J Mol Evol (1995) 41:784–794

JOURNAL OF MOLECULAR EVOLUTION

Human and Mouse Krüppel-Like (MOK2) Orthologue Genes Encode Two Different Zinc Finger Proteins

Michèle Ernoult-Lange,[1] Valérie Arranz,[1] Maryvonne Le Coniat,[2] Roland Berger,[2] Michel Kress[1]

[1] GMIFC CNRS-UPR 9044, Institut de Recherche sur le Cancer, 7 rue Guy Moquet, 94801 Villejuif Cedex, France
[2] INSERM U-301, Institut de Génétique Moléculaire, 27 rue Juliette Dodu, 75010 Paris, France

Received: 24 March 1995 / Accepted: 17 July 1995

Abstract. We have isolated the human homologue of *Mok2* gene encoding a Krüppel-like protein. The identification of three cDNAs and genomic clones reveals that the human protein shows substantial structural differences with the mouse MOK2 protein. The mouse MOK2 protein is composed of seven tandem zinc-finger motifs with five additional amino acids at the COOH-terminal. This structural feature is also present at the end of the human MOK2 protein. The seven zinc-finger motifs show 94% identity between the two proteins. In addition, the human protein contains three additional zinc-finger motifs in tandem with the others and a nonfinger acidic domain of 173 amino acids at the NH_2-terminal. The Southern analysis indicates that a single copy of these two genes is present in the genome. The human gene has been localized on chromosome 19 on band q13.2–q13.3. The comparison of human and mouse cDNA sequences reveals a strong identity in the sequences localized outside the seven highly conserved zinc-finger motifs. The divergence from their common ancestor results in the loss of a potential transcription activator domain in mouse MOK2 protein.

Key words: Zinc-finger protein — Human — Mouse — Chromosome location — Divergent evolution

Introduction

Eukaryotic nucleic-acid-binding proteins can be classified according to the conserved structural motifs they share, such as the C_2H_2-type zinc-finger repeat discovered in *Xenopus* transcription factor IIIA (TFIIIA) (Miller et al. 1985). The zinc-finger motif is unique in functioning both as a specific RNA- and a specific DNA-binding module (reviewed in El-Baradi and Pieler 1991). A large zinc-finger multigene family (ZFP), probably made up of several hundred different genes, has been characterized on the basis of structural similarity in the finger repeat and is represented in many different organisms (Bellefroid et al. 1989; Chowdhury et al. 1987; Köster et al. 1988; Nietfeld et al. 1989). A few mammalian ZFP proteins have been characterized in detail and shown to participate in the control of cell proliferation, cell differentiation, and development (Call et al. 1990; Chavrier et al. 1989; Crosby et al. 1991; Gessler et al. 1990; Kinzler et al. 1988; Rosenberg et al. 1986; Sukhatme et al. 1988). The vast majority of ZFP genes are classified as Krüppel-like on the basis that they share a conserved stretch of six amino acids (H/C-link) connecting consecutive finger motifs (Schuh et al. 1986).

We have previously identified the murine intronless *Mok2* gene that encodes a 22.8-kDa Krüppel-like protein (Ernoult-Lange et al. 1990). The murine MOK2 protein consists of seven tandem zinc-finger motifs highly similar to one another with only five additional amino acids at the COOH-terminal. At present, only one other human gene encoding a protein composed mainly of zinc-finger motifs has been isolated (Lechalony et al. 1994). The *Mok2* gene is preferentially expressed in cells transformed by chemical agents or by DNA tumor viruses such as SV40 or adenovirus and in two normal adult tissues, brain and testis (Ernoult-Lange et al. 1990). The promoter of mouse *Mok2* gene differs from the tissue-

Correspondence to: M. Ernoult-Lange

specific genes. It lacks TATA and CCAAT boxes and contains two short interspersed mouse genomic repeats (B2 sequences) which exert a negative cis-acting effect on *Mok2* promoter activity (Arranz et al. 1994).

We report here the characterization and the chromosomal localization of the human homologue of *Mok2* gene. The identification of three cDNAs and genomic clones reveals that the human protein is larger than the murine MOK2 protein. The seven tandem zinc-finger motifs similar to the murine protein with five additional amino acids are present at the COOH-terminal of the human protein. In addition, the latter contains three additional zinc-finger motifs and a nonfinger coding region at the NH_2-terminal. Our results suggest that these two genes present as single copy in the genome are orthologue. The comparison of human and mouse cDNA sequences reveals a strong identity in the sequences outside of the seven highly conserved zinc-finger motifs. The divergence from their common ancestor results in the loss of potential transcription activator domain in mouse MOK2 protein.

Materials and Methods

Cell Lines and Recombinant Plasmid Construction. A-431 (epidermoid carcinoma), Raji (Burkitt lymphoma), JEG-3 (choriocarcinoma), 293 (embryonal kidney transformed by adenovirus type 5), and HeLa (cervical carcinoma) human cell lines were obtained from the ATCC collection. The SW613-S C13 cell line was derived from a colon carcinoma (Modjatahedi et al. 1992). The A-431, Raji, JEG-3, 293, and SW613-S C13 cells were grown in Dulbecco's modified Eagle's minimal essential medium (DMEM) supplemented with 10% fetal calf serum. HeLa cells were grown in DMEM with 10% newborn calf serum.

Genomic and cDNA Libraries Screening. An EMBL3 genomic library prepared from human chronic myeloid leukemia was screened with a 600-bp *Ddel* fragment containing a zinc-finger region of mouse *Mok2* cDNA. Six filters (Hybond-N, Amersham; 20 cm by 20 cm) with approximately 10^5 phages were hybridized at 65°C for 16 h in 3× SSC, 1× Denhardt's solution, and 0.1% SDS with a probe concentration of 1×10^6 cpm/ml. Filters were washed four times at 65°C for 15 min in 2× SSC and 0.1% SDS and one time at 65°C for 15 min in 0.1× SSC and 0.1% SDS. Positive plaques were analyzed for restriction enzyme mapping. All clones were identical and one clone was called *GhsMOKcl1*. A human placenta genomic library cloned in lambda FIXII vector (Stratagene) was screened with *Sall-Ndel* fragment (330 bp) containing the nonfinger coding region from human genomic *GhsMOKcl1* clone. Hybridization and washing were performed as described above.

A lambda Zap human cDNA library from the Daudi cell line was screened by a nonradioactive PCR-based method with the 5′ primer (5′-ACAAGAACTGGCCTCGCC-3′) and the 3′ primer (5′-GGTTTTTCGCCTGTATGG-3′) located in the coding sequence of *MOK2* at nucleotides 1148–1166 and 1627–1645 (Uze et al. 1990; Israel 1993; Amaravadi and King 1994).

The nucleotide sequence determination was performed by the dideoxy-chain-termination method using deletion mutants obtained by unidirectional digestion with exonuclease III or DNAsel (Sanger et al. 1977; Henikoff 1984; Thomas and Surdin-Kerjan 1990).

Northern Blot and Southern Analysis. Total cellular RNAs were isolated from cell lines by the guanidium thiocyanate procedure (Chomczynski and Sacchi 1987). Polyadenylated RNAs were prepared using oligo(dT) cellulose (Type III, Collaborative Research) columns (Sambrook et al. 1989). The human multiple tissue Northern blot containing equal amounts of poly(A) RNA (2 μg) was purchased from Clontech laboratories. The Northern blots were hybridized with human *Sall-Ndel* fragment labeled with α^{32}P dCTP.

Southern blots containing genomic DNA digested with different restriction endonucleases were hybridized either with a mouse finger region (600-bp *Ddel* fragment, nt 882–1481; Ernoult-Lange et al. 1990) or with a human finger region (1,408-bp *Bgll* fragment, nt 1194–2602) and washed in conditions described above for the genomic screening. The hybridizations of human blot with *Sall-Ndel* fragment (330 bp) containing a part of the nonfinger coding region and the beginning of the first zinc-finger motif from human genomic *GhsMOKcl1* clone and the mouse blot with a part of the mouse 5′ noncoding region of *Mok2* cDNA (nt 443–720; Ernoult-Lange et al. 1990) were done at 42°C in 50 mM Tris-HCl, 1 M NaCl, 0.1% PPi, 10× Denhardt's solution, 10% dextran sulfate, 1% SDS, and 30% formamide for human blot or 50% formamide for mouse blot. The filters were washed four times at 65°C for 15 min in 6× SSC and 0.5% SDS.

RACE Analysis of the 5′ End. Poly(A) RNA (2.5 μg) isolated from HeLa and Raji cells was reverse transcribed with M-MLV reverse transcriptase (BRL). Reactions were primed with random 6-mers (Pharmacia). RACE analysis of the 5′ end was performed as described previously (Rueff-Juy et al. 1991). Two internal primers located at nucleotides 1237–1257 and 1146–1166 of *MOK2* sequence (5′-CCATCCATGAATGTCTTTGC-3′ and 5′-GGCGAGGCCAGTTCTTGTCC-3)′ were used.

In Situ Hybridization. High-resolution chromosome preparations were obtained from phytohemagglutinin-stimulated blood cell cultures of two healthy men after methotrexate synchronization. The *Bgll* fragment (1,408 bp) corresponding to the finger region and the 3′ untranslated sequence of *GhsMOKcl1* genomic clone (nucleotides 1194–2602) were labeled by random primer procedure. Using 20 mCi (^{3}H)-dCTP for 50 ng of DNA, the specific activity was 3×10^7 cpm/μg. The labeled probe was used at the concentrations of 20 and 80 ng/ml. The hybridization procedures were performed as described previously (Caubet et al. 1985). G bands were obtained with Wright's stain.

Computer Analysis. Sequence data treatment was carried out using computer facilities at "Pôle de bioinformatique de Villejuif" (Dessen et al. 1990). Symmetry analysis by dot matrix was performed according to the procedure described in the GCG program (Devereux et al. 1984). The phylogenetic tree was constructed by applying the neighbor-joining method of Saitou and Nei (1987) using PHYLIP package version 3.5. The initial alignment of nucleotide sequences was made using the Clustal W program (Thompson et al. 1994). The nucleotide sequences reported in this paper have been submitted to the EMBL Databank under the following accession numbers: X82126 for the *XbaI-Bgll* fragment from genomic clone, X82125 for the complete cDNA1, X82128 for the 5′ untranslated exon from cDNA2, and X82127 for the 5′ untranslated exon from cDNA3.

Results

Isolation and Structural Features of the Human MOK2 Gene

In order to isolate the human gene homologous to the murine intronless *Mok2* gene, a human chronic myeloid leukemia genomic library was screened under high strin-

gency with a 600-bp *Mok2* fragment that includes the zinc-finger region (nt 882–1481) (Ernoult-Lange et al. 1990). Four positive clones that hybridized with this probe were identified and each one was purified to homogeneity. Restriction and hybridization analysis of these four clones showed that they all contained the same 12-kb genomic fragment. One clone called *GhsMOKcl1* was chosen for further analysis. An *EcoRI-Sall* fragment of 2.6 kb (*Sall* is a restriction site of lambda EMBL3 vector) was subcloned into pBluescript and sequenced. Analysis of the sequence reveals a single long open reading frame showing a high similarity with murine MOK2 zinc-finger region but no in-frame stop codon within the 31 nucleotides preceding the putative initiator methionine. To isolate a 5′ extended clone, we rescreened another genomic library (placenta library) with a different probe obtained from the 5′ end of the human genomic *GhsMOKcl1* subclone. We isolated one clone containing 15 kb of human genomic DNA overlapping on 6.5 kb of the first clone, *GhsMOKcl1*. The sequence of this clone shows a larger open reading frame with two additional in-frame methionines. Several stop codons are present in different reading frames upstream from the first ATG codon. The organization of 10.3 kb of human *MOK2* gene is shown in Fig. 1A. The 9.3-kb fragment localized between the restriction sites *Xbal* and *Bgll* was completely sequenced.

Structural Features of the Human MOK2 Protein

The sequence of the 2,606-bp *Hindlll-Bgll* fragment which contains a single long open reading frame with three in-frame methionines at the 5′ end is shown in Fig. 1B. The first and the third ATGs, respectively at positions 778 and 1069, are in a better context for translation initiation (GNN[ATG]G) than the second ATG at position 919 (GNN[ATG]A) (Kozak 1991). From the first methionine codon, we deduced a protein of 458 amino acids with a pl value of 7.48 and a molecular mass of 51,542 daltons. One out of three cDNAs which have been isolated contain only the third ATG and therefore encode for a smaller protein (see below). This smaller protein consists of 361 amino acids with a molecular mass of 40,709 daltons. The amino acid sequences can be divided into two regions based upon features of the sequence. The NH_2-terminal domain does not contain zinc-finger motifs and consists of 173 amino acids from the first methionine codon. It is an acid domain with a pl value of 4.53. There is no significant overall homology between this NH_2-terminal domain and the available sequences in the protein databases. The COOH-terminal basic domain with a pl value of 8.58 contains 10 zinc-finger motifs. The protein end is identical to the murine MOK2 protein composed of seven tandem Krüppel zinc-finger motifs (consensus sequence: $TGEKPX_2CX_2CX_3FX_5LX_2HX_3H$) with only five additional amino acids at the COOH-terminal. The identity of these seven zinc-finger motifs between human and murine protein is 94%. The amino acids localized between the positions 17 and 23 which are involved in the recognition of DNA bases are highly conserved (Jacobs 1992). Only one amino acid change is observed in the last finger (Fig. 1C). The five amino acids at the COOH-terminal position are not conserved. The human protein contains three additional zinc-finger motifs in tandem with the others. One of these zinc-finger motifs, hsMOK2 –2, is degenerated because the first cysteine is missing and another one, hsMOK2 –3, has the consensus sequence $TEEKPCX_4CX_{10}HX_5H$ (Fig. 1C).

Identification of Human MOK2 cDNAs

Our attempts to screen two different cDNA libraries using standard radiolabeled probe techniques failed to result in the isolation of any *MOK2* cDNA clone. This may be due to the low level of expression of the *MOK2* mRNA. Consequently, we have used the nonradioactive PCR-based method for high-stringency screening of DNA libraries as described by Israel (1993) and Amaravadi and King (1994). Only one cDNA of 2,386 nt terminated with a poly(A) stretch was isolated from the library constructed with mRNA from the Daudi cell line. The size of this cDNA may correspond to a larger RNA of about 2.5 kb found in the brain (see below). The cDNA called cDNA1 is depicted in Fig. 2A. It has no typical polyadenylation signal located upstream from the polyadenylation tail as in the murine *Mok2* cDNA (Ernoult-Lange et al. 1990). The 1,825 nucleotides of this cDNA containing the entire coding sequence with the three ATGs and the 3′ untranslated sequence of 356 nucleotides can be perfectly aligned with the genomic sequence between nucleotides 686 and 2511 (Fig. 1). In the 93 nucleotides located upstream from the three ATGs, several stop codons are present in different reading frames. The 560 nucleotides at the 5′ end show no homology with the 7.45-kb sequenced DNA localized in 5′ of coding exon in genomic clones. Contrary to the mouse *Mok2* gene, the human *MOK2* gene contains a large intron greater than 7.45 kb. The first exon is an untranslated region. In order to determine the 5′ exon/intron boundary, we rescreened the placenta genomic library using this 5′ untranslated exon as a probe. We isolated one clone which contains this 560 nucleotides in a *Xbal* fragment of 5 kb (data not shown). The exon/intron boundaries match perfectly with the GT/AG consensus (Mount 1982).

The "RACE" method of anchored PCR was used in order to determine the 5′ end of cDNAs corresponding to the major mRNAs. After two amplifications with different specific internal primers localized in the coding region, two specific products of 5′ end RACE reactions

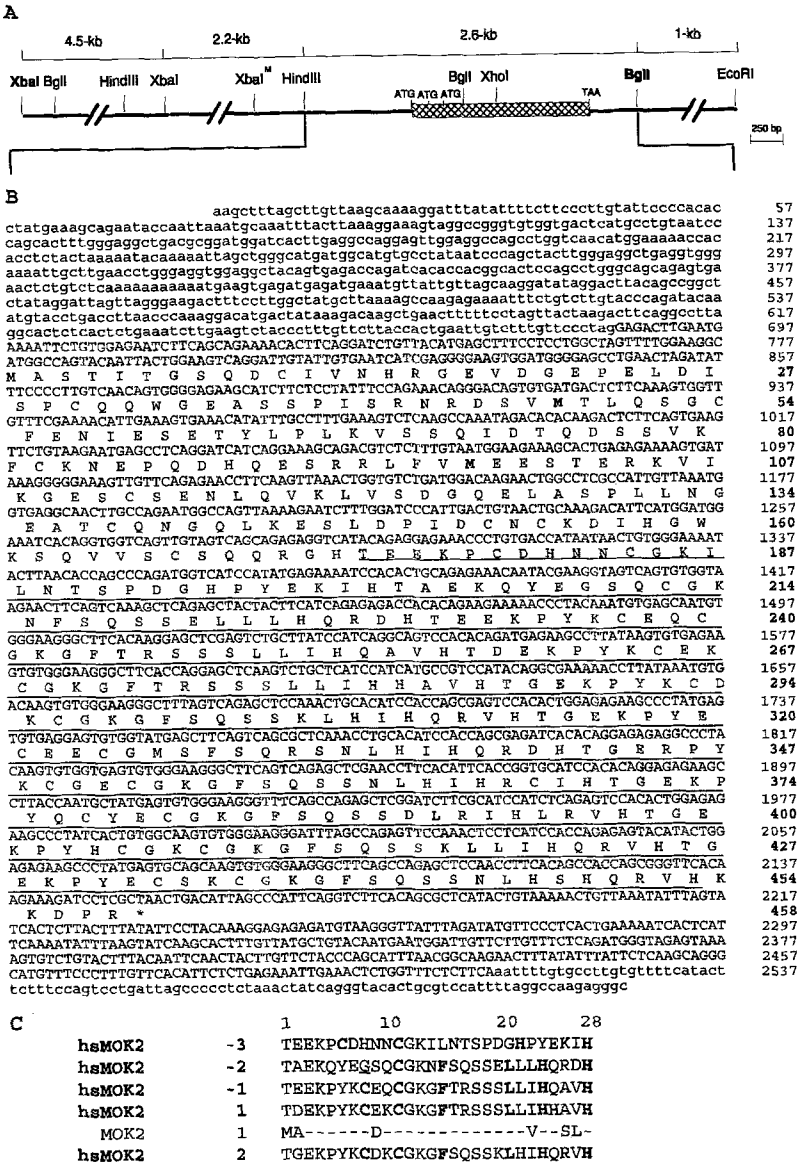

Fig. 1. Nucleotide sequence and predicted amino acid sequence of human MOK2 protein. **A** Schematic restriction map of genomic clone. The coding region is represented by the *hatched box*. The position of translation initiation ATGs and the stop codon TAA are indicated. *XbaI*^{M} indicates a methylated XbaI restriction site. **B** Nucleotide sequence of *HindIII-BglI* genomic fragment obtained from two overlapping genomic clones. The nucleotide sequence written in *capital letters* was identical for the cDNAs. The amino acid sequence is *numbered* from the first ATG codon and the termination codon is marked by an *asterisk*. The three ATGs are in *bold* and the zinc-finger domain is *underlined*. **C** The zinc-finger domain of human MOK2 protein arranged as 10 repeats of 28 amino acid residues; each is aligned with the seven zinc-finger motifs of murine MOK2 protein. The human zinc-finger motifs are called *hsMOK2*. The seven zinc-finger motifs conserved between human and mouse proteins are numbered from *1 to 7*. The three additional zinc-finger motifs in human protein are *numbered negatively*. The conserved residues in Krüppel/TFIIIA zinc-finger family are in *bold*.

were found with mRNA isolated from expressing HeLa cells. No band was observed with mRNA isolated from nonexpressing Raji cells. The two specific products obtained were cloned and sequenced. These two RNAs, represented in Fig. 2A (cDNA2 and cDNA3), have distinct 5′ ends which are different from that of the cDNA1 already isolated. The cDNA2 contains the entire coding exon and the last 123 nucleotides of 5′ untranslated exon of cDNA1. It has 89 nucleotides at the 5′ end which do not share identity with cDNA1, indicating that this exon is interrupted by one intron. The cDNA3 differs in coding sequence because it lacks the sequence that encodes the 97 NH_2-terminal amino acids present within the two other cDNAs. This deletion involves a smaller coding exon beginning at nucleotide 991 which contains only the third ATG described above (Fig. 1). This cDNA thus encodes a protein of 361 amino acids containing a smaller NH_2-acidic domain constituted of 77 amino acids. The 5′ untranslated exon of 122 nucleotides is not the same as the 5′ untranslated exon of cDNA1 and 2. The 3′ ends of these two cDNAs are unknown but they could terminate as cDNA1: the sizes of these cDNAs of 2037

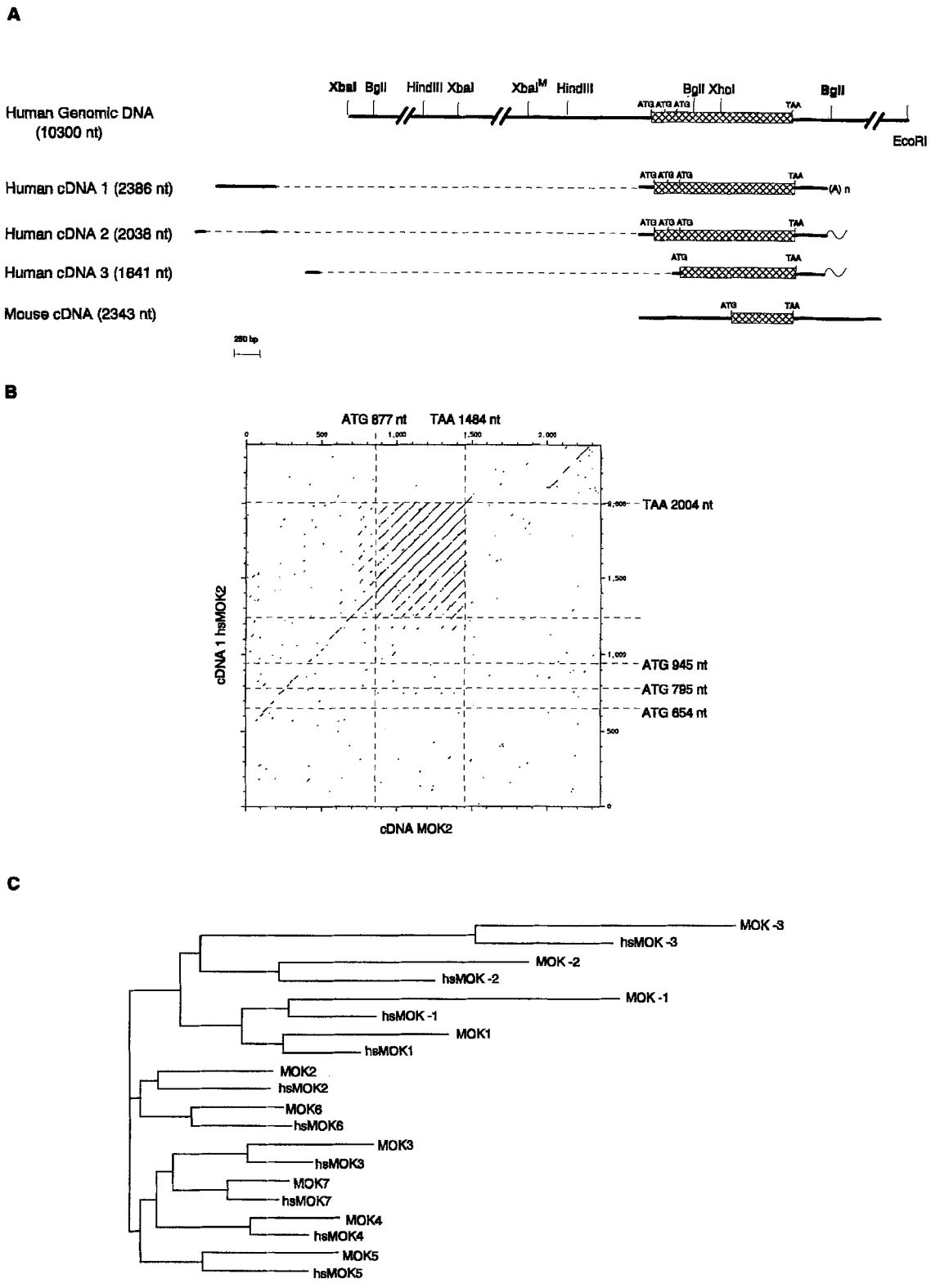

Fig. 2. **A** Schematic representation of the structure of human *MOK2* gene, three corresponding cDNAs, and murine *Mok2* cDNA. The coding region is represented by the *hatched box*. The position of translation initiation ATGs and the stop codon TAA are indicated. *XbaI*^{M} indicates a methylated XbaI restriction site. The *tilde* indicates an undeterminated 3′ end cDNA. **B** *Diagonal dot matrix* comparison between the mouse *Mok2* cDNA and the human *MOK2* cDNA1. The *window* for comparison is >14 identical nucleotides in 21 continuous nucleotides. **C** Phylogenetic analysis of the nucleotide sequence which encodes the ten zinc-finger motifs of human gene and the corresponding mouse sequence. The tree was constructed by applying the neighbor-joining method using the PHYLIP package.

and 1641 of this 3′ end were similar to the hybridizing doublet mRNA of about 2.15 kb (see below).

Comparison of Human and Mouse cDNA Sequences

Figure 2B shows a diagonal dot-matrix comparison between murine *Mok2* cDNA and the human cDNA1. In addition to the strong identity of the seven zinc-finger motifs, the dot-plot analysis reveals an identity in sequences outside of this finger region. The 877 nucleotides of 5′ untranslated *Mok2* sequence share 69.5% identity with the first 863 nucleotides of the human coding exon localized outside of the seven conserved zinc-finger motifs. This result indicates that these two regions are closely related. The 3′ untranslated regions of human and murine cDNAs are respectively 358 nucleotides and 869 nucleotides long. An identity of 76.8% is found between the entire human 3′ untranslated sequence (358 nt) and the end of murine 3′ untranslated sequence. The identity evidenced by the displaced diagonal suggests that either an insertion in the murine gene or a deletion in the human gene has occurred during evolution. Furthermore, the 5′ end of human cDNA1 corresponding to a 5′ untranslated exon in the human gene which does not exist in murine gene shares no obvious sequence identity with the murine *Mok2* cDNA. The same result was found with cDNA2 and cDNA3 (data not shown).

A phylogenetic analysis of the nucleotide sequence which encodes the ten zinc-finger motifs of human gene and the corresponding mouse sequence is shown in Fig. 2C. The tree was constructed by applying the neighbor-joining method using the PHYLIP package. Each of the seven conserved zinc-finger motifs exhibits between 83 and 91% identity, while the three other sequences show between 69 and 71% identity. The tree shows that the nucleotide sequences encoding each of the seven conserved zinc-finger motifs and the nucleotide sequences of the three additional zinc-finger motifs are arranged in pairs. Each pair is composed of a human sequence and the corresponding mouse sequence without any mixing of different zinc-finger sequences. This analysis confirms that the structure of human and murine *MOK2* genes was already present in their common ancestor. The murine sequence localized in 5′ of the seven conserved zinc-finger motifs became an untranslated sequence as, during evolution, numerous mutations have produced 28 stop codons in different reading frames.

Southern Blot Analysis

In order to determine if the human *MOK2* gene is related or homologous to murine *Mok2* gene, the same Southern blot containing human genomic DNA digested with *Hindlll, EcoRI,* or *BamHl* was hybridized under high stringency with zinc-finger probes isolated from mouse *Mok2* cDNA or human *MOK2* genomic clone (Fig. 3). The results show the presence of a unique sequence able to hybridize with the murine and human probes. The murine genomic DNA does not hybridize under low stringency with the *Sall-Ndel* fragment (330 bp) containing a part of the nonfinger coding sequence and the beginning of the first zinc-finger motif of the human genomic *GhsMOKcl1* clone. The two bands detected in human genomic DNA digested by *BamHl* correspond to the presence of a *BamHl* site in the nonfinger coding region. A Southern blot of human genomic DNA digested with nine other restriction enzymes hybridized under high stringency with probe corresponding to the finger region and the 3′ untranslated sequence of the *GhsMOKcl1* genomic clone (*Bgll* fragment, nucleotides 1194–2602) is shown in Fig. 4A. In different lanes, a unique band is hybridized and the size of fragments generated by *Rsal* (1262 bp), *Bgll* (1449 bp), and *Sspl* (3089 bp) corresponds to the size found by restriction mapping of isolated human genomic clones. These results suggest that the human *MOK2* gene is present as a single copy per haploid genome. Furthermore, the two human genomic clones which collectively span a 20.5-kb region contain a unique *MOK2* sequence. The human *MOK2* gene is thus the orthologue of murine *Mok2* gene.

The fact that the murine and human proteins show substantial differences raises the question of the presence of the murine *Mok2* gene as a single or multiple copy per haploid genome. To determine this, a Southern blot containing murine genomic DNA digested with 12 different restriction enzymes was hybridized with a probe corresponding to a part of the 5′ noncoding sequence isolated from *Mok2* cDNA (nt 443–720; Ernoult-Lange et al. 1990). The blot washed under low stringency is shown on Fig. 4B. In different lanes, a unique band is hybridized. The same pattern is found when the blot is hybridized under high stringency with a probe containing the zinc-finger region (data not shown). The size of fragments generated by *Hindlll, Ncol, Rsal,* and *Sacl* corresponds to the restriction map of isolated genomic *Mok2* clone (sizes of fragments generated by *Hindlll,* 1454 nt, *Ncol,* 2778 nt, *Rsal,* 1036 nt, and *Sacl,* 1841 nt). The hybridization signals of single bands in the Southern blot indicate that murine *Mok2* gene exists as a single-copy gene. Furthermore, the mouse genomic clone previously isolated, spanning a 15-kb region, contains a unique *Mok2* sequence localized in the middle of these 15 kb of mouse DNA (Ernoult-Lange et al. 1990). Thus, the murine *Mok2* gene previously isolated is not a pseudogene.

Chromosomal Localization of the Human MOK2 Gene

The chromosomal localization of the human *MOK2* gene was determined in the human genome; 111 metaphases were examined after in situ hybridization with a 1.4-kb

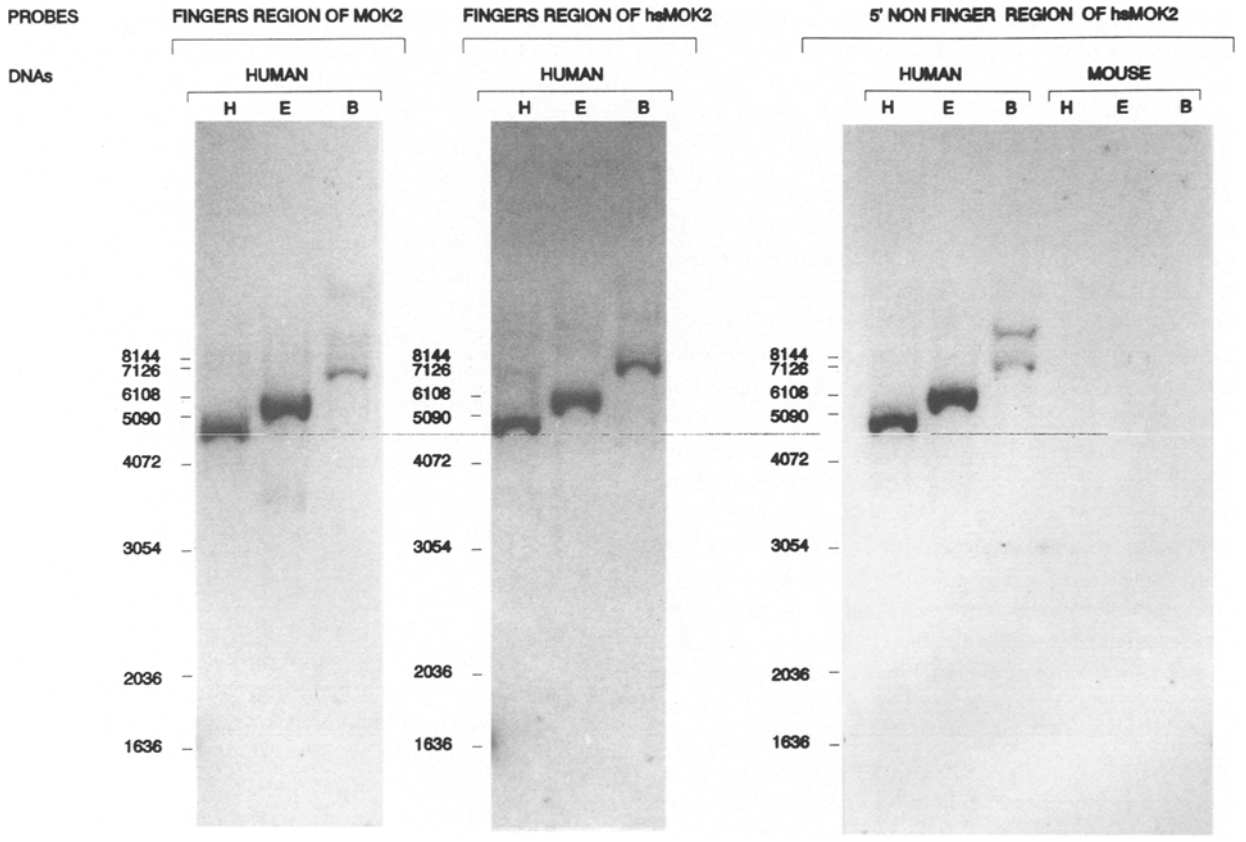

Fig. 3. Genomic Southern blot analysis; 20 μg of human or mouse genomic DNA was digested with different restriction endonucleases, *H* corresponding to *HindIII*, *E* to *EcoRI*, and *B* to *BamHI*. The same human genomic blot was successively hybridized under high stringency with either mouse or human finger region. Another blot with human and mouse genomic DNA was hybridized under low stringency with a fragment containing the nonfinger coding region and the beginning of the first zinc-finger motif from human genomic *GhsMOKcl1* clone (nucleotides 1038–1366, Fig. 1). Positions of size standards are indicated on the *left*.

MOK2 probe corresponding to the finger region and the 3′ untranslated sequence of the *GhsMOKcl1* genomic clone (nucleotides 1194–2602). On a total of 180 silver grains, 37 (20.6%) were located on chromosome 19, 19 (10.6%) were on band 19q13.2, and 12 (6.7%) were on band 19q13.3 (Fig. 5). No other significant peak could be detected. The localization was thus 19q13.2-q13.3.

Expression of Human MOK2 in Various Human Tissues and Cell Lines

The expression of human *MOK2* gene was determined by Northern blot analysis of poly(A) RNA isolated from different human organs and tumor cell lines. The blots were hybridized with a *SalI-NdeI* fragment (328 bp) containing the non-finger coding region and the beginning of the first zinc-finger motif of genomic *GhsMOKcl1* clone (nucleotides 1038–1366). A hybridizing doublet mRNA of about 2.15 kb was detectable at different levels in heart, brain, lung, skeletal muscle, kidney, and pancreas (Fig. 6). In addition, a larger transcript of about 2.5 kb was found in brain where the expression of *MOK2* gene is higher. A high level of *MOK2* transcripts was also detected in testis and breast (data not shown). Placenta and liver tissues do not contain detectable levels of *MOK2* transcripts. Rehybridization of the filters with a β-actin confirmed the RNA integrity and that similar amounts of poly(A) RNA were present in the different lanes. The expression of human *MOK2* gene compared to the expression of mouse *Mok2* gene appears less restricted to certain tissues: a weaker expression is detectable in human lung and kidney while in mouse lung and kidney the expression was undetectable (Ernoult-Lange et al. 1990). Expression of human *MOK2* gene was also examined in different tumor cell lines. The data showed that the *MOK2* gene is expressed at high levels in HeLa and 293 cell lines and at low levels in SW613, C13, and JEG-3 cell lines. The others cell lines do not contain detectable levels of *MOK2* transcripts. The murine *Mok2* gene is preferentially expressed in cells transformed by chemical agents or by DNA tumor viruses such as SV40 or adenovirus (Ernoult-Lange et al. 1990). In human tumor cell lines, an overexpression is only found in HeLa and 293 cells. These two cell lines are known to contain a DNA virus. HeLa is a permanent HPV-18-positive human cervical carcinoma cell line and 293 a permanent

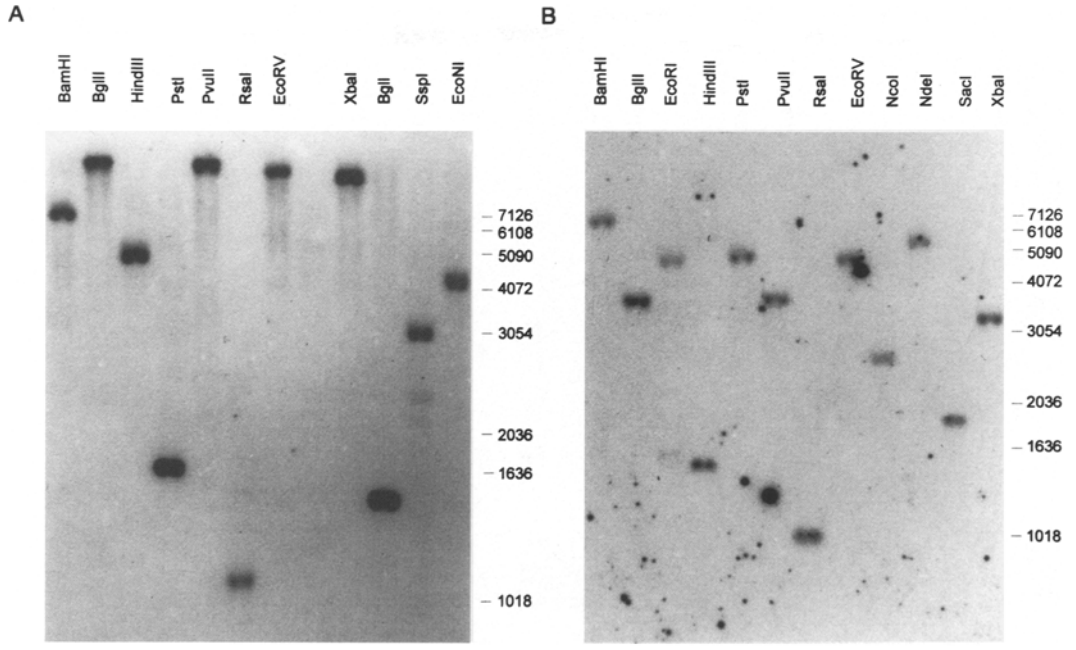

Fig. 4. Genomic Southern blot analysis. **A** Human genomic DNA (6 μg) was digested with 11 different restriction endonucleases. The genomic blot was hybridized under high stringency with the human *Bgl* fragment corresponding to the zinc-finger region and the 3′ untranslated sequence (nt 1194–2602) from *GhsMOKcl1* clone. **B** The genomic blot containing 6 μg of murine genomic DNA digested with 12 different restriction endonucleases was hybridized under low stringency with a fragment corresponding to a part of the 5′ noncoding region isolated from *Mok2* cDNA (nt 443–720; Ernoult-Lange et al. 1990). Positions of size standards are indicated on the *right*.

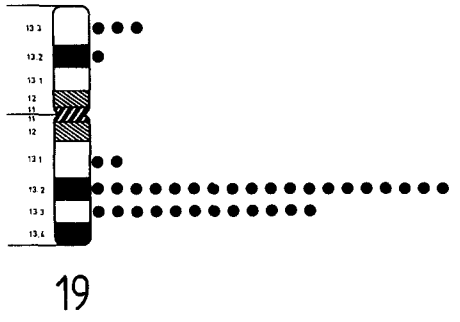

Fig. 5. Distribution of silver grains on human chromosome 19 indicating the distribution of *MOK2* hybridization grains in 111 metaphases.

line of primary human embryonal kidney cells transformed by human adenovirus type 5.

Discussion

In this paper, we describe the isolation and characterization of the human homologue of *Mok2* gene which encodes a Krüppel-like protein. The comparison of human genomic sequence with the three human cDNAs isolated shows that the structure of human *MOK2* gene is more complicated than the murine intronless *Mok2* gene. Indeed, the murine genomic DNA previously isolated had exactly the same sequence as *Mok2* cDNA (Ernoult-Lange et al. 1990). This *Mok2* cDNA has approximately the same size as *Mok2* mRNA. Furthermore, the 5′ flanking region of *Mok2* gene contains a functional promoter (Arranz et al. 1994). We show that human *MOK2* RNA maturation results in three mRNAs with a different 5′ untranslated exon. One of these mRNAs (cDNA2) contains two 5′ untranslated exons. A single exon encodes the MOK2 zinc-finger protein. Comparison of the splicing patterns of the three *MOK2* RNAs shows that they must be generated by alternative mRNA processing.

The human protein shows substantial structural differences with the murine protein. The murine MOK2 protein is composed of seven zinc-finger motifs and five additional amino acids at the COOH-terminal. This structural feature is present at the end of human MOK2 protein. The seven zinc-finger motifs show 94% identity between human and mouse protein. Only one amino acid change in the last finger is observed between positions 17 and 23 involved in the recognition of DNA sequences (Jacobs 1992). The human protein contains three additional zinc-finger motifs in tandem with the others and a nonfinger acidic domain of 173 amino acids at the NH_2-terminal. Members of the Krüppel family have been divided into subfamilies defined on the basis of conserved structural features found outside the finger motif. Such elements include the FAX domain present in a large number of *Xenopus* ZFPs (Knöchel et al. 1989) and the KRAB domain identified in human, mouse, and *Xenopus* ZFPs (Bellefroid et al. 1991; Rosati et al. 1991). The NH_2-terminal domain of human MOK2 protein does not present a significant homology with these two Krüppel subfamilies and with the available sequences in the pro-

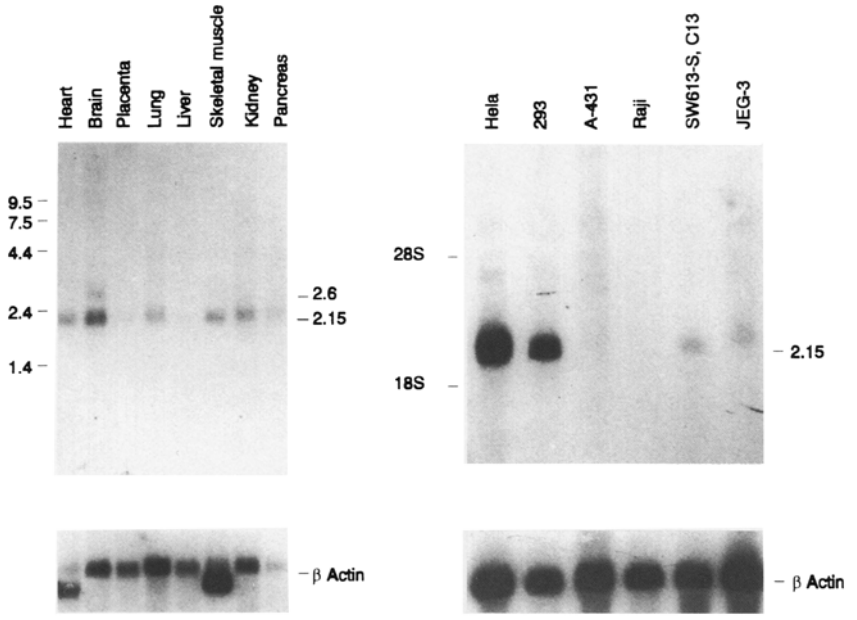

Fig. 6. Northern blot analysis of adult tissues and tumor cell lines. A human multiple tissue blot containing 2 μg of poly(A) RNA purchased from Clontech Laboratories and a multiple tumor cell line blot containing 5 μg of poly(A) RNA were hybridized with a 328-bp restriction fragment of *GhsMOKcl1* genomic clone covering a part of the nonfinger coding region and the beginning of the first zinc-finger domain (nucleotides 1038–1366, Fig. 1). The blots were then stripped and rehybridized with a β-actin probe.

tein databases. The human MOK2 protein contains a potential DNA/RNA-binding domain and a potential activation domain usually found in transcription factors. An acidic domain is characteristic of the activation regions of various transcription factors (Ptashne 1988). It has been proposed that the acidic domains adopt an amphipathic α-helical structure or exist as unstructured "acid blobs" (Giniger and Ptashne 1987; Sigler 1988). Recent reports now suggest that some acidic activation domains present a β-sheet rather than α-helical secondary structure (Leuther et al. 1993; Van Hoy et al. 1993). The human MOK2 acidic domain did not present a significant α-helical or β-sheet secondary structure. This NH_2-domain is lost in mouse MOK2 protein, which contains only the potential DNA- and/or RNA-binding domain. The Southern analysis suggests that these two genes are present as a single copy in the genome and are orthologue. These results indicate that the murine *Mok2* gene previously isolated is not a pseudogene. The human gene has been localized on chromosome 19 on band q13.2–q13.3. Numerous zinc-finger genes have been reported in clusters on chromosome 19 (Bellefroid et al. 1993; Lichter et al. 1992; Hoovers et al. 1992). Thirteen ZFP genes were mapped to 19q13.2-qter (Halford et al. 1995; Hromas et al. 1991; Rousseau-Merck et al. 1993). The mouse *Mok2* gene has been located on chromosome 6 between the markers D6Nds5 and D6Mit24. This region is a new region of homology with human chromosome 19 (Arranz et al. 1996).

The sequence comparison between the murine and human *MOK2* cDNAs shows a high identity in sequences localized outside of the conserved finger region. In particular, 69.5% identity is found between the 5′ untranslated region of murine *Mok2* cDNA and the 5′ untranslated, nonfinger coding region and the first three finger motifs of human coding exon. This result indicates that these two regions are evolutionarily related. The murine sequence became nonfunctional because numerous mutations have produced 28 stop codons in different reading frames during evolution. When we compare the three codons positions separately for the two sequences, the rate of substitution is approximately 10% (data not shown). This result suggests that the murine sequence is subject to weaker selective constraints. Studies on molecular evolution assume that once a gene becomes nonfunctional, the substitution rate is the same for all codon positions (Li et al. 1985). The divergent evolution observed between the human and mouse *MOK2* genes results in the loss of potential transcription activator domain in mouse gene. To our knowledge, this is the first example of an unconserved regulatory domain between two proteins of different species. Some differences were found between highly conserved proteins. For example, the polyserine and polyglycine repeats are absent in the zebrafish Egr1 zinc-finger transcriptional regulator protein, which has a high degree of homology with the mouse, rat, and human proteins (Drummond et al. 1994). Recently, substantial differences have been described at the NH_2-terminal coding region between the *Xenopus laevis* and mouse FIII/YY1 zinc-finger protein (Pisaneschi et al. 1994). This region contains conserved amino acids alternating with divergent regions and long gaps.

In conclusion, we have described a molecular evolution between the human and mouse orthologue genes involving an important modification of structural features of MOK2 protein with the loss of the potential transactivating domain in murine protein. This modification should result in an alteration of the function of the MOK2 protein between the two species. The next step toward understanding the precise function of the human

and mouse proteins will be to determine the specific nucleotide sequence they recognize.

Acknowledgments. We thank Michèle Meunier and Laurent Duret for helpful discussions and comments on the manuscript. This research was supported by a grant from Association de la Recherche sur le Cancer.

References

Amaravadi L, King MW (1994) A rapid and efficient nonradioactive method for screening recombinant DNA libraries. Bio techniques 16:98–104

Arranz V, Kress M, Ernoult-Lange M (1994) The gene encoding the MOK-2 zinc-finger protein: characterization of its promoter and negative regulation by mouse *Alu* type-2 repetitive elements. Gene 149:293–298

Arranz V, Kress M, Ernoult-Lange M (1996) Localization of zinc finger *Mok2* gene to mouse Chromosome 6, a new region of homology with Chromosome 19. Mammalian Genome (in press)

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 Krüppel type. Mol Cell Biol 8:377–387

Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA (1991) The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. Proc Natl Acad Sci USA 88:3608–3612

Bellefroid EJ, Marine JC, Ried T, Lecocq PJ, Riviere M, Amemiya C, Poncelet DA, Coulie PG, Dejong P, Szpirer C, Ward DC, Martial JA (1993) Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T-lymphoid cells. EMBO J 12:1363–1374

Call KM, Glasser T, Ito CY, Buckler AJ, Pelletier J, Haber DA, Rose EA, Kral A, Yeger M, Lewis AH, Jones C, Housman D (1990) Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilm's tumor locus. Cell 60:509–520

Caubet J, Mathieu-Mahul D, Bernheim A, Larsen CJ, Berger R (1985) Human proto-oncogene c-mos maps to 8q11. EMBO J 4:2245–2248

Chavrier P, Janssen-Timmen U, Mattei MG, Zerial M, Bravo R, Charnay P (1989) Structure, chromosome location, and expression of the mouse zinc finger gene Krox-20: multiple gene products and coregulation with the protooncogene c-fos. Mol Cell Biol 9:787–797

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guadinium thiocyanate-phenolchloroform extraction. Anal Biochem 162:156–159

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

Crosby SD, Puetz JJ, Simburger KS, Fahrner TJ, Milbrandt J (1991) The early response gene NGFI-C encodes a zinc finger transcriptional activator and is a member of the GCGGGGGCG (GSG) element-binding protein family. Mol Cell Biol 11:3835–3841

Dessen P, Fondrat C, Valencien C, Mugnier C (1990) BISANCE: a French service for access to biomolecular sequence databases. Comput Appl Biosci 6:355–356

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Drummond IA, Rohwernutter P, Sukhatme VP (1994) The zebrafish egr1 gene encodes a highly conserved zinc-finger transcriptional regulator. DNA and Cell Biology 13:953–961

El-Baradi T, Pieler T (1991) Zinc finger proteins: what we know and what we would like to know. Mech Dev 31:155–169

Engelke DR, Ng SY, Shastry BS, Roeder RG (1980) Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell 19:717–728

Ernoult-Lange M, Kress M, Hamer D (1990) A gene that encodes a protein consisting solely of zinc finger domains is preferentially expressed in transformed mouse cells. Mol Cell Biol 10:418–421

Gessler T, Poustka A, Cavenee W, Neve RL, Orkin SH, Bruns GAP (1990) Homozygous deletion in Wilm's tumours of a zinc finger gene identified by chromosome jumping. Nature 343:774–778

Giniger E, Ptashne M (1987) Transcription in yeast activated by a putative amphipatic α helix linked to a DNA binding unit. Nature 330:670–672

Halford S, Mattei M, Daw S, Scambler PJ (1995) A novel C2H2 zinc-finger protein gene (ZNF160) maps to human chromosome 19q13.3-q13.4. Genomics 25:322–323

Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351–359

Hoovers JMN, Mannens M, John R, Bliek J, Vanheyningen V, Porteous DJ, Leschot NJ, Westerveld A, Little PFR (1992) High-resolution localization of 69 potential human zinc finger protein genes—a number are clustered. Genomics 12:254–263

Hromas R, Collins SJ, Hickstein D, Raskind W, Deaven LL, Ohara P, Hagen FS, Kaushansky K (1991) A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. J Biol Chem 266:14183–14187

Israel DI (1993) A PCR-based method for high stringency screening of DNA libraries. Nucleic Acids Res 21:2627–2631

Jacobs GH (1992) Determination of the base recognition positions of zinc fingers from sequence analysis. EMBO J 11:4507–4517

Kinzler KW, Ruppert JM, Bigner SH, Vogelstein B (1988) The GLI gene is a member of the Krüppel family of zinc finger proteins. Nature 332:371–374

Knöchel W, Poting A, Köster M, El-Baradi T, Nietfeld W, Bouwmeester T, Pieler T (1989) Evolutionarily conserved modules associated with zinc fingers in *Xenopus laevis*. Proc Natl Acad Sci USA 86:6097–6100

Kozak M (1991) An analysis of vertebrate mRNA sequences: intimations of translational control. J Cell Biol 115:887–903

Köster M, Pieler T, Pöting A, Knöchel W (1988) The finger motif defines a multigene family represented in the maternal mRNA of Xenopus laevis oocytes. EMBO J 7:1735–1741

Lechalony C, Prosperi MT, Haluza R, Apiou F, Dutrillaux B, Goubin G (1994) The OZF gene encodes a protein consisting essentially of zinc finger motifs. J Mol Biol 236:399–404

Leuther KK, Salmeron JM, Jonston SA (1993) Genetic evidence that an activation domain of GAL4 does not require acidity and may form a β-sheet. Cell 72:575–585

Li WH, Luo CC, Wu CI (1985) In: Molecular evolutionary genetics: evolution of DNA sequences, vol 1. MacIntyre RJ (eds) Plenum Press, New York, pp 1–94

Lichter P, Bray P, Ried T, Dawid IB, Ward DC (1992) Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile site regions of human chromosomes. Genomics 13:999–1007

Miller J, McLachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J 4:1609–1614

Modjatahedi N, Haddada H, Lamonerie T, Lazar E, Lavialle C, Brison O (1992) TGF-alpha production correlates with tumorigenicity in clones of the SW613-S human colon carcinoma cell line. Int J Cancer 52:483–490

Mount SM (1982) A catalogue of splice junction sequences. Nucleic Acids Res 10:459–472

Nietfeld W, El-Baradi T, Mentzel H, Pieler T, Köster M, Pöting A, Knöchel W (1989) Second-order repeats in Xenopus laevis finger proteins. J Mol Biol 208:639–659

Pelham HR, Brown DD (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. Proc Nat Acad Sci USA 77:4170–4174

Pieler T, Bellefroid E (1994) Perspectives on zinc finger protein function and evolution—an update. Mol Biol Rep 20:1–8

Pisaneschi G, Ceccotti S, Falchetti ML, Fiumicino S, Carnevali F, Beccari E (1994) Characterization of FIII/YY1, a Xenopus laevis conserved zinc-finger protein binding to the first exon of L1 and L14 ribosomal protein genes. Biochem Biophys Res Commun 205: 1236–1242

Ptashne M (1988) How eucaryotic transcriptional activators work. Nature 335:683–689

Rosati M, Marino M, Franze A, Tramontano A, Grimaldi G (1991) Members of the zinc finger protein gene family sharing a conserved N-terminal module. Nucleic Acids Res 19:5661–5667

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

Rousseau-Merck MF, Hillion J, Jonveaux P, Couillin P, Seite P, Thiesen HJ, Berger R (1993) Chromosomal localization of 9 Kox zinc finger genes—physical linkages suggest clustering of Kox genes on chromosomes 12, 16, and 19. Hum Genet 92:583–587

Rueff-Juy D, Marche P, Drapier A, Cazenave P (1991) Junctional diversity of H and L chains allows the coexpression of two mutually exclusive idiotopes (Idl104 and Idl558). J Immunol 146:4024–4030

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Ecol 4:406–425

Sambrook J, Fritsch E, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schuh R, Aicher WK, Gaul S, Cote S, Preiss A, Maier D, Seifert E, Nauber U, Schröder C, Kemler R, Jäckle H (1986) A conserved family of nuclear proteins containing structural elements of the finger protein encoded by Krüppel, a Drosophila segmentation gene. Cell 47:1025–1032

Sigler PB (1988) Acid blots and negative noodles. Nature 333:210–212

Sukhatme VP, Cao XM, Chang LC, Tsai-Morris CH, Stamenkovich D, Ferreira PC, Cohen DR, Edwards SA, Shows TB, Curran T (1988) A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. Cell 53:37–43

Thomas D, Surdin-Kerjan Y (1990) An improved strategy for generating of unidirectional deletions on large DNA fragments. GATA 4:87–90

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Uze G, Lutfalla G, Gresser I (1990) Genetic transfer of a functional human interferon alpha receptor into mouse cells: cloning and expression of its cDNA. Cell 60:225–234

Van Hoy M, Leuther KK, Kodadek T, Johnston SA (1993) The acidic activation domains of the GCN4 and GAL4 proteins are not α helical but form β sheets. Cell 72:587–594