Volume 16 Number 21 1988 Nucleic Acids Research

Specific and ubiquitous expression of different Zn finger protein genes in the mouse

Kamal Chowdhury*, Heidi Rohdewohld+ and Peter Gruss

Max-Planck Institute for Biophysical Chemistry, 3400 Göttingen, FRG

Received August 23, 1988; Revised and Accepted October 13, 1988 Accession nos X12592-X12594 (incl.)

ABSTRACT

Zinc finger proteins (Zfp) are members of a multigene family encoding Zn mediated nucleic acid binding proteins. They have been isolated from various organisms including yeast(1-3), Drosophila (4-6), Xenopus(7-10) mouse(11-12) and human (13-14). All Zfp share the 28-30 amino acid long finger repeats containing conserved residues at specific positions. Some of these proteins(7,13) have been identified as transcriptional regulatory factors. In this paper,we describe the isolation,DNA sequence determination and the expression pattern in developing embryos and adult tissues of 3 new members of the mouse Zfp. All of them are expressed as multiple transcripts. Unaltered level of mkr5 expression could be detected in 10-15 day whole embryo RNAs but its level started to decrease from day 16. In the adult animal,predominant expression was detected only in the ovary. In contrast,mkr3 and 4 were expressed at a constant level in all embryos and tissues tested. These data suggest the presence of both tissue specific and ubiquitious Zfp in the mouse.

INTRODUCTION

Many eucaryotic genes involved in the process of differentiation and development are expressed in a temporally and spatially restricted fashion(15). Considering that regulation of such genes also occur at the transcriptional level,the identification of different transcriptional regulatory factors should provide further insight into the molecular biology of development. One possible way how these factors might function is through their binding to specific DNA sequences. These DNA-protein complexes may also subsequently interact with other involved factors via protein-protein or DNA-protein binding to regulate gene expression.

Depending on the structural motifs,the known DNA binding regulatory proteins can be classified mainly into two groups. The first group

contains the helix-turn-helix motif present in λ repressor or other prokaryotic DNA binding proteins(16) and in the eukaryotic homeo domain proteins (17,18). The DNA binding ability of homeo domain proteins has also been demonstrated (19,20).

The other group contains a different DNA binding motif,commonly known as the Zn fingers. The potential finger like structures were first recognized to be present in the transcription factor III A (TFIIIA) of Xenopus laevis by Brown et al (21) and Miller et al(8). This motif containing the consensus amino acid sequence F/YXCXXCXXXFXXXXXLXXHXXXHTGEKP is present as multiple copies in many proteins(22-25) and usually depicted as a folded finger like peptide structure around a Zn atom. In addition to the conserved residues present in the consensus motif,it has been postulated to contain an α-helical region(see table 1). In TFIIIA, a DNA,RNA and Zn binding transcriptional factor (8,26,27), there are 9 such fingers,each 30 amino acids long containing 2 Cys and 2 His residues at invariant positions. These 2 Cys and 2 His residues are thought to build collectively the tetrahedral core around a Zn ion (28) and the sequences in between loop out to form a finger structure. A second and

TABLE 1- COMPARISON OF MOUSE Zn FINGER MOTIFS

	α-Helix
mkr3 (15)	- f/ye **C** k e **C** G K a **F** t x S S x **L** x x **H** x R x **H** **T** **G** **E** **K** **P**
mkr4 (14)	- Y v **C** x x **C** G K A **F** t q k S x **L** x x **H** q x I **H** **T** **G** **E** **K** **P**
mkr5 (10)	- f e **C** x x **C** g k a **F** x x x x x **L** x x **H** q R i **H** **T** **G** **E** **K** **P**
mkr1 (7)	- y e **C** x e **C** g k t **F** x x x s n **L** i x **H** q r i **H** **T** **G** **E** **K** **P**
mkr2 (9)	- Y x **C** x e **C** g K a **F** x x x s s **L** t x **H** q r i **H** **T** **G** **E** **K** **P**
krox4 (5)	- y e **C** x e **C** G k x **F** x q k S x **L** t x **H** q r t **H** **T** **G** **E** **K** **P**
krox6 (6)	- Y v **C** x E **C** g K A **f** x c S S y **L** t k **H** q R I **h** **x** **g** **e** **k** **p**
krox8 (3)	- Y x **C** p x **C** G K s **F** s x r S n **L** x a **H** x R t **H** **x** **G** v **k** **p**
krox9 (3)	- Y E **C** x e **C** g k a **F** R L l S q **L** t q **H** q S I **H** **t** **G** **E** **k** **P**
krox20- (2)	f x **C** x x **C** x R x **F** s R S D e **l** t r **H** i r i **H** **T** **G** x **K** **P**
Consensus	Y e C x e C G K a F x x x S x L t x H Q R I H T G E K P

Table1.

Comparison of the consensus mouse Zinc finger motifs. In order to compare the sequences from these multifinger genes, consensus finger sequences for each gene was first determined. Mkr1 and 2 sequences are from (11). The Krox sequences were taken from (12). Capital letters indicate more than 75% conservation and lower case letters indicate 50% or less conservation. Bold letters indicate strict conservation of amino acid residues among all fingers.

more complex model involving an antiparallel β-ribbon followed by an α-helix has been proposed(29) that differ from the common Zn finger structure. In a more recent model Gibson et al(25) utilized computer simulated molecular dynamics and interactive model building to propose a somewhat different three dimensional structure.It consists of a two stranded β hairpin stabilizing a C-terminal α-helix by both Zn ligands and hydrophobic interaction.

Several other DNA binding proteins containing clustered Cysteins have been postulated to contain 1 or 2 Zn fingers. They include HIV tat protein, Adenovirus E1A and several nuclear receptors(reviewed in 30). These proteins differ from the multiple Zn finger containing proteins (TFIIIA, Kr, mkr,Sp1etc) in lacking Zn finger(s) mediated specific DNA binding (31). Furthermore,they lack the conserved Phe and Leu residues in the finger loop and the H-C link peptide TGEKP.

Interestingly, when Krüppel(Kr), the Drosophila segmentation gene of the gap class(32) was isolated and sequenced,it was found to contain 5 Zn fingers(5),thereby implicating a role of DNA binding finger structures(33) in gene regulation during Drosophila development. Mutation in one of the metal binding Cys residue to Serine resulted in the loss of Kr^+ function(34). At least some of the Zn finger proteins have been shown to be transcriptional activators or factors involved in the regulation of transcription(7,13).

To understand the potential role of multifinger proteins in higher eucaryotes,it is necessary to isolate and analyze the corresponding genes. Several murine(11,12),human(14),and X.laevis(10) genes encoding multiple fingers and belonging to a multigene family have been described. One of the genes,mkr2 is expressed specifically in the neuronal cells of developing and adult mice (35). Another gene,Krox-20 is activated during G0/G1 cell cycle transition in cultured cells(36).Human testis determining factor(TDF) has also been shown to contain 13 Zn fingers(37). In this paper,we describe the isolation, sequence characterization and expression pattern of 3 more mouse genes encoding multiple fingers. Mkr3 and mkr4 transcripts are present as multiple mRNAs in developing mouse embryos and in all adult mouse tissues tested but in contrast mkr5 expression is predominantly restricted in the adult animal to the ovary. These data suggest the occurance of both ubiquitious and spatially restricted Zn finger containing proteins in the adult mouse tissues.

MATERIALS AND METHODS

Cloning of mouse genes with multiple Zinc fingers

An 8.5 day p.c. C57BL whole mouse embryo cDNA library in λ gt10 (38) was screened with a cloned 562 bp EcoR1 finger encoding fragment

from mkr1,another mouse multifinger gene (11). Approximately $4X10^5$ phages were plated out and after transferring the plaques to nitrocellulose filters (Schleicher & Schüll),the phage DNAs were denatured and immobilized according to the standard procedures(39). Hybridization to oligo labeled ^{32}-P probe was performed at 65^0 C in 3XSSC,5X Denhardt.0.1%SDS and 100µg/ml denatured Salmon sperm DNA for 18h(1XSSC= 0.15M NaCl,0.015M Na-Citrate). The filters were washed twice at 65^0 C in 2XSSC,0.1%SDS, and once at room temp. in 0.1XSSS,0.1%SDS, for 30 minutes, air dried and exposed at -70^0 C for 2 days. The positives phages were plaque purified by replating and hybridization.

Mini preparations of phage DNA from high titer lysates for Southern blotting (40).

Ten gm of DEAE Sephacel(Pharmacia) was suspended in 200 ml of 0.05N HCl and the pH of the suspension was neutralized by adding about 400 µl of 10N NaOH. After decanting the smaller particles, the DEAE Sephacel was washed 3 - 4 times with 5 volumes of 0.14 M NaCl. Finally, the particles were resuspended such that it contained 75% DEAE and 25% 0.14 M NaCl (V/V). High titer liquid or plate lysates (0.6 ml) were mixed with equal volumes of well resuspended DEAE Sephacel in eppendorf tubes,mixed thoroughly and the bound bacterial DNA and RNA was precipated by spinning in a microfuge for 5 minutes at room temperature. In case of plate lysates,it is necessary to reextract the supernatant with 0.6 ml of DEAE. To denature the phage protein 120 µl each of 10mM Tris(pH8.0),2.5%SDS and 0.25M EDTA was added to each tube and incubated at 70^0 C for 15 minutes. The samples were allowed to cool to room temperature (10min) and the protein was precipitated for 15 minutes on ice by adding 75 µl of 5M potassium acetate. After spinning for 5 minutes at room temp. the protein pellets were discarded and the phage DNA from the supernatant(0.8 ml) was precipitated by mixing with 0.5 ml of isopropanol for 2 minutes. The DNA pellets were collected by spinning for 5 minutes at room temp.,washed with 70% ethanol, dried and suspended in 50µl of 10mM Tris(7.5),1mM EDTA. Depending on the titre of the lysate,the recovery of phage DNA was between 1-4 µg.

For southern blotting 1µg of individual phage DNA was cleaved with EcoR1,electrophoresed and blotted onto nitrocellulose papers according to the standard procedures(39). The probe,hybridization and washing conditions were the same as described for screening the cDNA library.

DNA sequencing

All inserts from the cDNA clones were excised with different restriction enzymes and recloned in plasmid vector. The overlapping

DNA fragments were sequenced by a combination of Maxam and Gilbert (41) and the chain termination method using the Sequenase sequencing kit from USB (Cleveland,Ohio).

RNA Isolation (42)

RNA samples were isolated from cultured cells by lysing in 4M guanidium thiocyanate. Frozen adult mouse tissues or embryos were mechanically disintegrated with polytron homogenizer in guanidium thiocyanate. After passing several times through sterile needles to shear the cellular DNA, the lysate was loaded on a 4 ml cushion of 5.7 M CsCl,25mM sodium acetate (pH 5.0) and spun in a Beckman SW 41 rotor at 20^0 C and 31000 rpm for 22 hours. The total RNA pellets were dissolved in 10mM Tris (pH 7.0),1mM EDTA and after phenol treatment precipitated with 2.5 volumes of ethanol. Poly A^+ RNAs were isolated by retention on oligo dT(collaborative research)columns.

Northern blotting.

About 5 µg of cellular or tissue poly A^+ RNAs were loaded in each lane in a 1% agarose gel in 3.7% formaldehyde and MOPS buffer (20 mM morpholine propane sulfonic acid,50 mM sodium acetate,10 mM EDTA, pH 7.0). The electrophoresis buffer also contained the same amount of formaldehyde and MOPS and was recirculated during elctrophoresis. After blotting the RNAs onto Gene screen plus filter papers (Dupont-NEN) in 10XSSC,the membrane was baked at 80^0 C for 2 hours and hybridized for 18 h at 42^0 C to ^{32}-P oligo labeled individual probes. The hybridization mixture contained 50% formamide,1M NaCl,1% SDS and 100ug/ml denatured salmon sperm DNA. Filters were washed twice in 2XSSC,1%SDS at 65^0 C for 30 minutes and once in 0.1XSSC,1%SDS at 65^0 C for 30 minutes. The same adult mouse tissue or embryonic poly A^+RNA blot was used for hybridization with mkr3,mkr4 and mkr5 probes. For removal of the hybridized probe,the filters were treated with 1%SDS in water at 70^0C for 2 hours,air dried and exposed to detect the absence of any remaining signal. The following cloned DNA fragments were used as probes for northern blotting hybridyzation. Mkr3 : nt.1-700; mkr4 : nt.293-1258; mkr5 : nt.390-980. All nucleotide numbers are from figure 1.

RESULTS

Isolation of 3 mouse cDNA clones encoding multiple Zn fingers

Previous cloning and Southern blotting data (11) indicated the presence of a multigene family encoding repeated Zn fingers in the mammalian genome. Subsequently,it has been possible to isolate several such genes (11,12) from the mouse genome under low stringency hybridization with the Drosophila multiple finger containing Kr gene

(5) as a probe. To isolate finger genes expressed early during mouse embryogenesis, we have screened an 8.5 day p.c. whole mouse embryo cDNA library. A DNA segment from the mouse multifinger gene mkr1(11) encoding several fingers was used as the hybridization probe. The resulting positive phages were isolated and plaque purified. For further screening by nucleic acid hybridization, their DNA were isolated from 0.6 ml of high titre plate lysates. Restriction enzyme analysis,Southern blotting and hybridization to mkr1 probe (data not shown) indicated the presence of 3 potentially new mouse multifinger genes. To establish their identity,they were further analysed by DNA sequencing and this led to the identification of 3 new mouse multiple Zn finger containing genes designated mkr3, mkr4 and mkr5. The sizes of the cDNA clones were 2517 bp,1258 bp and approximately 2.5 kbp for mkr3, mkr4 and mkr5 respectively.

Size determination of the respective transcripts by northern blotting revealed that none of the cDNA clones were full length (see below). The DNA and the conceptually translated protein sequences of the 3 clones are depicted in figure 1. Each clone encoded multiple 28 amino acid long finger domains containing the conserved metal binding Cys and His residues at specific positions as in Drosophila Kr(5) ,mkr1 and mkr2 (11,35) and as represented by the consensus sequence Y/FXCXXCXXXFXXXXXLXXHXXXHTGEKP (X represents a variable amino acid). In addition,the Tyr/Phe,Phe,Leu and the H-C link sequence TGEKP is also conserved.

Structural analysis of mkr3,mkr4 and mkr5

Mkr3 contains a total of 428 amino acids including 14 full and 1 partial fingers(fig.1 and 2). At the 3′ end,there is a termination codon at nt. 1307 preceeded by 27 unique non finger amino acids. Downstream from the termination codon,there is a long untranslated sequence of 1207 nucleotides containing several stop codons in all reading frames and at nt 1361 the ATTTA sequence motif which is potentially involved in the regulation of mRNA stability by degradation (43). No poly A tail was present at the 3′end of mkr3. But judging from the absence of an open reading frame after the stop codon at nt 1307,we conclude that the mkr3 protein terminates at this position. However,the actual 3′end of mkr3 lies most likely beyond nt 2517 as indicated by the lack of poly adenylation signal and poly A tail in the untranslated region.

The mkr4 partial protein sequence consists of 419 amino acids including 13 full and 1 partial fingers (fig.1,2). Similar to the Drosophila Hunchback gene (6) it can be divided into two finger domains separated by a unique domain. The N terminal portion consists of 15 unique amino acids followed by 3 fingers. The middle region consists of 28 unique non-finger amino acids which is followed by 10 fingers.Since the

isolated mkr4 cDNA sequence do not contain any start or stop codons in the open reading frame,we consider this protein sequence to be incomplete at both N- and C- terminal ends. A potential nuclear localization signal PRKKV (44) is present in the unique middle region.

The mkr5 DNA sequence shown in fig.1 and 2 is a composite of 1

A. mkr3

```
1     GAATTCATGT CTTACGGTCA AGG ATA AAA ACT CAC ACT GGT GAG AAG CCC TTC CGG TGT AAA GTA TGT GCA AGG ACT
1                               Ile Lys Thr His Thr Gly Glu Lys Pro Phe Arg Cys Lys Val Cys Ala Arg Thr

78    TTT AGA AAT TCC TCA TGC CTT AAG ACT AAC TTT CGA ATT CAC ACT GGA ATA AAA CCA TAC AAA TGT AAT TAC TGT
19    Phe Arg Asn Ser Ser Cys Leu Lys Thr Asn Phe Arg Ile His Thr Gly Ile Lys Pro Tyr Lys Cys Asn Tyr Cys

153   TGG AAA GCC TTC ACT GCA CGC TCA GGC CTT ACT AAG CAT GTA CTA ATT CAC AAT GGT GAG AAG CCC TAT GAG TGT
44    Trp Lys Ala Phe Thr Ala Arg Ser Gly Leu Thr Lys His Val Leu Ile His Asn Gly Glu Lys Pro Tyr Glu Cys

228   AAG GAG TGT GGG AAA GCC TTC AGT ACA TCT TCC GGC CTT GTT GAA CAT ATA AGA ATT CAT ACA GGA GAG AAG CCC
69    Lys Glu Cys Gly Lys Ala Phe Ser Thr Ser Ser Gly Leu Val Glu His Ile Arg Ile His Thr Gly Glu Lys Pro

303   TTT GAA TGT TAC CAG TGT GGG AAA GCC TTG GTC CAT TCC TCA TCC CTT GTT GGA CAT TTA AGA ACT CAC ACT GGA
94    Phe Glu Cys Tyr Gln Cys Gly Lys Ala Leu Val His Ser Ser Ser Leu Val Gly His Leu Arg Thr His Thr Gly

378   GAG AAA CCC TTT GAG TGT AAT CAG TGT GAC AAA ACT TTT ACA CGA TCT TCT TAT CTT CGT ATT CAT ATG CGA ACT
119   Glu Lys Pro Phe Glu Cys Asn Gln Cys Asp Lys Thr Phe Thr Arg Ser Ser Tyr Leu Arg Ile His Met Arg Thr

453   CAC ACC GGA GAG AAA CCA TAT GAA TGT AAA GAG TGT GGG AAA ACT TTC CCT GAG CGC TCA TGC CTT ACT AAA CAC
144   His Thr Gly Glu Lys Pro Tyr Glu Cys Lys Glu Cys Gly Lys Thr Phe Pro Glu Arg Ser Cys Leu Thr Lys His

528   ATA AGG ACA CAC ACT GGT GAA AGG CCC TAT GAA TGT AAG GAG TGT GAC AAA GGC TTT ATT AGC TTT GCT CAG CTT
169   Ile Arg Thr His Thr Gly Glu Arg Pro Tyr Glu Cys Lys Glu Cys Asp Lys Gly Phe Ile Ser Phe Ala Gln Leu

603   ACT GTA CAC ATA AAA ACT CAT AGT TCT GAG AGA CCC TTT CAG TGT AAG GTA TGC ACA AAA TCT TTT AGA AAC TCT
194   Thr Val His Ile Lys Thr His Ser Ser Glu Arg Pro Phe Gln Cys Lys Val Cys Thr Lys Ser Phe Arg Asn Ser

678   TCA TCC CTT GAG ACC CAC TTC CGA ATT CAC ACT GGA GTA AAA CCC TAT AAA TGC AGT TAC TGT GGG AAA GCC TTC
219   Ser Ser Leu Glu Thr His Phe Arg Ile His Thr Gly Val Lys Pro Tyr Lys Cys Ser Tyr Cys Gly Lys Ala Phe

753   ACT GCT CGT TCA GGC CCT ACC ATC CAT TTA CGC AAT CAC ACT GGG GAG AAA TCC TAT GCA TGC CAA GAG TGT GGA
244   Thr Ala Arg Ser Gly Pro Thr Ile His Leu Arg Asn His Thr Gly Glu Lys Ser Tyr Ala Cys Gln Glu Cys Gly

828   AAG GCC TTT AGC ACT TCC TCA GGC CTT ATT GCA CAT ATA AGA AGT CAC AAA GGA GAG AAA CCC TTT GAA TGT GAC
269   Lys Ala Phe Ser Thr Ser Ser Gly Leu Ile Ala His Ile Arg Ser His Lys Gly Glu Lys Pro Phe Glu Cys Asp

903   CAC TGT GGG AAG GCC TTT CCT TCT TCA TCT TAT CTT AAT GTG CAT TTG AAA ATT CAC ACT GGG GAA AAA CCC TTT
294   His Cys Gly Lys Ala Phe Pro Ser Ser Ser Tyr Leu Asn Val His Leu Lys Ile His Thr Gly Glu Lys Pro Phe

978   CGG TGT ACG GTA TGT GGG AAA ACA TTT ACG TGT TCT TCT TAC CTT CCG GTT CAC ATG CGA ACT CAC ACT GGA GGG
319   Arg Cys Thr Val Cys Gly Lys Thr Phe Thr Cys Ser Ser Tyr Leu Pro Val His Met Arg Thr His Thr Gly Gly

1053  AGG CCT TTT CGG TGT ATA ATA TGT GGA AGG TCA TTT TTG TGG TCC TCG TAC CTT AGA GTT CAC ATG CGA ATT CAC
344   Arg Pro Phe Arg Cys Ile Ile Cys Gly Arg Ser Phe Leu Trp Ser Ser Tyr Leu Arg Val His Met Arg Ile His

1128  ACT GGA GAG AAA CCC TAT GTA TGT CAG TAC TGT GGA AAA GCT TTT ACA GAG CAC TCA GGC CTT AAT AAA CAT TTA
369   Thr Gly Glu Lys Pro Tyr Val Cys Gln Tyr Cys Gly Lys Ala Phe Thr Glu His Ser Gly Leu Asn Lys His Leu

1203  CGG AAA CAC ACA GGA GAG AAA CCA TAT GAA TAT AAG GAA TGT GGG GAA AAC TTC ACT ACT TCT GCT GAT GCT AAT
394   Arg Lys His Thr Gly Glu Lys Pro Tyr Glu Tyr Lys Glu Cys Gly Glu Asn Phe Thr Thr Ser Ala Asp Ala Asn

1278  GAA CAT GAA ACT CCC CAC TGG GGA GAC CTT TGA ATGTAAGGGA TCAGAAAATT GATAGCCGTT GACCATCATT TGCAATCAAA
419   Glu His Glu Thr Pro His Trp Gly Asp Leu Opa

1361  ATTTATAGTC TATACAAGTC TTATTGATAA GCAACGTATA AAAAACATTC GGTTGCTATG TGCTGTAAAA ACAACTCATT CTGAGGATGC
1451  TCTCAGTTTA AGGATTGTGG GAGAAATCAA AATTTCATGA TATATGTATG AACTGAAGGT TTGTAATATA GGGAACGAAA ACCTTTAGTG
1541  TTTCTATGAG TGTTGATGCA AGACATGGAG AAATGTTTCT ATGTTCACGT GTGGTCTCAC AGTTCGAGAA ATGTCTTTCT GAAGTTTGGG
1631  AAGTAACTTT ACATTTGAAC ACAAGCTACA TGCGTGTAAG AGGGGATTTG CTTCAGTGTC TTGGATGGCA ACACTTGGCT TTCTTGGGTG
1721  CTTCTCTATG AACTTTGAAT ATTTCTCCAT TTGTGTATTT TTATGACCTC AATTCCATTG TTAGTTTTCC TTTTTAAAAT ATTCCTGTTT
1811  GATGTGATTG AAACTTGTTT TTATCTAGTA TCTGTTATAT GTCCCATGTT GGCATTGTGA ACCTAATGAC ACAAACTCCC ATCTCCTCAC
1901  TACTTGTTAG TTTATTTGTT GTTGTTTTTT TCCCCTCTTT TCTTTGGAGA CAGTCTCACT TTCTGGCCTA GTATAACATG AAAGTTATGT
1991  AGCCTAGACT GGTCCCTAGC TCATGGCAGT CTAGACCCAG CCTTCTGAGT GCATAGATTA CAGGTGCAAG CCACCATACC T*GCTCTTTC
2081  TACTGTTTTA AGTATAAACT TCACTAGCAC TGAAAAAACC ATCATGTTTT ACAATTGTGT TCAGGTTCTT AACACAAGTA GGTTTGGATC
2171  TGACCAATAC AATCCCTTAC TCTGAGTTCT AAGTCTTTGG TAGCCATTCC ATTCTCTGCT TGGAGTTCAG CTTTTCTTTA GGCTTTGCTT
2261  CTTAGATCAT GTAGTATCTG TATTAACTCT ACTTAGCTTT TTTATTTTTG TGACAGGGTC TTGCTGTAGC TCAAACTGAC CTCAAACTCG
2351  CAAACCTTTT TGCCTCAGAC TTACTGCTAC AGTTATAGGC ATGTGCCTTC ATGCCTAGCT GGCGTATATC AATCTACATA ATGTTCTCAA
2441  GGTGGAACTG TGTTATGGTC AGTTATTTTA AGACATCTAA TTATCCTTTG GCCTCCTTGA CCGTAAGACA TGAATTC
```

B. mkr4

```
1      AAA GTC CTT TTT CCA AGT ATC TTC TCT GTT CAG GCA CCG GAG ATC CAC ACT GGA GAA AAG CTC TAC GAC TGT AGC
1      Lys Val Leu Phe Pro Ser Ile Phe Ser Val Gln Ala Pro Glu Ile His Thr Gly Glu Lys Leu Tyr Asp Cys Ser

76     CAT TGT GGG AAA GGC TTC TCT TAC AAC TCA GAC CTC AGG ATA CAT CAG AAG ATC CAT ACA GGA GAG AAA CGC CAT
26     His Cys Gly Lys Gly Phe Ser Tyr Asn Ser Asp Leu Arg Ile His Gln Lys Ile His Thr Gly Glu Lys Arg His

151    GGC TGC GTC GAC TGT GGC AAA GCC TTC ACA CAA AAG TCC ACA CTG AGG ATG CAC CAG AAG ATC CAC ACG GGC GAG
51     Gly Cys Val Asp Cys Gly Lys Ala Phe Thr Gln Lys Ser Thr Leu Arg Met His Gln Lys Ile His Thr Gly Glu

226    AGG GCC TAT GTC TGT ATT GAG TGT GGA CAG GCT TTT ATC CAG AAG ACA CAC CTG GTC GCA CAC CGA AGA ATT CAC
76     Arg Ala Tyr Val Cys Ile Glu Cys Gly Gln Ala Phe Ile Gln Lys Thr His Leu Val Ala His Arg Arg Ile His

301    TCA AGA GTG AGG CCC TGT GTA TCT CTC GAC CGT GCG AAA CCC TTC AGC AGC GCT CCC AAC CTC CTT CCG CGT AAG
101    Ser Arg Val Arg Pro Cys Val Ser Leu Asp Arg Ala Lys Pro Phe Ser Ser Ala Pro Asn Leu Leu Pro Arg Lys

376    AAA GTT CAA ATG AGA GAG AAA TCG TCC ATC TGC GCG GAG TGT GGG AAG GCC TTT ACC TAC AGG TCG GAG CTA ATC
126    Lys Val Gln Met Arg Glu Lys Ser Ser Ile Cys Ala Glu Cys Gly Lys Ala Phe Thr Tyr Arg Ser Glu Leu Ile

451    ATT CAT CAG AGA ACT CAC ACG GGA GAG AAG CCT TAT CAG TGC GGG GAC TGT GGC AAA GCC TTC ACC CAG AAG TCA
151    Ile His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Gln Cys Gly Asp Cys Gly Lys Ala Phe Thr Gln Lys Ser

526    GCC CTC ACC GTG CAC CGA AGA ATC CAC ACG GGG GAG AAG TCG TAT GTG TGT GTG AAG TGC GGG CTA GCC TTC GTG
176    Ala Leu Thr Val His Arg Arg Ile His Thr Gly Glu Lys Ser Tyr Val Cys Val Lys Cys Gly Leu Ala Phe Val

601    CAA AGG GCA CAC TTG GAT GCA CAC CAG GTG ATT CAC ACT GGA GAA AAA CCT TAC CAG TGT GGT CAT TGT GGG AAA
201    Gln Arg Ala His Leu Asp Ala His Gln Val Ile His Thr Gly Glu Lys Pro Tyr Gln Cys Gly His Cys Gly Lys

676    TTC TTC ACT TCC AAG TCG CAA CTC CAC GTG CAC AAG CGG ATT CAC ACG GGG GAA AAA CCC TAT GTG TGC AGT AAC
226    Phe Phe Thr Ser Lys Ser Gln Leu His Val His Lys Arg Ile His Thr Gly Glu Lys Pro Tyr Val Cys Ser Asn

751    TGT GGG AAG GCC TTT GCC AAC AGG TCA AAT CTC ATC ACA CAT CAG AAA ACT CAT ACA GGG GAG AAA GCC TAT GTC
251    Cys Gly Lys Ala Phe Ala Asn Arg Ser Asn Leu Ile Thr His Gln Lys Thr His Thr Gly Glu Lys Ala Tyr Val

826    TGT GCA AGG TGT GGG AAA GCA TTC ACT CAG AGG TCA GAA CCT GTT ACA CAC CAG AGG ATA CAT ACC GGA GAG AAG
276    Cys Ala Arg Cys Gly Lys Ala Phe Thr Gln Arg Ser Glu Pro Val Thr His Gln Arg Ile His Thr Gly Glu Lys

901    CCA TAT GGG TGT AGA CCC TGC GGG AAA GCC TTT ACC CAG AAA TCA CAC CTC AGT ATA CAC GAG AAG ATT CAC ACC
301    Pro Tyr Gly Cys Arg Pro Cys Gly Lys Ala Phe Thr Gln Lys Ser His Leu Ser Ile His Glu Lys Ile His Thr

976    GGA GAG AGA CAG TAC GGA TGC CGA GAC TGT GGG AAA GCC TTC AAC CAG AAA TCC ATA CTC ATC GTG CAC CAG AAA
326    Gly Glu Arg Gln Tyr Gly Cys Arg Asp Cys Gly Lys Ala Phe Asn Gln Lys Ser Ile Leu Ile Val His Gln Lys

1051   ATC CAC ACG GGG GAG AAA CCG CAG GTG TGT GCC GAG TGC GGG CGA GCT TTC ATC CGC AAG TCC AAC TTC ATC ACC
351    Ile His Thr Gly Glu Lys Pro Gln Val Cys Ala Glu Cys Gly Arg Ala Phe Ile Arg Lys Ser Asn Phe Ile Thr

1126   CAT CAG CGC ATC CAT ACT GGC GAG AAA CCT TAT GGA TGC ACC GAC TGC GGC AAG TCC TTT ACC TCC AAG TCG CAG
376    His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Gly Cys Thr Asp Cys Gly Lys Ser Phe Thr Ser Lys Ser Gln

1201   CTC CTA GTG CAC CGG CCA ATC CAC ACC GGC GAG AAA CCT TAC GTG TGT GCC GAG TGT G
401    Leu Leu Val His Arg Pro Ile His Thr Gly Glu Lys Pro Tyr Val Cys Ala Glu Cys
```

C. mkr5

```
1      CAG CAG TCT GTA CAG GAA GCC CAG GAT CTG CTT CCA AGG CAG GAC TCA CAT GCT GAA CGG GTA ACA GGC AGA ACT
1      Gln Gln Ser Val Gln Glu Ala Gln Asp Leu Leu Pro Arg Gln Asp Ser His Ala Glu Arg Val Thr Gly Arg Thr

76     TGG AGC ACT AAA CTT GAG TGC TCC ACT TTC AGA GAT CAG GAT TCT GAG TGT ACG TTT GAA AGG AAT GAG CAA GAG
26     Trp Ser Thr Lys Leu Glu Cys Ser Thr Phe Arg Asp Gln Asp Ser Glu Cys Thr Phe Glu Arg Asn Glu Gln Glu

151    ACA GTC ACT CCG AAC AGA GCC TTC TCA GAG GGG AGA GAT GGT ATG TGT ATC GAA TCT GGA AGA TGG TTT CAT TTG
51     Thr Val Thr Pro Asn Arg Ala Phe Ser Glu Gly Arg Asp Gly Met Cys Ile Glu Ser Gly Arg Trp Phe His Leu

226    AAC AGT TCA GAC GAG AGA AGT CAT AAC TGT GAC TCA GGT AAA AGT TTC TCC TCA AAT CCA GTA GTT GTA AAG GAA
76     Asn Ser Ser Asp Glu Arg Ser His Asn Cys Asp Ser Gly Lys Ser Phe Ser Ser Asn Pro Val Val Val Lys Glu

301    ACC GGA ATC TGT TCA GGA AAA AAA CTT TTC CAA TGT AAT GAA TGT AAG AAA ACT TTC ACC CAG AGC TCC TCC CTT
101    Thr Gly Ile Cys Ser Gly Lys Lys Leu Phe Gln Cys Asn Glu Cys Lys Lys Thr Phe Thr Gln Ser Ser Ser Leu

376    ACC GTT CAC CAG AGA ATT CAT ACT GGA GAG AAG CCC TAT AAA TGC AAT CAG TGT GGG AAG GCC TTC AGT GAC GGG
126    Thr Val His Gln Arg Ile His Thr Gly Glu Lys Pro Tyr Lys Cys Asn Gln Cys Gly Lys Ala Phe Ser Asp Gly
```

```
451   TCC TCC TTC GCA CGG CAC CAG AGG TAC CAT ACA GGC AAG AAG CCG TAC GAG TGC CCG GAG TGT GGC AAA GCT TTC
151   Ser Ser Phe Ala Arg His Gln Arg Tyr His Thr Gly Lys Lys Pro Tyr Glu Cys Pro Glu Cys Gly Lys Ala Phe

526   ATA CAG AAC ACC TCG CTC GTT CGT CAC TGG AGG TAC TAT CAC ACT GGG GAG AAA CCC TTC GAT TGC ATC GAC TGT
176   Ile Gln Asn Thr Ser Leu Val Arg His Trp Arg Tyr Tyr His Thr Gly Glu Lys Pro Phe Asp Cys Ile Asp Cys

601   GGG AAA GCC TTC AGT GAC CAC ATA GGG CTT AAT CAA CAC AGG AGG ATT CAC ACT GGA GAG AAA CCA TAC ACG TGT
201   Gly Lys Ala Phe Ser Asp His Ile Gly Leu Asn Gln His Arg Arg Ile His Thr Gly Glu Lys Pro Tyr Thr Cys

676   GAA GTG TGT CAC AAA TCC TTT AGG TAC GGC TCA TCC CTC ACT GTG CAC CAA AGG ATT CAT ACT GGA GAG AAA CCG
226   Glu Val Cys His Lys Ser Phe Arg Tyr Gly Ser Ser Leu Thr Val His Gln Arg Ile His Thr Gly Glu Lys Pro

751   TAC GAG TGT GAG ATT TGC AGA AAA GCC TTC AGC CAC CAT GCA TCC CTC ACT CAG CAT CAA CGA GTG CAT TCT GGA
251   Tyr Glu Cys Glu Ile Cys Arg Lys Ala Phe Ser His His Ala Ser Leu Thr Gln His Gln Arg Val His Ser Gly

826   GAA AAG CCT TTT AAA TGT AAA GAG TGT GGG AAA GCT TTT AGG CAG AAT ATA CAC CTT GCT AGT CAC TGG AGG ATC
276   Glu Lys Pro Phe Lys Cys Lys Glu Cys Gly Lys Ala Phe Arg Gln Asn Ile His Leu Ala Ser His Trp Arg Ile

901   CAT ACC GGG GAG AAG CCC TTT GAG TGT GGG GAA TGT GGG AAA TCT TTC AGC ATC AGC TCA CAG CTT GCC ACG CAT
301   His Thr Gly Glu Lys Pro Phe Glu Cys Gly Glu Cys Gly Lys Ser Phe Ser Ile Ser Ser Gln Leu Ala Thr His

976   CAG AGA ATT CAT ACA GGA GAG AAG CCC TTT GAG TGT AAG GTT TGT AGG AAA GCT TTT AGG CAG AAT ATA CAC CTT
326   Gln Arg Ile His Thr Gly Glu Lys Pro Phe Glu Cys Lys Val Cys Arg Lys Ala Phe Arg Gln Asn Ile His Leu

1051  GCT AGT CAC TGG AGA ATC CAT ACC GGG GAG AAG CCC TTT GAG TGT GGG GAA TGT GGG AAA TCT TTC AGC ATC AGC
351   Ala Ser His Trp Arg Ile His Thr Gly Glu Lys Pro Phe Glu Cys Gly Glu Cys Gly Lys Ser Phe Ser Ile Ser

1126  TCA CAG CTT GCC ACG CAT CAG AGA ATT CATACAGG AGAGAAGCCT TATGAATGTA AGGTTTGTAG GAAAGCATTT ACGCAGAAGG
376   Ser Gln Leu Ala Thr His Gln Arg Ile

1211  CTCACCTTGC ACAACATCAG AAGACTCATA CGGGGGAGAA GCCATACGAG TGCAAGGAAT GTGGCAAGGC CTTCAGCCAG ACCACTCACC
1301  TCATTCAACA CCAGAGGGTT CATACTGGAG AGAAGCCCTA TAAATGCCTG GAATGTGGGA AGGCCTTTGG TGATAACTCA TCCTGTACTC
1391  AACATCAGAG ACTTCACACT GGCCTGAGGC CTTATGAGTG TGTGGAGTGT GGAAAGACAT TCAAGACCAA GTCATCCCTT ATCTGTTATT
1481  GTAGATGTCA CACGGGAGAG AAACCTTATG AGTGCAGTGC ATGTGGTAAA GCCTTTAGCC ACTGGCAGTC CCTTAGTGTA CATCAGAGA

1570  ATT CAT TCG GGG AAG AAA CCG TAT GAA TGT AAG GAA TGT CGG AAA ACC TTC ATC CAG ATT GGA CAC CTT AAC CAA
385   Ile His Ser Gly Lys Lys Pro Tyr Glu Cys Lys Glu Cys Arg Lys Thr Phe Ile Gln Ile Gly His Leu Asn Gln

1645  CAT AAG AGA GTC CAT ACC GGG GAG AGA ACG TAC AAC TAC AAG AAG GGC AGG AGG GCC TTC AGG CAG ACT GCA CAC
410   His Lys Arg Val His Thr Gly Glu Arg Thr Tyr Asn Tyr Lys Lys Gly Arg Arg Ala Phe Arg Gln Thr Ala His

1720  TTT GCT CAC CAT CAG CAG ATT CAT TCC GGA AAG TCA CCT GCT CAC CAC TCT TTG CCT TCC ACA TCG AAT CCT GTG
435   Phe Ala His His Gln Gln Ile His Ser Gly Lys Ser Pro Ala His His Ser Leu Pro Ser Thr Ser Asn Pro Val

1795  GAT CTC TTC TCC AAA TTC GTC TGG AAT CCA TCC TCA CTG CCA TCA TCA TAA CCTCA TATTTTCACC ACACTTGTTT
460   Asp Leu Phe Ser Lys Phe Val Trp Asn Pro Ser Ser Leu Pro Ser Ser Och

1871  ATCAACTTGT CTCTTTGCCC CTTTGTTCAT TCTTGTCCCT TATGAGTCCT TCATATTACA ACCAGGTTGG ATTATTTTTG GAGTGTAAAT
1961  GTAATTAATT TGCTCATTTA AAGCTTATAT TCACTGGGGA AAAAAAAAGA AGTTAGCTGA AACATAAAGA AACCTCATGA ATCTAAGAAA
2051  TGTTCCATAT CTTGCTCTGG GTAGTATATC TGTAAATCTT AGCAAGCCCT GTATGTAGTA TTTTAAGAAC TTTCTGACTT TGTTATCCTT
2141  TAATTTTTAA TGCTGTTATA TTAAGAAAAA CACATTTATA ATAAATATGG TATTTTAAAT AAAAAGTATT ATATCTTATA GCTATTTTTC
2231  AGATAAACAA AAACAAATTT GATACAAAGT AAAATATGTA TAAGGTTGGT GTTCCAGAAC AAAATCTGCA GGAATATTTA CAGCATGGTA
2321  TAAAGCAACG AAACCAGGTC ATCCACCAAC ATTTTGCCTT AGATTTTCTA ATATTTCAGC ATATGAGTAG CACAGTATTA AAACACTATA
2411  GGTAGTAGCA TTTCACAGAA TTC
```

Figure 1.

The nucleotide and the conceptually translated protein sequences of mkr3, 4 and 5. Mkr3 and 4 are cDNA sequences and mkr5 represents a composite of cDNA and genomic sequences.The genomic sequence shown hier has not been conceptually translated, since its colinearity with the cDNA sequence remains to be established. The first and the last 23 nucleotides in the mkr3 DNA sequence represent the oligonucleotide linker sequence used for the cDNA cloning. Poly adenylation signals AATAAA and the sequence motif ATTTA, potentially involved in the mRNA degradation are underlined. Potential glycosylation sites(NXS) are marked by stars. The finger domains are bracketed.

mkr3

```
                                    --IKT H TGEKP
FR C KV C ART F RNSSC L KT N FRI H TGIKP
YK C NY C WKA F TARSG L TK H VLI H NGEKP
YE C KE C GKA F STSSG L VE H IRI H TGEKP
FE C YQ C GKA L VHSSS L VG H LRT H TGEKP
FE C NQ C DKT F TRSSY L RI H MRT H TGEKP
YE C KE C GKT F PERSC L TK H IRT H TGERP
YE C KE C DKG F ISFAQ L TV H IKT H SSERP
FQ C KV C TKS F RNSSS L ET H FRI H TGVKP
YK C SY C GKA F TARSG P TI H LRN H TGEKS
YA C QE C GKA F STSSG L IA H IRS H KGEKP
FE C DH C GKA F PSSSY L NV H LKI H TGEKP
FR C TV C GKT F TCSSY L PV H MRT H TGGRP
FR C RI C GRS F LWSSY L RV H MRI H TGEKP
YV C QY C GKA F TTSAD A NE H ETP H WGDL*
```

mkr4

```
                    --KVLFPSIFSVQAPEI  H  TGEKL
YD C SH C GKG F SYNSD L RI   H QKI   H  TGEKR
HG C VD C GKA F TQKST L RM   H  QKI  H  TGERA
YV C IE C GQA F IQKTH L FA   H  RI   H   SRVRP
CVSLDRAKPFSSAPNLLPRKKVQMREKS
SI C AE C GKA F TYRSE L II   H QRT   H  TGEKP
YQ C GD C GKA F TQKSA L TV   H RRI   H  TGEKP
YV C VK C GLA F VQRAH L DA   H QVI   H  TGEKP
YQ C GH C GKF F TSKSQ L HV   H KRI   H  TGEKP
YV C SN C GKA F ANRSN L IT   H QKT   H  TGEKA
YV C AR C GKA F TQRSEPVT     H QRI   H  TGEKP
YG C RP C GKA F TQKSH L SI   H EKI   H  TGERQ
YG C RD C GKA F NQKSI L IV   H QKI   H  TGEKP
QV C AE C GRA F IRKSNFIT     H QRI   H  TGEKP
YG C TD C GKS F TSKSQ L LV   H RPI   H  TGEKP
YV C AE C  --
```

mkr5

```
                  ---QQSVQEAQDLLPRQDSHAERVTG
RTWSTKLECSTFRDQDSECTFERNEQETVTPNRAFSEGRDNC
IESGRWFHLNSSERSHNCDSGKSFSSNPVVVKETGICSGKKL
FQ C NE C KKT F TQSSS L TV H QRI   H TGEKP
YK C NQ C GKA F SDGSS F AR H QRW   H TGKKP
YE C PE C GKA F IQNTS L VR H WRWW H TGEKP
FD C ID C GKA F SDHIG L NQ H RRI   H TGEKP
YT C EV C HKS F RYGSS L TV H QRI   H TGEKP
YE C EI C RKA F SHHAS L TQ H QRV   H SGEKP
FK C KE C GKA F RQNIH L AS H WRI   H TGEKP
FE C GE C GKS F SISSQ L AT H QRI   H TGEKP
FE C KV C RKA F RQNIH L AS H WRI   H TGEKP
FE C GE C GKS F SISSQ L AT H QRI  --------
------------------------------ I  H SGKKP
YE C KE C RKT F IQIGH L NQ H KRV   H TGERT
YNYKKGRRAFRQTAHFAHHQQIHSGKSPAHHSLPSTSNPVDL
FSKFVWNPSSLPSS*
```

Figure2.

Protein sequences of mkr3, 4 and 5 in one letter amino acid code showing the 28 residue long Zn finger repeats. Dashes represent the sequences not done. Stars represent stop codon.

genomic and 2 cDNA clones.The longer cDNA (c4-2) sequence starts at nt 1 and ends at the EcoR1 site at nt. 1148. The smaller cDNA (G3-1) sequence is located between the EcoR1 site at nt 1568 and the poly A tail of 25 A residues at nt 1997. The genomic DNA sequence contains 9 A residues at this position in the untranslated 3'end. To bridge the sequence gap present between the two cDNA clones,a genomic DNA clone from a mouse genomic library was isolated using a mkr5 finger containing fragment as a probe and partially sequenced. The genomic DNA sequence starts at nt 389 and is colinear with the cDNA sequence between nts 389 to 1148 and 1569 to 1999. An additional sequence of 417(139 in frame amino acids) bp was present between the two cDNA clones. Furthermore,at the 3' end additional 434 bp was sequenced beyond the poly A site at nt 1999 in G3-1. Available mkr5 cDNA sequence encodes 475 amino acids including 109 unique residues at the N-terminal end followed by 11 fingers and terminating with 55 unique C-terminal amino acids. The stop codon is located at nt 1842. If the genomic DNA sequence is also included,the total coding capacity of mkr5 would be 614 amino acids. Interestingly,one of the fingers encoded by the genomic sequence is of C_2- CH type. The unique C-terminal end contains a potential nuclear localization signal KKGRR,previously shown to be present in another Zn finger protein rfp (45). The untranslated 3'end of the cDNA and genomic DNA contains several poly adenylation signals at nts 1960,2177 and 2195.The sequence motif ATTTA, present at nt 2171 is potentially involved in mRNA regulation by degradation (43).

Amino acid sequence comparison of various mouse Zn finger motifs

In addition to the previously described mkr1 and mkr2 clones, Chavrier et al have reported the isolation of several other mouse genomic and cDNA sequences encoding multiple Zn fingers (12). To determine whether mkr3,4 and 5 represent new mouse finger genes,a comparison with all the published mouse finger sequences was performed. A summary of the data is shown in table1. Clearly, the sequences of mkr3,4 and 5 are different from mkr1,2(11) and krox4,6,8,9 and 20(12). They therefore represent new mouse finger genes. For simplicity,only comparison of the consensus finger motifs from different genes are shown. The unique region did not contain any obvious homology and was therefore omitted from the table.The consensus finger motifs are highly homologous, but clearly not identical.Most of the variability was found in the sequences located between the conserved Cys and His residues comprising the "finger tips". This region also contains the α-helical structure postulated previously(25).The individual finger sequences present in each clone was also different from each other. Strikingly,in addition to the Zn binding Cys and His residues,in all the mouse finger genes described so far,the

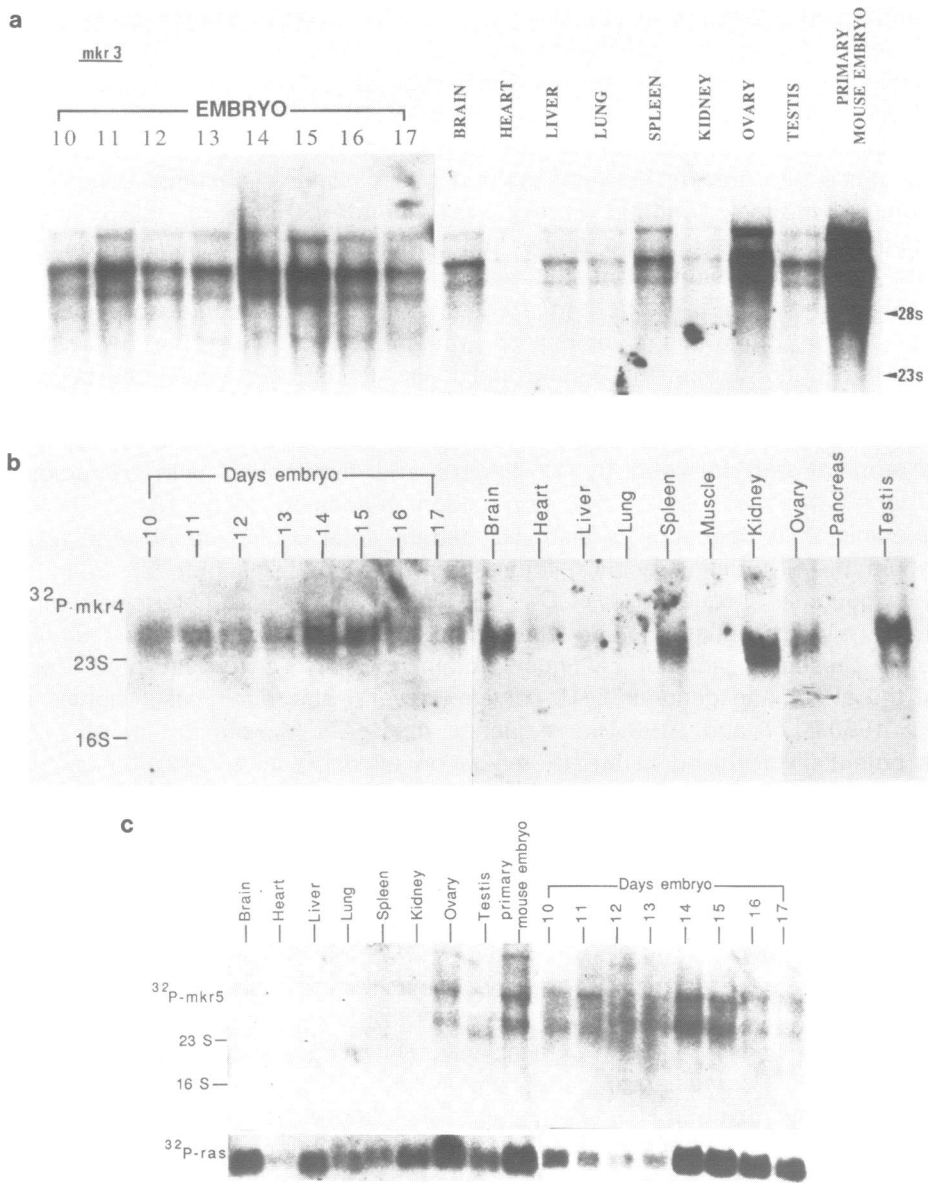

Figure3.

Expression of mkr3, 4 and 5 in developing mouse embryos and adult tissues analysed by northern blotting technique.In all cases,except for the adult mouse tissue blot in figure 3a,the same filters were used as described in materials and methods. Internal hybridization control for the amount of RNA loaded in each lane is shown in figure 3c.

Phe,Leu and the strech of TGEKP peptides were conserved (Table1), suggesting a possible functional role.

Mkr3.4 and 5 are expressed during murine embryogenesis and in adult mouse tissues as multiple transcripts

Since the clones were isolated from an 8.5 day mouse embryo cDNA library, they can be expected to be expressed early during mouse embryogenesis. To analyze the fate and nature of these transcripts during further development of the mouse embryo, northern blotting experiments were performed using poly A^+ RNA isolated from 10-17 day old total mouse embryos and adult mouse tissues. In each case, the cloned probe used for hybridization was first tested by southern blotting experiments with stringent washing (0.1XSSC,65^{0}C) and found to detect only a single band in the mouse genomic DNA cleaved by various restriction enzymes (data not shown). RNA blotting revealed multiple messages of different sizes produced by all 3 clones.

Mkr3 probe detected in 10,11and 12 day embryos 3 descrete size messages of 11.5 kb,8.5 kb and 6.5 kb (fig.3). Interestingly,a new transcript of 4.5 kb appeared first in day 13 embryo and was found to be present in all subsequent day embryos tested. All transcripts were expressed in developing embryos at constant level. However,the levels of 11.5 and 4.5 kb transcripts were in each case 3-5 fold lower than the 8.5 and 6.5 kb transcripts. A similar pattern of expression comprising the 4 transcripts was also observed in all adult mouse tissues tested (fig.3).

Mkr4 probe detected 2 closely migrating transcripts of 4.2 and 3.8 kb(fig.3). They were also expressed at a constant level from day 10-17. The levels of these 2 transcripts in embryonic and adult tissues are extremely low. Since,using the same blot and probe of similar specific activity,mkr3 transcripts were visible on the autoradiogram after 30 hour exposure. But in contrast,under similar experimental conditions only faint mkr4 trancripts could be seen even after 7 day exposure.Both mkr4 transcripts were present in all adult mouse tissues tested.

Two transcripts of 3.1 kb and 4.4 kb were detected by the mkr5 probe in 10-17 day embryos. Their level remained constant in 10-15 day embryos. An overall decrease of 1-2 fold was observed in 16-17 day embryos(fig.3). In contrast to mkr3 and mkr4, a tissue specific expression was observed for mkr5. Of all the adult mouse tissues tested,only ovary had a predominant expression of the two mkr5 transcripts. The only other tissue expressing them at a much lower level was testis. In addition,testis had a slightly faster migrating message.

DISCUSSION

All Zn finger proteins(Zfp) share a structurally similar nucleic acid binding motif. However,in contrast to the proteins of the

helix-turn-helix family ,the Zfp are capable of binding not only to DNA but also to metal and RNA as shown for TFIIIA. Furthermore, helix turn helix motif recognizes as dimers only a very small region of DNA comprising several nucleotides,whereas Zfp may bind to different length of DNA depending on the number of fingers present. Compared to the helix turn helix motif, the finger regions may be quite long. In case of Xfin (40), 88% of the entire protein consists of 36 such repeats. Why such large number of fingers are required for the recognition of specific DNA sequence is unclear. This kind of multiple contact may be necessary for the binding stability of the complex which is subsequently utilized by RNA polymerase for transcription. Moreover, they may be involved in the binding of specific RNA sequences and serve as RNA storage proteins. A clear understanding of the mechanism of action of Zfp would also require knowledge about the tertiary structure of the proteins and their DNA- protein complexes.

Further still unsolved questions are the roles of unique and finger regions in the activation of transcription by TFIIIA. It is alligned on the internal control region (ICR) of 5S RNA gene with C-terminus towards the 5'end of the gene(8). By analysis of truncated TFIIIA molecules it could be shown that only 19 amino acids outside the DNA binding finger domain at the C-terminus are critical for maximal transcription (46). Deletion of the first 17 amino acids from the N-terminus that include 12 unique and 5 amino acids from the first finger reduces the overall binding affinity of TFIIIA over the entire length of ICR. Observation of finger sequences from different Zfp reveal that inspite of the presence of highly conserved residues,each finger contains many variable amino acids. Whether they are involved in discriminating the specificity of binding is not known. Finger swapping experiments or the use of many single residue mutants will be necessary to answer this question.

Mkr3,4 and 5 add to the growing list of a large multigene family of Zfp in different organisms,indicating that such proteins may be widely used in the transcriptional or post transcriptional regulation. In table 1 we have compared the protein sequences of published mouse finger genes to show that mkr3,4 and 5 represent new genes. Their transcription pattern is also different from all of them. Interestingly,all 3 clones encode large number of fingers and the sequences of the N-terminal regions of these proteins remain to be determined by isolating larger cDNA clones. However, the C-terminal sequences are complete for both mkr3 and mkr5 but not for mkr4. All these and other published Zfp contain typically small non finger unique regions. In the primary structure of mkr4, two finger domains are separated by a unique region, similar to that of Hunchback (6), a Drosophila segmentation gene of the gap class. Both mkr3 and 5 c-DNAs have long untranslated 3'ends, as in mkr2 described

previously (35) and contain the sequence motif ATTTA potentially involved in the degradation of mRNA(43). But without nuclear run on experiments, it is not possible to distinguish between a transcriptional or post transcriptional mechanism of mkr mRNA metabolism. Two potential nuclear localization signals were detected in mkr4 and mkr5 suggesting that they may be nuclear proteins analogous to the Drosophila Krüppel family of nuclear finger gene products(47).Wheather or not they bind to Zn and DNA remains to be shown.

Northern blotting data revealed that all clones were expressed as multiple transcripts in developing mouse embryos and adult mouse tissues. It is highly unlikely that these descrete multiple messages are due to the result of unspecific hybridization. Several facts argue against this reasoning. Firstly,all hybridizations were done with probes that detect unique bands in southern blotting experiments using mouse genomic DNA. Secondly,all messages detected for mkr3,4, 5 and as well as for mkr1 and 2(11,34) are of different sizes. There are obviously no transcripts of similar size among all these clones. Examples of such transcription pattern are known for Drosophila Antennapedia gene (48) or for mouse int-2 gene(49). These genes produce multiple messages due to different promotors and termination processing signals but share the same coding region. However,the possibility of differential splicing for the generation of different proteins remain open for mkr3,4 and 5. This question can be addressed by cloning and analysing the corresponding c-DNAs.

Judging from the pattern of expression, mkr3 and 4 are ubiquitious proteins. If they turn out to be transcriptional regulatory factors,then they would be similar to SP1 (13) and TFIIIA(7,8). In contrast mkr5 gene expression was predominantly restricted to ovary. Other examples of tissue specific Zfp are mkr2 (35) whose expression is limited to the neuronal cells and TDF,the human testis determining factor(37). These proteins may represent tissue specific regulatory factors. We are now in the process of characterizing the 5′ regions of these genes to determine whe ther they contain any tissue specific regulatory signals. Ultimately,it would be necessary to isolate the target genes being controlled by mkr family of genes to shed light on their functional role in murine embryogenesis and development.

ACKNOWLEDGEMENTS

We would like to thank Barbara Meyer for help with the screening of the cDNA library, Bernd Föhring for computer work and Thomas Schowalter for technical assistance. H.R. was a recipient of Böhringer Ingelheim postdoctoral fellowship.This project was supported by a

financial grant from the Deutsche Forschungsgemeinschaft (DFG, Gr 650/3-1).

*To whom correspondence should be addressed.

+Present address: Schering AG, Institute für Biochemie, Postfach 650311, D-1 Berlin, 65, GDR

REFERENCES

1. Hartshorne,T.,Blumberg,H., and Young,E. (1986). Nature,320,283-287.
2. Blumberg,H.,Eisen,A.,Sledziewski,A.,Bader,D.,and Young,E(1987). Nature, 328,443-445.
3. Stillman,D.,Bankier,A.,Sedon,A.,Groenhout,E.,and Nasmyth,K.(1988).EMBO J.,7, 485-494.
4. Vincent,A., Colot,H., and Rosbash,M. (1985). J. Mol. Biol.,185,146-166.
5. Rosenberg,U.B., Schröder,C., Preiss,A., Kienlin,A., Cote,S., Riede,E.,and Jäckle,H.(1986). Nature, 319,336-339.
6. Tautz,D., Lehman,R.,Schnürch H.,Schuh,R.,Seifert,E., Kienlin.A., Jones.K., and Jäckle,H. (1987). Nature, 327,383-389.
7. Ginsberg,A., King,B., and Roeder,R. (1984). Cell, 39,479-489.
8. Miller,J., Mclachlan,A.D., and Klug,A. (1985). EMBO J. 4,1609-1614.
9. Ruiz i Altaba,A.,Perry-O'Keefe,H.,and Melton,D.(1987). EMBO J. 6,3065-3070.
10. Köster,M.,Pieler,T.,Pötting A.,and Knöchel,W.(1988).EMBO J.,7(6),1735-1741.
11. Chowdhury,K., Deutsch,U., and Gruss,P. (1986). Cell 48,771-778.
12. Chavrier,P., Lemaire,P., Revelant,O., Bravo,R., and Charnay,P.(1988).Mol. Cell.Biol. 8(3),1319-1326.
13. Kadonaga,J., Carner,K., Masiaz,F.,and Tjian.R.(1987). Cell,51,1079-1090.
14. Pannuti,A.,Lanfrancone,L.,Pascucci,A.,Pier-Giuseppe,P.,LaMantia,G. and Lania,L.(1988). Nucl.Acids.Res. 16(10),4227-4237.
15. Akam,M. (1987). Development, 101,1-22.
16. Pabo,C., and Sauer,R.(1984). Ann.Rev.Biochem., 53, 293-321.
17. Gehring,W., and Hiromi,Y.(1986). Ann.Rev.Genet. 20,147-173.
18. Colberg-Poley,A., Voss,S., and Gruss,P.(1987). in Oxford surveys on eukaryotic genes, 4,92-115.
19. Desplan,C., Theis,J., and O'Farrel,P.(1985). Nature, 318, 630-635.
20. Hoey,T., and Levine,M.(1988). Nature, 332, 858-861.
21. Brown,R.S., Sander,C., and Argos,P.(1985) . FEBS Lett. 186, 271-274.
22. Berg,J. (1986) . Science, 232, 485-486.
23. Rhodes,D., and Klug,A.(1988) in Nucleic acid and molecular biology. Eckstein,F.,and Lilley,D. (Eds),Springer Verlag,Berlin,Heidelberg,1988.
24. Evans,R., and Hollenberg,S. (1988). Cell 52, 1-3.
25. Gibson,T., Postma,J.,Brown,R., and Argos P.(1988). Protein engeneering, in press.

26. Hanas,T.S., Hazuda,D.J., Bogenhagen,D.F., Wu,F.,and Wu,C.W. (1983).J.Biol. Chem. 258,14120-14125.
27. Frankel,A.,Berg,J.,and Pabo,C.(1987). Proc.Natl.Acad.Sci. USA. 84,4841-4845.
28. Diakun,G.,Fairall,L.,and Klug,A.(1986). Nature, 324, 698-699.
29. Berg,J. (1988). Proc.Natl.Acad.Sci. USA. 85, 99-102.
30. Frankel,A., and Pabo,C. (1988). Cell, 53, 675.
31. Frankel,A., Bredt,D., and Pabo,C. (1988). Science, 240, 70-73.
32. Nüsslein-Vollhard,C. and Wieshaus,E. (1980). Nature, 287, 795-801.
33. Ollo,R.,and Maniatis,T.(1987). Proc.Natl.Acad.Sci.USA.84,5700-5704.
34. Redemann,N., Gaul,U., and Jäckle,H. (1988). Nature, 332, 90-92.
35. Chowdhury,K., Dressler,G., Breier,G., Deutsch,U., and Gruss,P. (1988). EMBO J. 7(5),1345-1353.
36. Chavrier,P., Zerial,M., Lemaire,P., Almendral.,J.,Bravo,R., and Charnay,P., (1988). EMBO J. 7(1),29-35.
37. Page,D., Mosher,R., Simpson,E.,Fisher,E.,Mardon,G.,Pollack,J.,McGillivray, B., De la Chapelle,A., and Brown,L.(1987). Cell, 51,1091-1104.
38. Fahrner,K., Hogan,B., and Flavell,R.(1987). EMBO J., 6,1265-1271.
39. Maniatis,T., Fritsch., and Sambrook,J.,(1982). Molecular clonong: a laboratory manual, Cold Spring Harbor Laboratory press, CSH, NY.
40. Ozaki,L.,and Cseko,Y.(1984). Genes and antigens of parasites,a laboratory manual, Department of Biochemistry and Molecular Biology,Institut Oswaldo Cruz, Brazil.
41. Maxam,A., and Gilbert,W. (1977). Proc.Natl.Acad.Sci. USA. 74, 560-564.
42. Chirgwin,J.M., Przybyla,A.E., Macdonald,R.J., and Rutter,W.J.(1979). Biochemistry, 18, 5294-5299.
43. Shaw,G., and Kamen,B. (1986). Cell, 46, 659-667.
44. Kalderon,D., Roberts,B., Richardson,W., and Smith,A.(1984).Cell,39,499-509
45. Takahashi,M.,Inaguma,Y.,Hiai,H.,and Hirose,H.(1988) Mol.Cell.Biol.8(4), 1853-1856.
46. Vrana,K.,Churchill.M.,Tullius,T.,and Brown,D.(1988). Mol.Cell.Biol. 8(4), 1684-1696.
47. Schuh,R., Aicher,W., Gaul,U., Cote,S.,Preiss,A.,Maier,D.,Seifert,E.,Nauber, U., Schhröder,C., Kemler,R., and Jäckle,H.(1986). Cell 47, 1025-1032.
48. Schnewly,S., Kuroiwa,A., Baumgartner,P.,and Gehring,W.(1986). EMBO J. 5(4),733-739.
49. Mansour,S., and Martin,G.(1988). EMBO J. 7(7), 2035-2041.