger genes, in order to analyze their possible function as transcription regulators. In the present study, we isolated two zinc finger–encoding cDNAs from a cDNA library constructed from a human hepatoblastoma cell line (HEP-G2) by using probes of the *Kruppel*-like zinc finger gene HZF-3 (Saleh et al., in press). HZF-3 represents a human genomic clone mapping to 11q23 and encoding a zinc finger protein. This work defines a novel zinc finger–encoding gene which gives rise to multiple transcription products and is located on human chromosome 1q44. An alternate splicing event generates a smaller gene product, in which there is a fusion between two of the zinc finger domains of the larger cDNA, and potentially generates a protein with an altered DNA-binding specificity. This gene is expressed in multiple human tissues, and alternate transcripts are observed in various transformed human cell lines.

Material and Methods

PCR Amplification of Zinc Finger–Encoding Domains

PCR primers were synthesized (PCR-Mate; Applied Biosystems, Foster City, CA) corresponding to the genomic sequence of an open reading frame contained in the human gene, HZF-3 (Saleh et al., in press), which has tandemly repeated zinc finger motifs (sense primer 1, 5′-GAT-TAC-TTC-ATA-TGG-TTT-CTC-TCC-AGT-ATG-3′; antisense primer 2, 5′-AAC-TCT-ATG-AAT-GTA-ATG-AAC-GTT-CTA-AAG-3′). The primers were used to amplify a 279-bp fragment from 100 ng of human genomic DNA. The amplification reaction was carried out in 30-μl reactions with 1 unit *Thermus aquaticus* polymerase with appropriate buffer (Stratagene, La Jolla, CA) and with 2.5 mM of each deoxynucleoside triphosphate. The reaction was carried out for 35 cycles of 1 min at 94°C, 1 min at 65°C, and 1 min at 72°C. Amplification of the PCR product was verified by visualization on 2% agarose gels. The PCR product was purified from the gel and was labeled to a specific activity of 5×10^8 cpm/μg by the random primer reaction (Feinberg and Vogelstein 1983) using [α-^{32}P]dCTP.

cDNA Library Screening

An adult human hepatoblastoma (i.e., HEP-G2) cDNA library prepared in lambda ZAP II vector (Stratagene, La Jolla, CA) was screened using the 279-bp PCR fragment corresponding to HZF-3. Approximately 6×10^5 plaques were screened, and hybridization was performed in 1 mM EDTA, 0.5 M $NaHPO_4$ pH 7.2, 7% SDS (Church and Gilbert 1984) for 16 h at 65°C. The filters were washed in 0.5 × SSC, 0.1% SDS at 65°C. The positive plaques were purified by replating and hybridization. The in vivo excision and circularization of the cloned insert in the lambda vector to form a phagemid (Bluescript vector) retaining the cloned insert was performed exactly as recommended by the manufacturer.

DNA Sequencing Determination

cDNA clones were sequenced directly off the Bluescript vector (phagemid; Stratagene, La Jolla, CA) by

using the T3 and T7 sequencing primers (Stratagene, La Jolla, CA). DNA sequencing was performed using double-stranded DNA templates with the Sequenase version 2.0 Sequenase enzyme (U.S. Biochemicals, Cleveland).

Southern Blot Analysis

Human genomic DNA was isolated from normal peripheral blood cells and was digested with a panel of restriction enzymes including *Eco*RI. Genomic DNA from a variety of species (Old World monkey, chimpanzee, chicken, rabbit, pig, frog, trout, and mouse [Oncor, Gaithersburg, MD]) were also digested with the same enzyme, and all the genomic DNAs were electrophoresed on a 1% agarose gel. The DNA was transferred to a GeneScreen membrane (DuPont, Wilmington, DE), and hybridization with the HZF-16 cDNA clones was performed in 1 mM EDTA, 0.5 M $NaHPO_4$ pH 7.2, 7% SDS at 65°C (Church and Gilbert 1984) for 16 h. The filters were washed in three changes of 2 × SSC, 0.1% SDS at 50°C and in 0.1 × SSC, 0.1% SDS at 65°C, where required, for higher stringency. DNA was radiolabeled by random hexamer-primed synthesis using [α-^{32}P]dCTP (Feinberg and Vogelstein 1983).

Northern Blot Analysis

Total RNA was isolated from 10^6–10^8 cells of each cell line. The cell pellet was resuspended in solution D (4 M guanidinium thiocyanate, 25 mM sodium citrate pH 7.0, 0.5% sarcosyl, 0.1 M β-mercaptoethanol), and then phenol and chloroform were added. The sample was then mixed vigorously, placed on ice for 15 min, and centrifuged at 14,000 *g* for 5 min. The aqueous phase was precipitated by the addition of isopropanol. The RNA was pelleted by centrifugation at 14,000 *g*, redissolved in solution D, then reprecipitated by isopropanol, and dissolved in 0.5% SDS. Routinely, 20 μg of total RNA was electrophoresed on a 1% agarose, 3% formaldehyde RNA gel and was transferred to GeneScreen membrane. The filters were hybridized with the HZF-16 cDNA inserts, as for Southern analysis described above, and were washed in 0.1 × SSC, 0.1% SDS at 55°C. Northern blots containing 2 μg of polyA$^+$ mRNA of normal human tissue were obtained from Clontech (Palo Alto, CA) and hybridized as described above. A 1.6-kb γ-actin fragment was also used to probe the filters to visually quantify the RNA concentration in each lane.

Human Cell Lines

The human tumor cell lines used for RNA analysis were A16 (bladder carcinoma), CCRF-CEM (T-cell), and SKNSH (neuroblastoma) and were obtained from K. D. Pischel (Salk Institute), and the HEP-G2 cell line (liver hepatoblastoma) was a gift from C. Biazak (Scripps Institute). The human fibroblast line CRL-1634 with a normal karyotype was obtained from Human Genetic Mutant Cell Repository (Camden, NJ). Fibroblast cell division was arrested by maintaining the cells in serum-free medium for 48 h. FCS (10%; Tissue Culture Biologicals, Tulare, CA) was then added to the cells to allow the commencement of cell division.

Chromosomal In Situ Suppression Hybridization (CISSH)

Slide preparation and CISSH were performed according to protocol described elsewhere (Selleri et al. 1991). Metaphase chromosomes were prepared by mitotic blockage of actively growing normal human fibroblast cells (CRL1634; American Type Culture Collection) and were spread on microscope slides (Verma and Babu 1989). The slides were simultaneously hybridized with human HZF-16.1 (1.6 kb) and HZF-16.2 (2.0 kb) cDNAs labeled with biotin-11-dUTP (Enzo, New York) and biotin-11-dCTP (Enzo, New York) by using a random primer-extension reaction (Selleri et al. 1991). Hybridization was visualized by treating the slides with fluoresceinated avidin and biotinylated goat anti-avidin (Vector Laboratories), both at 5 μg/ml (Lichter et al. 1988, 1990). Images were produced using a laser scanning confocal microscope (BioRad MRC 600). Narrow band-pass filters were used to obtain separate images for fluorescein isothiocyante (550 nm) and propidium iodide (610 nm), which were then superimposed electronically.

Results

Isolation of Zinc Finger–Encoding cDNAs from an HEP-G2 Library

It has been previously shown that the genomic sequence in the clone HZF-3 (Saleh et al., in press) contains an open reading frame that codes for previously undescribed zinc finger domains. The two tandemly repeated zinc finger domains comply with the consensus sequence of the *Drosophila Kruppel* gene and also have the characteristically conserved H/C link sequence that is present between subsequent finger mo-

tifs. Northern analysis has shown that HZF-3 is expressed as multiple transcripts in several adult human tissues, heart, placenta, brain, lung, liver, skeletal muscle, and kidney. PCR primers were synthesized according to the unique sequence of the open reading frame of HZF-3, so that the amplification product would span the sequence coding for the two zinc finger domains. These primers were used to amplify human genomic DNA whereby the product could serve as a probe specific for the HZF-3 zinc finger motif.

The 279-bp PCR-amplification product was used to screen 6×10^5 plaques of an adult human HEP-G2 cDNA library, under conditions of stringency that would allow the detection of highly homologous zinc finger–encoding sequences. Four cDNA clones were plaque purified and subjected to restriction-enzyme digestion for comparison of the inserts. This analysis revealed that three of the four cDNAs had a 2-kb insert and had an identical restriction pattern, indicating that three copies of the same cDNA (i.e., HZF-16.2) were isolated. The fourth cDNA had a 1.6-kb insert (i.e., HZF-16.1), yet it contained many of the same restriction sites as did the 2-kb cDNA. Therefore, sequence analysis was performed on both HZF-16.2 and HZF-16.1 cDNA clones.

The HZF-16.2 cDNA Contains Nine Zinc Finger Motifs

Sequence analysis of the 2,080-bp HZF-16.2 cDNA insert revealed that there is an open reading frame that extends for 888 nucleotides (fig. 1). There is an in-frame methionine residue at nucleotide 511, indicating a potential initiation site as well as an in-frame termination codon at nucleotide position 1399. There are also a number of stop codons upstream of this methionine residue in this reading frame. On the basis of the supposition that this methionine is the initiation site, the open reading frame would then extend for 296 amino acids. The 3′-untranslated sequence consists of 683 nucleotides ending in a polyA tail (14 nucleotides) and contains several AATAAA sites representing possible polyadenylation signals (fig. 1). The cDNA sequence is highly homologous but not identical to the genomic zinc finger sequence of HZF-3, at the nucleotide and amino acid levels.

The deduced amino acid translation of HZF-16.2 shows that there are nine tandemly repeated C_2-H_2 zinc finger motifs encoded in the open reading frame that conform with the basic consensus sequence of the *Drosophila Kruppel* finger motif (Rosenberg et al. 1986). The H/C link consensus sequence is also highly conserved in this cDNA (Schuh et al. 1986), with the exception of the amino acid sequence between finger domains 1–2, 5–6, and 6–7 (fig. 1 and fig. 2, *top*). In the H/C link sequence of the third finger domain, the proline residue is replaced by arginine. In addition to the conserved cysteine and histidine residues, the strictly conserved hydrophobic phenylalanine and leucine residues are present in all the finger domains of HZF-16.2, except in domain 2, where leucine is substituted by phenylalanine (fig. 2, *top*). Of the nine zinc finger domains encoded in HZF-16.2, the first domain has the highest level of divergence from the *Kruppel* consensus sequence (fig. 1 and fig. 2, *top*) in that there are an additional eight amino acid residues between the cysteine and histidine residues in the finger loop region (fig. 1 and fig. 2, *top*). There are also five residues, instead of three, between the two histidine residues; phenylalanine is not conserved; and leucine is located six residues, instead of three, prior to histidine (fig. 1 and fig. 2, *top*). It is unclear whether these variations in the motif would affect the conformation of the zinc finger domain and/or the DNA-binding capability of the domain. Thus, it is possible that the first domain may not be functional in DNA recognition and contact in this zinc finger protein.

HZF-16.2 cDNA Is Related to Other Zinc Finger Genes

The nucleotide and deduced amino acid sequences of the HZF-16.2 cDNA were compared with sequences of other zinc finger genes in GenBank and Pir + SwissProtein data bases (Altschul et al. 1990). These comparisons revealed that the sequence of HZF-16.2 cDNA is not identical to any previously identified zinc finger genes. Comparison of the 5′-untranslated sequence failed to identify any genes of significant homology at either the nucleotide or the amino acid levels. On the other hand, the 3′-untranslated sequence revealed 60%–65% (over a region 120–160 nucleotides long) identity with the human hox13 and Zfp36 mRNAs and with the mouse mfg1 and mfg2 mRNAs. As would be expected, the open reading frame containing the zinc finger domains of HZF-16.2 showed high levels of homology with other zinc finger genes. At the amino acid level, there is 50%–60% identity in a region of 200 residues with the human clone ZNF 8 and human finger-protein clone 647 and with the mouse mkr1 and mkr3 proteins.

```
   1  TTTTTTTTTTTTTTTTAACAATCGGATCTTTCAGGAACTAATAGAGCGAGAAGTCACTCA   60
  61  TTACCACAACAGTGCCACTTATGTGGAATCTGCTCCCATGACCCAAGCACCTCACACCAG  120
 121  GTCATACCTCCAACATGAGGATCAAATTTCAGCATGAAATTTGAAGGGAGAAAATACCCA  180
 181  AACTGTATTCAATACTAAGAAACTGCATATGAGAATATTACTGTATTGTTAATAGCTATA  240
 241  GGGGAGGGAGCCATGTTGTAGACTAATCAATCCATTTATGTTCAATTTGTTTATGTTAGA  300
 301  AAACCTGCACTTTCTCTGATATTGGTAGCAGTGTAAGTTCAGACTTAGTAATAAAAGAAA  360
 361  ATAACTAATAAACCATTAATGATGGGGTTGTCATTTTTTGCAGAAGTCATATGATAGACA  420
 421  TACTGTGTAAAATTAAATAAGTCAGTGTAGAAAAACCTCCAGATGCCAAATTTTAATCTG  480
 481  AACAAAAAATTCCTGCTAGAGTAAAACCACATGAATGCATTGTGTGTGAAAAATTCTTCA  540
   1                              M  N  A  L  C  V  K  N  S  S       10

 541  TACGTTCATTCATCCCTTCATAGGCACATCATATCTCATTCTGGAAACAACCCATATGGG  600
  11  Y  V  H  S  S  L  H  R  H  I  I  S  H  S  G  N  N  P  Y  G    30

 601  TGTGAGGAATGCGGAAAGAAGCCATGTACATGTAAACAATGTCAGAAAACTTCCCTTTCT  660
  31  C  E  E  C  G  K  K  P  C  T  C  K  Q  C  Q  K  T  S  L  S    50
       1
 661  GTCACAAGGGTTCACAGAGACACAGTAATGCACACTGGAAATGGACATTATGGTTGTACA  720
  51  V  T  R  V  H  R  D  T  V  M  H  T  G  N  G  H  Y  G  C  T    70
                                                          2
 721  ATATGTGAGAAAGTTTTTAATATTCCCAGTTCATTTCAGATACATCAGAGAAATCACACT  780
  71  I  C  E  K  V  F  N  I  P  S  S  F  Q  I  H  Q  R  N  H  T    90

 781  GGAGAGAAACCCTATGAATGTATGGAATGTGGGAAAGCCTTAGGTTTTTCCCGTTCTCTT  840
  91  G  E  K  P  Y  E  C  M  E  C  G  K  A  L  G  F  S  R  S  L   110
                          3
 841  AATAGACATAAAAGGATTCACACTGGAGAAAAACGCTATGAATGTAAGCAATGTGGGAAA  900
 111  N  R  H  K  R  I  H  T  G  E  K  R  Y  E  C  K  Q  C  G  K   130
                ▼                                   4
 901  GCCTTCAGTCGTTCCAGTCACCTTCGTGACCATGAAAGAACTCATACTGGAGAGAAACCC  960
 131  A  F  S  R  S  S  H  L  R  D  H  E  R  T  H  T  G  E  K  P   150

 961  TATGAATGTAAGCACTGTGGGAAAGCCTTCCGTTACTCCAATTGCCTTCATTACCATGAA 1020
 151  Y  E  C  K  H  C  G  K  A  F  R  Y  S  N  C  L  H  Y  H  E   170
             5
1021  AGAACTCACACTGGAGAGAAACCTTATGTGTGCATGGAATGTGGCAAAGCTTTCAGTTGT 1080
 171  R  T  H  T  G  E  K  P  Y  V  C  M  E  C  G  K  A  F  S  C   190
                                   6
1081  CTCAGTTCCTTGCAAGGACATATAAAGGCTCATGCTGGTGAAGAACCCTATCCATGTAAG 1140
 191  L  S  S  L  Q  G  H  I  K  A  H  A  G  E  E  P  Y  P  C  K   210
                                                             7
1141  CAATGTGGGAAAGCCTTCAGATACGCCAGTTCCCTTCAGAAACACGAGAAAACTCATATT 1200
 211  Q  C  G  K  A  F  R  Y  A  S  S  L  Q  K  H  E  K  T  H  I   230

1201  GCACAGAAACCCTATGTATGTAACAATTGTGGTAAAGGCTTCAGATGTTCCAGTTCCCTT 1260
 231  A  Q  K  P  Y  V  C  N  N  C  G  K  G  F  R  C  S  S  S  L   250
                         8
1261  CGTGACCATGAAAGGACTCATACTGGAGAGAAACCCTATGAATGTCAGAAATGTGGCAAA 1320
 251  R  D  H  E  R  T  H  T  G  E  K  P  Y  E  C  Q  K  C  G  K   270
                ▼                                   9
1321  GCCTTTAGTCGTGCTAGTACCCTTTGGAAGCATAAAAAAAACTCATACTGGAGAAAAGCCC 1380
 271  A  F  S  R  A  S  T  L  W  K  H  K  K  T  H  T  G  E  K  P   290

1381  TATAAATGTAAAAAAATGTAAAGGCTTTAATCACTACAGTTTTTGTCAAAAACATGAACA 1440
 291  Y  K  C  K  K  M  *                                          296

1441  GTCACATACTTGAGAGAAACTGTGAATGTAAGGTGTAGGAAAGTACTTAATTTTCCCAGA  1500
1501  TTTCCTCAAATACATGAAACGAATCAAACTGGAGATAAACCCTATGACTATAAGCAATAA  1560
1561  GGTAAAGCATTCAATTTTTCCATTTCTTTTTGAAAACTTGAAAGGACTCACTGAAGAAAA  1620
1621  TCCATATGAATGTTTAAAATGTGGTAAGGCCTGCAGTTGTTCCAGTGGTATTTGATGGTA  1680
1681  TAACATAACTCATTCTAGAGAAAAACTTTATGAAGGTATCGAATGTGAGAATGCCTTCAT  1740
1741  TTATCCTATAACTCACTCAGAGACACATGGTAACACATACTCCAGATTGACCTTATAAAT  1800
1801  AAAAGAATGCCCACCAGATTGAAATCCTAGAAATACAGAAAATTCTGAATTTTAACAATT  1860
1861  ACTTTAAAGGTCATGTGAAAACTCCCACTGGAAATAAATCCTGTAAATGTAAATATTATG  1920
1921  GAAAACCTTATGGAAAATAATTATAGTACAATTCTCAAGAAAATTCACTATGTGAATGAG  1980
1981  ATGTTTTAGGAGCTATAAATAAATTGAATATATTGGTTGCTCATTTTTGAAGAGAGTCTC  2040
2041  TAGAATATAGATTTTCACTTCTTTTGAAAAAAAAAAAAAA  2080
```

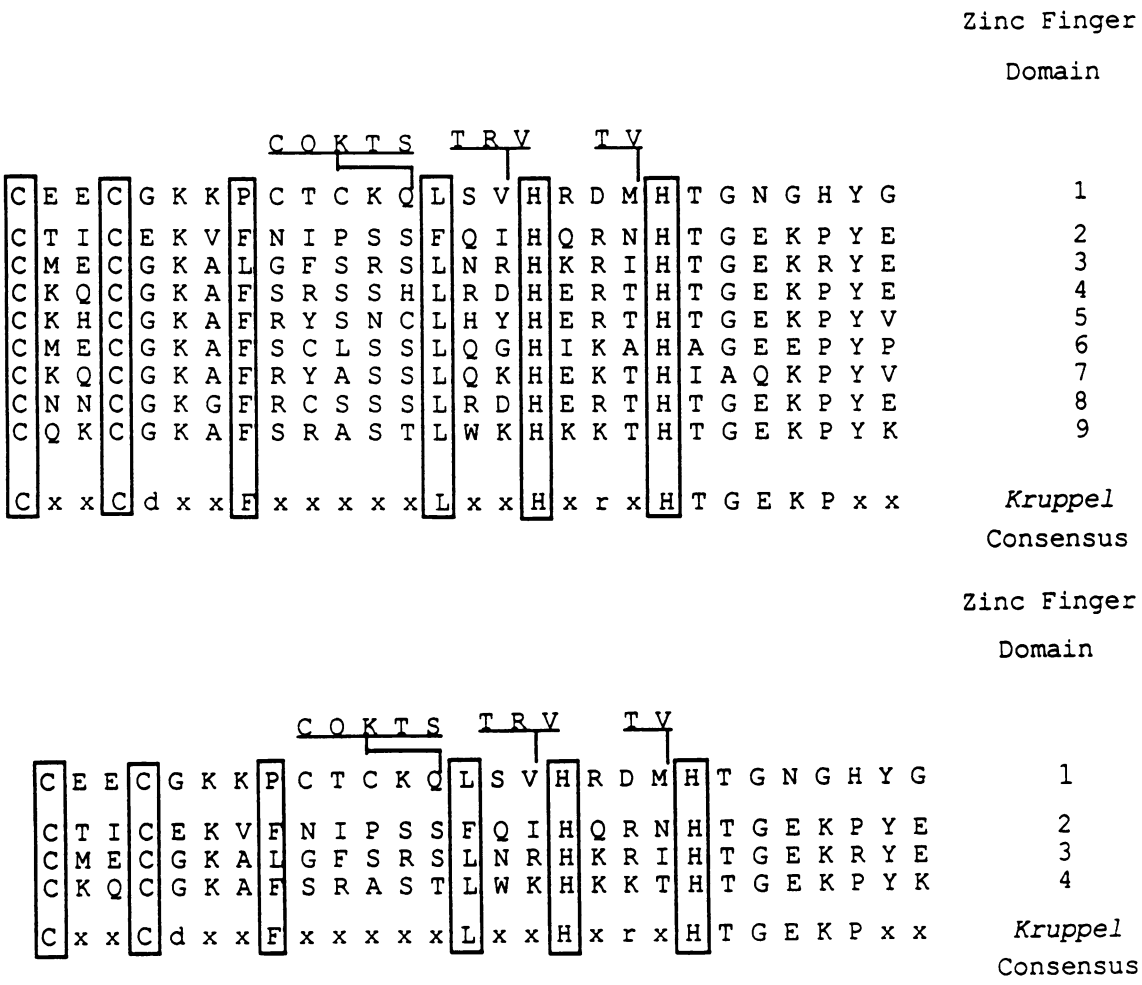

Figure 2 Comparison of the deduced amino acid translation of the tandemly repeated zinc finger domains encoded by HZF-16.2 (*top*) and HZF-16.1 (*bottom*) cDNAs with the *Kruppel* consensus sequence. The sequence is presented in one-letter code and aligned to show the repeated units. Amino acids strictly conserved in the zinc finger motif are boxed. In the *Kruppel* consensus sequence, uppercase letters denote strictly conserved residues, lowercase letters indicate residues conserved over 50%, and (x) denotes residues conserved less than 50%. Underlined amino acids represent those present in domain 1 but absent in domains 2–9.

HZF-16.1 cDNA Represents an Alternatively Spliced Product of a Single HZF-16 Gene

The HZF-16.1 cDNA has an insert size of 1,633 nucleotides, as determined by sequence analysis (fig. 1). The nucleotide sequence of HZF-16.1 is identical to that of HZF-16.2 for the first 921 nucleotides, which therefore includes the 5′-untranslated sequence, the proposed initiation site, and the first three complete zinc finger motifs. The sequence identity is interrupted halfway into the fourth zinc finger domain of HZF-16.2 at nucleotide 913. However, it appears that the sequence of HZF-16.1 resumes in perfect identity with HZF-16.2 at nucleotide 1333, which is halfway into the ninth finger motif of HZF-16.2. The sequence

Figure 1 Nucleotide sequence analysis of the HZF-16.1 cDNAs. The amino acid translation in single-letter code is shown for the open reading frame beginning at the first methionine residue indicating a potential initiation site. There is an in-frame termination codon (denoted by an asterisk), and the sequence is shown to the polyA tail. The zinc finger motifs are underlined, and the respective complete motifs are numbered 1–9. The possible polyadenylation signals (AATAAA) in the 3′-untranslated region are underlined. Inverted black triangles denote where the sequence of HZF-16.1 is interrupted and resumes, in comparison with HZF-16.2.

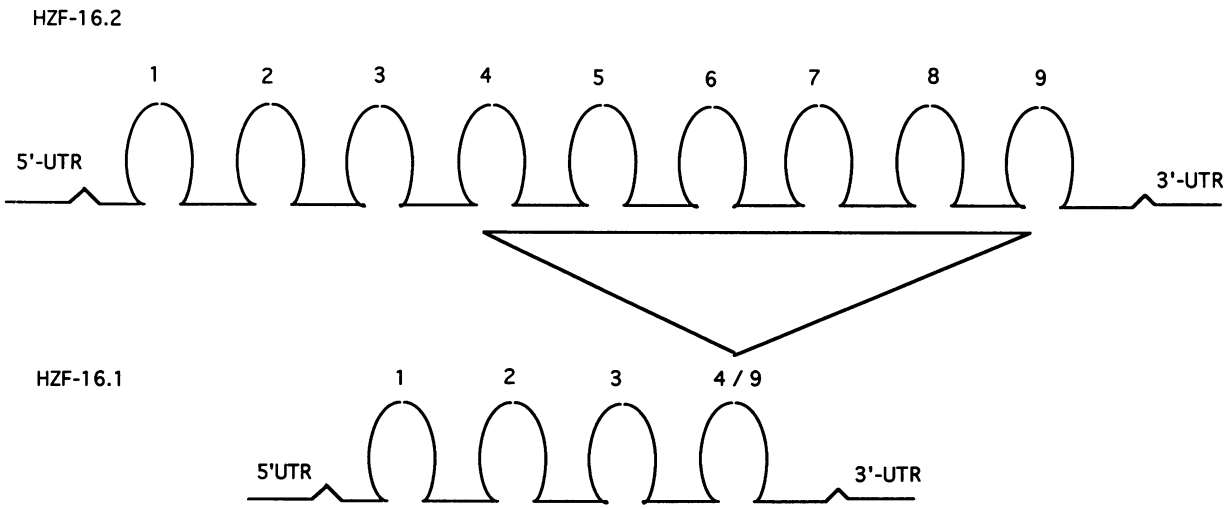

Figure 3 Schematic representation of the zinc finger domains of both HZF-16.2 and HZF-16.1 (not drawn to scale). The domains are numbered 1–9, and the domains of HZF-16.2 that are not present in HZF-16.1 are indicated by a triangle. All remaining sequences of the two cDNAs are identical. The last domain of HZF-16.1 is shown as a combination of sequences corresponding to domains 4 and 9 of HZF-16.2.

identity continues through to the in-frame stop codon and the entire 3′-untranslated sequence and polyA tail (fig. 1). The deduced amino acid sequence of HZF-16.1 shows that the open reading frame codes for four complete zinc finger motifs that include the H/C link sequences. The first three domains of HZF-16.1 are identical to the first three domains of HZF-16.2, while the fourth domain is identical to the first half of the fourth domain and to the last half of the ninth domain of HZF-16.2 (fig. 2). Therefore, there appears to be an internal splicing in HZF-16.1 of 420 nucleotides from the sequence of HZF-16.2 (fig. 1), which does not interrupt the reading frame, and the sequence resumes in such a way that a complete zinc finger motif is still encoded (fig. 1 and fig. 2, *bottom*). A schematic representation of the zinc finger domains of the two cDNA clones is shown in figure 3.

In order to determine whether HZF-16.1 and HZF-16.2 cDNAs are transcribed from the same gene, the insert of HZF-16.2 was used as a probe to hybridize total human genomic DNA, under high-stringency conditions. The human genomic DNA was digested with a panel of enzymes including *Eco*RI, *Bam*HI, *Bgl*II, *Hin*dIII, and *Pst*I. This probe was found to hybridize to a single fragment of human DNA rather than to multiple fragments (fig. 4; only the *Eco*RI digest is shown). Digests showing more than one band hybridizing corresponded to the existence of internal restriction sites in the cDNA. Hybridization of total human DNA with the insert of the HZF-16.1 cDNA resulted in the detection of the identical single fragment (data not shown), indicating that the two cDNAs are recognizing the same genomic sequence. This experiment was also repeated with a *Pst*I/*Eco*RI fragment that only contains a 3′-untranslated sequence that is identical in both cDNAs. Even under low-stringency conditions (2 × SSC, 0.1% SDS at 50°C), only a single band was detected on human genomic DNA, which was the same size as that detected using the entire cDNA insert (data not shown). These findings suggest that it is most likely that the two cDNAs are transcribed from the same gene.

The HZF-16 Gene Is Conserved across Species

In order to examine whether the sequences coding for the HZF-16 gene are conserved between species, the cDNA insert of HZF-16.2 cDNA was used as a probe to hybridize zoo blots. Low-stringency conditions of washing (2 × SSC, 0.1% SDS at 50°C) revealed multiple bands in human, Old World monkey, chimpanzee, pig, rabbit, and mouse, and only single bands in chicken and frog (fig. 4). Increasing the stringency of washing (0.1 × SSC, 0.1% SDS at 55°C) reduced many of the multiple bands detected in human and Old World monkey so that essentially one band remained; the single band in frog was still detected, whereas the bands in chimpanzee, pig, mouse, and chicken were no longer detected (data not shown). A

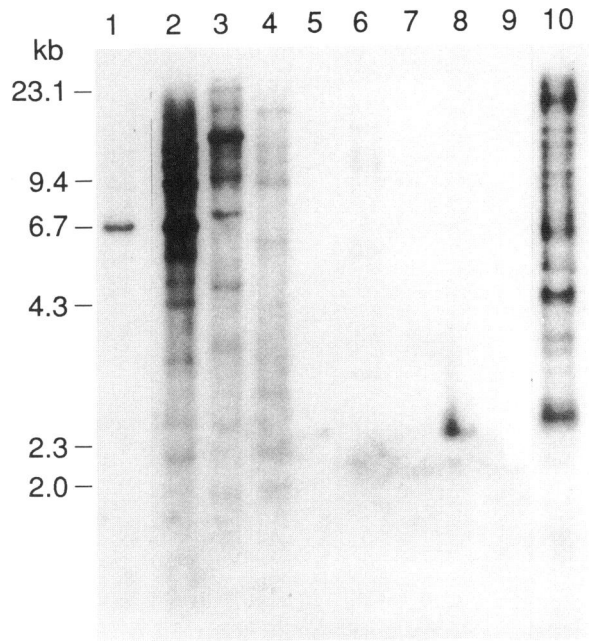

Figure 4 Southern analysis of genomic DNA from a variety of species that was digested with *Eco*RI and hybridized with a 2-kb insert of the human HZF-16.2 cDNA. Human genomic DNA (lane 1) was analyzed under high-stringency conditions (0.1 × SSC, 0.1% SDS at 65°C). The remaining DNA was analyzed under less stringent conditions (2 × SSC, 0.1% SDS at 50°C). The genomic DNAs analyzed are human (lane 2), Old World monkey (lane 3), chimpanzee (lane 4), chicken (lane 5), rabbit (lane 6), pig (lane 7), frog (lane 8), trout (lane 9), and mouse (lane 10).

further increase in stringency (0.1 × SSC, 0.1% SDS at 65°C) resulted in only a single band in human DNA being detected (fig. 4). Identical results were obtained when hybridization of the zoo blots was performed with the insert of the HZF-16.1 cDNA as a probe (data not shown). These results suggest that the gene coding for the human zinc finger gene HZF-16 is conserved between distant species.

The HZF-16 Gene Is Expressed as Alternate Gene Products in Transformed Human Cell Lines and Various Normal Human Tissues

In order to determine whether HZF-16.1 and HZF-16.2 cDNAs are indeed coexpressed in the adult HEP-G2 liver hepatoblastoma cell line (from which the cDNA library was constructed), northern analysis was performed on this cell line by using the 2-kb insert of HZF-16.2 as the probe. Four apparent transcripts, seen as bands on northern blots and appearing as two sets of doublets (fig. 5, *left*) with transcript sizes of 2.5 kb, 2.2 kb, 1.5 kb, and 1.3 kb, were detected. An additional band at approximately 3.8 kb, which coincides with the 28s rRNA band, likely corresponds to an additional, longer transcript. In these cell lines, the distribution and abundance of this transcript is difficult to determine in light of the coincidence with the ribosomal RNA band. Northern analysis was also performed on the human cell lines A16 (bladder carci-

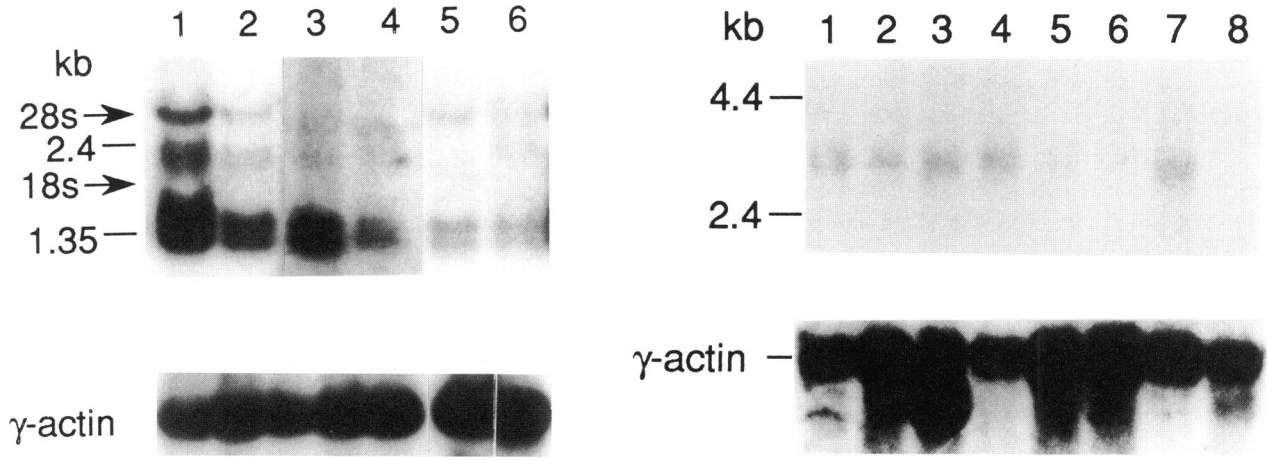

Figure 5 Northern analysis of human cell lines and adult human tissues. *Left,* Total RNA analyzed from the human cell lines liver hepatoblastoma (lane 1), A16 bladder carcinoma (lane 2), CEM (lane 3), SKNSH neuroblastoma (lane 4), and CRL 1634 fibroblast in G_0 (lane 5) and G_1 (lane 6). Hybridization was performed with the 2-kb insert of the HZF-16.2 cDNA. There is cross hybridization of the HZF-16.2 probe with 28s ribosomal RNA. *Right,* Normal human tissues analyzed. Each lane contains 2 μg of polyA$^+$ RNA from the normal human tissues: heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7), and pancreas (lane 8). RNA size markers (in kb) are shown. A γ-actin fragment (detecting a 1.6-kb transcript) was also hybridized to both membranes to demonstrate the RNA concentration of each tissue loaded on the gel. The γ-actin probe recognizes two species of cytoplasmic actin in normal human placenta.

noma), CEM (T-cell), and SKNSH (neuroblastoma), and an identical set of four transcripts was also detected (fig. 5, *left*). The normal human fibroblast line CRL1634 was also analyzed at an actively dividing (G_1) and arrested (G_0) cell stage, and, again, identical transcripts were observed in both cases. In all cases, the shorter transcripts are expressed in a higher relative abundance compared with the longer transcripts, and, in fibroblasts, the longer transcripts are just visible.

Northern analysis was also performed on normal adult human tissues (heart, placenta, brain, lung, liver, skeletal muscle, kidney, and pancreas) where only a single 3.8-kb transcript was detected in all the tissues examined, with the exception of pancreas (fig. 5, *right*). This transcript is longer than the transcripts detected in the cell lines examined. In order to establish that the hybridization signal is specific rather than due to cross-hybridization of homologous sequences, northern analysis was repeated with a *Pst*I/*Eco*RI fragment that only contains 3′-untranslated sequence. This probe was previously shown to detect only a single human genomic DNA fragment, under low-stringency conditions (described above). Transcripts detected by using this probe were identical to those detected with the entire HZF-16.2 cDNA insert. These results argue against the possibility that the transcripts detected represent homologous rather than identical sequences. However, despite high-stringency hybridizations, it remains possible that transcripts of high homology rather than of perfect identity are detected.

The HZF-16 Gene Is Located at Chromosome Band 1q44

CISSH of normal human fibroblast metaphases was performed with the HZF-16.1 (1.6-kb) and HZF-16.2 cDNA (2.0-kb) inserts hybridized together to human normal metaphases in order to determine the chromosomal localization of this gene. These experiments revealed that HZF-16.2 maps to the telomeric tip of human chromosome 1q (fig. 6). Fluorescent signals on both sister chromatids of chromosome 1q were observed in 30%–40% of metaphases examined. Fluorescent signals on a single chromatid or on both sister chromatids of chromosome 1 were observed at qter in over 40% of metaphases analyzed (fig. 6). Chromosome 1 was identified on the basis of morphological criteria. The region of the centromeric chromatin that is characteristic of the long arm of chromosome 1 was evident by microscopic examination of the propidium iodide–stained chromosomes (fig. 6), thus allowing the localization of the fluorescent signals to be assigned to 1q. The HZF-16 gene could be further localized to 1q44, as the fluorescent signals were observed at the extreme tip of 1q in all metaphases that had hybridization signals. To confirm the localization, PCR analysis was carried out using a set of monochromosomal somatic cell hybrids (data not shown).

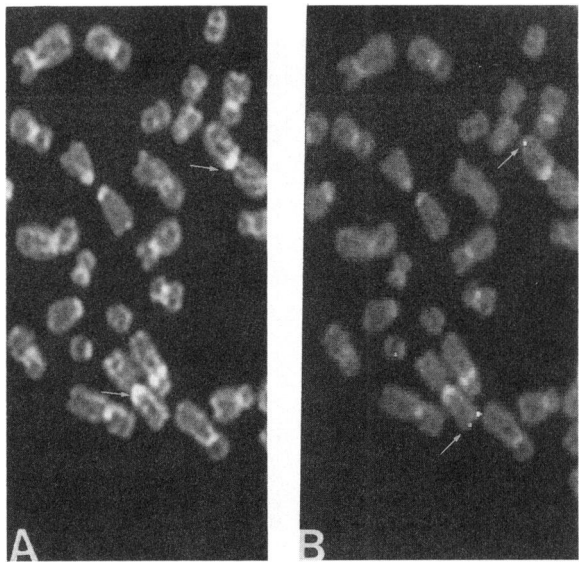

Figure 6 CISSH of HZF-16.1 and HZF-16.2 cDNAs on normal human fibroblast metaphases. Images were produced using a laser scanning confocal microscope (BioRad MRC 600), and black-and-white images were generated by a Mavigraph video printer (Sony). *A*, Image of the propidium iodide–stained chromosomes before merging with the fluorescein isothiocyanate image. Morphology and clear visualization of the centromeric chromatin (arrows) allow identification of chromosome 1q. *B*, Electronically reconstructed picture of propidium iodide– and fluorescent isothiocyanate–superimposed images. Fluorescent signals (indicated by arrows) are evident at the tip of chromosome 1q.

Discussion

DNA-binding proteins have been shown to play a fundamental role in the complex molecular interactions that lead to the precise regulation of gene expression. The structural motifs conserved through evolution in the different classes of DNA-binding proteins have allowed the isolation of a growing number of potential transcription factors from a variety of species. We have taken advantage of the conserved zinc finger motif of the C_2-H_2 class of DNA-binding proteins to isolate new zinc finger genes that have potential significance in development, on the basis of the

assumption that proteins that share the zinc finger–structural motifs will also have important functional roles as transcriptional regulators.

We screened a human liver hepatoblastoma (i.e., HEP-G2) cDNA library with a PCR product amplified from a human genomic fragment (i.e., HZF-3) previously shown to encode *Kruppel* zinc finger motifs (Saleh et al., in press). Two cDNAs were isolated (2 kb, HZF-16.2; and 1.6 kb, HZF-16.1), and their DNA sequence was determined. The sequence of HZF-16.2 revealed an open reading frame encoding nine zinc finger motifs that comply with the C_2-H_2 *Kruppel* consensus sequence. The second cDNA, HZF-16.1, was found to contain a sequence identical to the 5′-untranslated region and the first three zinc finger domains of HZF-16.2. The fourth domain of HZF-16.1 corresponds to the first half of the fourth domain and to the last half of the ninth domain of HZF-16.2, and thus it suggests an alternate mRNA splicing product resulting in a deletion of a portion of HZF-16.1 and a potentially new DNA-binding specificity. The 3′-untranslated regions are also identical between HZF-16.2 and HZF-16.1.

There are three possible explanations for the nature of the sequences observed between the two cDNAs: (i) the two cDNAs are transcribed from two different genes; (ii) the shorter cDNA is a cloning artifact; and (iii) the two cDNAs represent alternatively spliced products transcribed from a single gene. The first possibility is highly unlikely because of the identical sequences of both the 5′- and 3′-untranslated regions, as well as the observation that only a single human genomic fragment, rather than multiple fragments, is detected by the cDNAs. The second possibility, that HZF-16.1 is an artifact produced during library construction, is also excluded, as both the 2-kb and 1.6-kb transcripts are detected in the human cell lines analyzed. Thus, the results favor the third possibility—that the two cDNAs are alternatively spliced products transcribed from a single gene. The detection of another two transcripts in human cell lines, as shown by northern analysis, indicates that there may be additional spliced products of the same gene, besides the two cDNAs cloned. Furthermore, the single 3.8-kb transcript detected in the normal human tissues analyzed is longer than those observed in human cell lines. This finding suggests that HZF-16 may be a gene with a complex splicing mechanism that gives rise to all five different transcripts observed.

Alternative splicing of products transcribed from the one gene have been reported elsewhere for zinc finger genes, as with the human *Zfx* gene (Schneider-Gadick et al. 1989). In this case, alternative splicing and polyadenylation give rise to three cDNAs that have structurally distinct 5′- and 3′-untranslated regions, although two of the cDNAs have identical open reading frames (Schneider-Gadick et al. 1989). Thus, the HZF-16.1 and HZF-16.2 cDNAs described here represent an unusual and complex pattern of alternative splicing, in terms of both the structure of the cDNAs and the differential expression of the spliced products between normal human tissues and human cell lines. From the limited number of tissues and cell lines analyzed, it appears that the spliced products are expressed not only in transformed cell lines but also in cultured fibroblast cells, although in a significantly lower relative abundance. This may suggest that culture conditions may somehow induce alternative splicing of the HZF-16 gene transcript, whereas only the full-length transcript is transcribed in normal human tissues. The differential expression of the transcripts does not appear to be related to cell division, since northern analysis of RNA isolated from actively dividing (G_1) and arrested (G_0) cells resulted in the detection of the same transcripts, although at a significantly lower relative abundance. A larger panel of both human cell lines and normal human tissues, including tissues of different developmental stages, needs to be examined to gain a more complete understanding of the expression patterns of the spliced products. The other transcripts observed in the cell lines and tissues must now be isolated and analyzed, in order to determine the manner in which their splicing events occur. In addition, the isolation of the genomic sequence and determination of the organization of the HZF-16 gene will be crucial in understanding the mechanism of alternative splicing that occurs in this gene.

The HZF-16 gene was located within chromosomal band 1q44 by high-resolution CISSH. Abnormalities of chromosome 1q have been associated with breast cancer (Chen et al. 1989), gastric carcinoma (Sano et al. 1991) and other gastrointestinal cancers (Fey et al. 1989), testicular and germ-cell carcinomas (Murty et al. 1990), a severe dysmorphology syndrome (Mascarello et al. 1989), and idiopathic myelofibrosis (Donti et al. 1990). In human leukemia B-cell clones grown in nude mice, 1q trisomy has been associated with increased tumorigenicity, growth, and metastatic potential (Ghose et al. 1990). Van der Woude syndrome, an autosomal dominant syndrome characterized by cleft lip, cleft palate, hypodontia, and abnormalities in craniofacial development, has been linked to genetic

markers at 1q32-44 (Murray et al. 1990); and Usher syndrome type II, characterized by congenital hearing loss, retinitis pigmentosa, and other vestibular problems, has been located through linkage analysis to the distal long arm of chromosome 1 (Kimberling et al. 1990). Additional analysis of this gene on chromosome 1q will shed light on the functional relationship of the HZF-16 gene and related genes to human development and pathology.

Acknowledgments

We wish to thank G. G. Hermanson, M. Giovannini, and M. Djabali (Salk Institute, La Jolla, CA) for invaluable discussion, and we thank C. Biazak (Scripps Clinic and Research Foundation, La Jolla, CA) for the human liver hepatoblastoma cell line. This work was supported by NIH grants GM33868 and HG00202 (to G.A.E.) and by funds from the G. Harold and Leila Y. Mathers Foundation. L.S. is a recipient of an NIH Fogarty International Fellowship.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman J (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Bellefroid EJ, Lecocq PJ, Benhida A, Poncelet DA, Belayew A, Martial JA (1989) The human genome contains hundreds of genes coding for finger proteins of the *Kruppel* type. DNA 8:377–387

Brown RS, Sander C, Argos P (1985) The primary structure of the transcription factor TFIIIA has twelve consecutive repeats. FEBS Lett 186:271–274

Call KM, Glaser T, Ho CY, Buckler AJ, Pelletier J, Haber DA, Rose EA, et al (1990) Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilm's tumor locus. Cell 60:509–520

Chavrier P, Lemaire P, Revelant O, Bravo R, Charnay P (1988) Characterization of a mouse multi-gene family that encodes zinc finger structures. Mol Cell Biol 8:1319–1326

Chen LC, Dollbaum C, Smith HS (1989) Loss of heterozygosity on chromosome 1q in human breast cancer. Proc Natl Acad Sci USA 86:7204–7207

Chowdhury K, Deutch U, Gruss P (1987) A multi-gene family encoding several finger structures is present and differentially active in mammalian genomes. Cell 48:771–778

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81:1991–1995

Cuncliffe V, Koopman P, McLaren A, Townsdale J (1990) A mouse zinc finger gene which is transiently expressed during spermatogenesis. EMBO J 9:197–205

Diakun GP, Fairall L, Klugg A (1986) EXAFS study of the zinc binding sites in the protein transcription factor IIIA. Nature 324:689–691

Donti E, Tabilio A, Bocchini F, Falzetti F, Martelli MF, Grignani F, Donti GV (1990) Partial trisomy 1q in idiopathic myelofibrosis. Leuk Res 14:1035–1040

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragment to high specific activity. Anal Biochem 132:6–13

Fey MF, Hesketh C, Wainscoat JS, Gendler S, Thien SL (1989) Clonal allele loss in gastrointestinal cancers. Br J Cancer 59:750–754

Gehring WJ (1987) Homeoboxes in the study of development. Science 236:1245–1252

Gessler M, Poutzka A, Lavenee W, Neve RL, Orkin SH, Bruns GAP (1990) Homozygous deletion in Wilm's tumor of a zinc finger gene identified by chromosome jumping. Nature 343:774–778

Ghose T, Lee CL, Fernandez LA, Lee SH, Raman R, Colp P (1990) Role of 1q trisomy in tumorgenicity, growth and metastasis of human leukemic B-cell clones in nude mice. Cancer Res 50:3737–3742

Hanas JS, Hazuda DJ, Bogenhagen DF, Wa Y-H, Wa C-W (1983) *Xenopus* transcription factor A requires zinc for binding to the 5S RNA gene. J Biol Chem 258:14120–14125

Harding K, Levine M (1988) Gap genes define the limits of antennapedia and bithorax gene expression during early development in *Drosophila*. EMBO J 7:205–214

Huebner K, Druck T, Croce CM, Thiesen H-J (1991) Twenty-seven nonoverlapping zinc finger cDNAs from human T cells map to nine different chromosomes with apparent clustering. Am J Hum Genet 48:726–740

Ingham PW, Ish-Horowicz D, Howard K (1986) Correlative changes in homeotic and segmentation gene expression in Kruppel mutant embryos of *Drosophila*. EMBO J 5: 1659–1665

Ingham PW, Martinez-Arias A (1986) The correct activation of antennapedia and bithorax complex gene requires the fushi-tarazu gene. Nature 324:592–597

Kimberling WJ, Weston MD, Moller C, Davenport SL, Shugart YY, Priluck IA, Martini A, et al (1990) Localization of Usher syndrome type II to chromosome 1q. Genomics 7:245–249

Lichter P, Chang TCC, Call K, Hermanson GG, Evans GA, Housman D, Ward DC (1990) High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. Science 247:64–69

Lichter P, Cremer T, Tang CC-J, Watkins P, Manuelidis L, Ward DC (1988) Rapid detection of human chromosome 21 aberrations by *in situ* hybridization. Proc Natl Acad Sci USA 85:9664–9668

McGuire EA, Hochell RD, Pollock KM, Bartholdi MF, O'Brien SJ, Korsmeyer SJ (1989) The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts including Ttg-1, a gene encoding a potential zinc finger protein. Mol Cell Biol 9:2124–2132

Mascarello JT, Osborn C, Kadota RP (1989) A dysmorphic child with myelodysplasia characterized by a duplication

of 1q and multiple duplications of 3q. Cancer Genet Cytogenet 38:9–12

Miller J, Mclachlan AD, Klugg A (1985) Repetitive zinc-binding domains in the transcription factor TFIIIA from *Xenopus* oocytes. EMBO J 4:1609–1614

Murray JC, Nishimura DY, Buetow KH, Ardinger HH, Spence MA, Sparkes RS, Falk RE, et al (1990) Linkage of an autosomal dominant clefting syndrome (Van der Woude) to loci on chromosome 1q. Am J Hum Genet 46: 486–491

Murty VV, Dmitrovsky E, Bosl GJ, Chaganti RS (1990) Nonrandom chromosome abnormalities in testicular and ovarian germ cell tumor cell lines. Cancer Genet Cytogenet 50:67–73

Pabo CO, Sauer RT (1984) Protein-DNA recognition. Annu Rev Biochem 53:293–321

Preiss A, Rosenberg VB, Klenin A, Seifert E, Jackle H (1985) Molecular genetics of *Kruppel,* a gene required for segmentation of the *Drosophila* embryo. Nature 313:27–32

Rosenberg UB, Schroder C, Preiss A, Kienlin A, Cote S, Riede I, Jackle H (1986) Structural homology of the product of the *Drosophila Kruppel* gene with *Xenopus* transcription factor IIIA. Nature 319:336–339

Saleh M, Selleri L, Little PFR, Evans GA. Isolation and expression of linked zinc finger gene clusters on human chromosome 11q. Genomics (in press)

Sano T, Tsujino T, Yoshida K, Nakayama H, Haruma K, Ito H, Nakamura Y, et al (1991) Frequent loss of heterozygosity of chromsomes 1q, 5q and 17p in human gastric carcinomas. Cancer Res 51:2926–2931

Schneider-Gadick A, Berr-Romero P, Brown LG, Mardon G, Luoh SW, Page DC (1989) Putative transcription activated with alternate isoforms encoded by human ZFX gene. Nature 342:708–710

Schuh R, Aicher W, Gaul U, Cote S, Preiss A, Maler D, Seifert E, et al (1986) A conserved family of conserved nuclear proteins containing structural elements of the finger protein encoded by *Kruppel,* a *Drosophila* segmentation gene. Cell 47:1025–1032

Selleri L, Hermanson GG, Eubanks JH, Lewis KA, Evans GA (1991) Molecular localization of the t(11;22)(q24; q12) translocation of Ewing sarcoma by chromosomal *in situ* suppression hybridization. Proc Natl Acad Sci USA 88:887–891

Tautz D, Lehmann R, Schnurch H, Schuh R, Seifert E, Kienlin A, Jones K (1987) Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation gene. Nature 327:383–389

Verma RS, Babu A (1989) Tissue culture techniques and chromosome preparation. In: Human chromosomes: manual of basic techniques. Pergamon, New York, pp 4–44

Zuo P, Stanojevic D, Colgan J, Han L, Levine M, Manley JL (1991) Activation and repression of transcription by the gap proteins hunchback and *Kruppel* in cultured *Drosophila* cells. Genes Dev 5:254–264