The KUP gene, located on human chromosome 14, encodes a protein with two distant zinc fingers

Pierre Chardin*, Geneviève Courtois[1], Marie-Geneviève Mattei[2] and Sylvie Gisselbrecht[1]
Institut de Pharmacologie du CNRS, Route des Lucioles, 06560 Valbonne, [1]Institut Cochin de Génétique Moléculaire, 27 rue du Faubourg Saint-Jacques, 75014 Paris and [2]INSERM U.242, Hopital d'Enfants de La Timone, 13385 Marseille Cedex 5, France

Received January 16, 1991; Revised and Accepted March 6, 1991 EMBL accession no. X16576

ABSTRACT

We have isolated a human cDNA (kup), encoding a new protein with two distantly spaced zinc fingers of the C_2H_2 type. This gene is highly conserved in mammals and is expressed mainly in hematopoietic cells and testis. Its expression was not higher in the various transformed cells tested than in the normal corresponding tissues. The kup gene is located in region q23-q24 of the long arm of human chromosome 14. The kup protein is 433 a.a. long, has a M.W. close to 50 kD and binds to DNA. Although the structure of the kup protein is unusual, the isolated fingers resemble closely those of the Krüppel family, suggesting that this protein is also a transcription factor. The precise function and DNA motif recognized by the kup protein remain to be determined.

INTRODUCTION

Zinc fingers, found in many nucleic acid binding proteins, were first described in TFIIIA, a transcription factor required for the initiation of transcription of *Xenopus laevis* 5S RNA genes and binding both to genomic DNA and to 5S RNA itself (1). A large number of other proteins containing zinc fingers of the Cys-Cys/His-His type (C_2H_2 finger proteins) has now been identified, including the yeast regulatory proteins ADR1 (2) and MIG1 (3), *Aspergillus nidulans* brlA (4), *Drosophila* Krüppel (5), serendipity (6), hunchback (7), snail (8), engrailed (9), odd-skipped (10), glass (11) and terminus (12), *Xenopus laevis* Xfin (13) and other finger proteins (14), mouse mkr (15), Krox (16, 17) and mfg (18), human EGR (19), Sp1 (20), HF10 and HF12 (21), GLI (22), PLK (23), ZFX and ZFY (24, 25) and several other finger proteins (26, 27). All of these proteins possess imperfect tandem repeats of the consensus sequence: $Cx_2Cx_3Fx_5Lx_2Hx_3H$. This motif is able to fold in a 'finger like' domain where the two cysteines on one side and the two histidines on the other coordinate the central zinc ion. In most of these proteins, such as TFIIIA, Krüppel, ADR1, SP1, the zinc finger motifs are separated by a short conserved motif: the H/C link, connecting the final histidine of one finger to the first cysteine of the next finger. The sequence of this H/C link is usually close to the consensus: HTGEKP(Y/F)xC. In contrast, for some other proteins such as hunchback or serendipity, the fingers are separated by a sequence that does not fit this consensus, but has a similar length.

Furthermore it has been suggested that several genes encoding zinc fingers containing proteins might be frequently implicated in transformation, including: myeloid leukemia in mouse (28), cell proliferation (29), melanoma (30) and HTLV transformation (31) and in excision repair of DNA (32). It has also been shown that the gene deleted in Wilms tumors encodes a protein with four zinc fingers (33, 34) that seems to bind to the same DNA consensus sequence as EGR1 (35), the human homolog of mouse Krox24.

We report here the isolation and characterization of a new cDNA encoding a protein with two zinc fingers closely related to the ones of the Krüppel family, but where the H/C link is replaced by a much longer sequence.

MATERIALS AND METHODS

Plasmids and Libraries

The YRPSCD25a plasmid is a gift from J.Camonis and M.Jacquet (36), the human pheochromocytoma library in λGT10 is a gift from A.Lamouroux and J.Mallet (37).

Molecular cloning and sequencing of the kup cDNA

In search of a human homolog of the yeast *ras* exchange factors SCD25 and CDC25, the XbaI-EcoRI 1503 b.p. fragment, encoding the C-terminal part of the SCD25 protein was cut from the YRPSCD25a plasmid (36), and subcloned in pGEM3, XbaI-EcoRI. After plasmid amplification this fragment was purified by electro-elution, ^{32}P labelled with a multiprime labelling kit (Amersham, U.K.) and used to screen a human pheochromocytoma cDNA library in λGT10 (37) at very low stringency (5× SSPE, 5× Denhart, 0.1% SDS, at 58°C for pre-hybridization and hybridization, 1× SSC, 0.1% SDS at 58°C, 15 min for the final wash, methods described in detail in ref. 38). Out of about 500,000 plaques, 14 positive clones were

* To whom correspondence should be addressed

isolated belonging to four different classes, the largest cDNA of each class was sequenced, none of these sequences was found to encode a protein related to SCD25 or CDC25, however one of these clones encoded a protein with some homology to several zinc finger containing proteins of the Cys-Cys/His-His type and at least two zinc-fingers were identified. The complete sequence of this 1951 nucleotides cDNA was determined (see Figure 1).

Northern Blot Analysis of kup expression

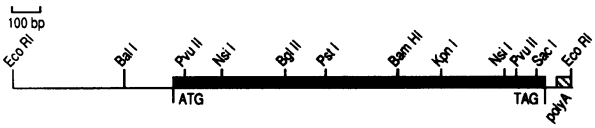
RNAs from various mouse, rat and human organs and cell lines were extracted by the guanidinium thiocyanate method (38) and analysed by Northern blotting, as described previously (47).

Chromosomal localization of kup in human

In situ hybridization was carried out on chromosome preparations obtained from phytohemagglutinin-stimulated human lymphocytes cultured for 72 hours. 5-bromodeoxyuridine was added for the final seven hours of culture (60 μg/ml of medium), to ensure a post-hybridization chromosomal banding of good quality. A 1148 b.p. PvuII fragment from the coding region of the kup cDNA (see Fig.1a) devoid of repetitive sequences, was tritium labelled by nick-translation to a specific activity of 7. 10^7 d.p.m./μg. The radiolabelled probe was hybridized to metaphase chromosome spreads at a final concentration of 25 ng per ml of hybridization solution as previously described (39). After coating with nuclear track emulsion (Kodak NTB_2), the slides were exposed for 14 days at +4°C and developed. To avoid any slipping of silver grains during the banding procedure, chromosome spreads were first stained with buffered Giemsa solution and metaphases photographed. R-banding was then performed by the fluorochrome-photolysis-Giemsa (F.P.G.) method and metaphases photographed again before analysis.

E.coli Expression of the kup protein and DNA binding ability

A BclI site, immediatly after the initiating ATG of kup was introduced by site directed mutagenesis, the BclI-HindIII fragment including the complete coding sequence of kup was then subcloned in pT7 BclI, a new expression vector (Chardin et al., in preparation), derived from the pET3c vector, a T7 RNA polymerase based expression vector (40). This pT7BclI-KUP vector was transfered into the *E.coli* pLysS strain, and T7 RNA polymerase production induced by 0.2 mM IPTG. For the ^{35}S methionine labelling experiment (see Fig.5), transcription from bacterial promoters was blocked by the addition of Rifampicin, an antibiotic that blocks bacterial RNA polymerases but not T7 RNA polymerase. An early log phase culture was induced with 0.2 mM IPTG for 20 min., cells were centrifuged, washed and resuspended in M9 minimal media containing: 1 mM $MgSO_4$, 1 mM $CaCl_2$, 0.1 mM Thiamine, 0.2% Glucose, 0.1 mM of all amino-acids except Met and Cys, 200 μg/ml Rifampicin and incubated for 20 min., ^{35}S labelled Methonine (37 TBq/mmol) was added at a final concentration of 10 μM and incubated for an additional 20 min., cells were then collected by centrifugation and directly lysed by boiling in Laemmli sample buffer for immediate analysis by 12% PAGE, or gently lysed, as described (41) except that lysis was performed by the freeze/thaw method, in the binding buffer, to prepare the soluble fraction for further analysis. DNA binding ability was demonstrated essentially as described (42) The soluble extract from ^{35}S methionine labelled *E.coli* was incubated with double-strand DNA-cellulose (Pharmacia), in binding buffer (20 mM Hepes, pH 7.8/0.05 M NaCl/2 mM $MgCl_2$/10 μM $ZnCl_2$/0.2 mM Phenyl-methyl-




```
                                                                    atcagttgttaccagttgcggaggcagcgcatagcacaacgtagtacctgga -505
acctgacggcacctagcatgtttcaacagttagcgcccgcatcaccccaagggattaaatttggcattcttt -433
ctcttgagactcaagtcctggcagagctgttgatcctctcacaagattctgtcatagaagacaccaggcctg -361
aagagcgtgatgtagctggagctagctacacctcttgatagctaagactccaaggctccagacaccttggtg -289
attgcactctccaccatgccaggaccgcagtgccagtttaccccactcacctgtagagaatgtatccggcgg -217
aggtcattatatatacttggcttgattcaggttatggagaagtcaggcatggccactgtgtgggtttgatgg -145
tatttctaggtgagatggagtttcactcttgttgcccagactggagtgcaatgacacggtctcggctcacca  -73
caacctctgcctcctgggttcaagcgattctcccgcctcagcctcccgagtagccgggattatagaaccaca   -1
 M  D  T  A  S  H  S  L  V  L  L  Q  Q  L  N  M  Q  R    18
ATG GAC ACT GCC AGC CAT AGC CTT GTT CTT CTC CAG CAG CTG AAC ATG CAG CGA     54
 E  F  G  F  L  C  D  C  T  V  A  I  G  D  V  Y  F  K    36
GAA TTT GGT TTT CTG TGT GAT TGC ACA GTT GCA ATT GGA GAT GTT TAC TTC AAA    108
 A  H  R  A  V  L  A  A  F  S  N  Y  F  K  M  I  F  I    54
GCC CAC AGA GCA GTG CTT GCT GCT TTT TCT AAC TAT TTC AAG ATG ATA TTT ATT    162
 H  Q  T  S  E  C  I  K  I  Q  P  T  D  I  Q  P  D  I    72
CAC CAA ACA AGT GAA TGC ATA AAA ATA CAA CCA ACT GAC ATC CAA CCT GAC ATA    216
 F  S  Y  L  L  H  I  M  Y  T  G  K  G  P  K  Q  I  V    90
TTC AGC TAT TTG TTG CAC ATT ATG TAC ACG GGG AAA GGG CCA AAA CAG ATT GTG    270
 D  H  S  R  L  E  E  G  I  R  F  L  H  A  D  Y  L  S   108
GAT CAT AGT CGT TTG GAG GAA GGG ATT CGA TTT CTT CAC GCC GAC TAC CTT TCT    324
 H  I  A  T  E  M  N  Q  V  F  S  P  E  T  V  Q  S  S   126
CAC ATT GCA ACT GAA ATG AAT CAA GTG TTC TCA CCA GAG ACT GTG CAG TCC TCA    378
 N  L  Y  G  I  Q  I  S  T  T  Q  K  T  V  V  K  Q  G   144
AAT TTA TAT GGC ATT CAG ATC TCA ACA ACC CAA AAA ACA GTT GTC AAA CAA GGA    432
 L  E  V  K  E  A  P  S  S  N  S  G  N  R  A  A  V  Q   162
CTG GAG GTC AAA GAA GCT CCT TCC AGT AAC AGT GGA AAC AGA GCT GCT GTC CAG    486
 G  D  L  P  Q  L  S  L  A  I  G  L  D  D  G  T  A  D   180
GGT GAC CTC CCC CAG TTG TCT CTT GCT ATT GGT CTG GAT GAT GGC ACT GCA GAC    540
 Q  Q  R  A  C  P  A  T  Q  A  L  E  E  H  Q  K  P  P   198
CAG CAG AGG GCC TGT CCT GCC ACC CAG GCC CTG GAG GAG CAC CAG AAG CCC CCA    594
 V  S  I  K  Q  E  R  C  D  P  E  S  V  I  S  Q  S  H   216
GTT TCC ATC AAG CAG GAG AGA TGT GAC CCA GAA TCT GTG ATC TCC CAG AGC CAC    648
 P  S  P  S  S  E  V  T  G  P  T  F  T  E  N  S  V  K   234
CCC TCA CCC TCA TCA GAG GTG ACA GGC CCC ACT TTT ACT GAA AAC AGT GTC AAA    702
 I  H  L  C  H  Y  C  G  E  R  F  D  S  R  S  N  L  R   252
ATA CAC TTA TGC CAT TAC TGT GGG GAA CGT TTT GAT TCC CGT AGT AAC CTA AGG    756
 Q  H  L  H  T  H  V  S  G  S  L  P  F  G  V  P  A  S   270
CAA CAT CTC CAT ACA CAT GTG TCT GGA TCC CTG CCA TTC GGT GTC CCT GCT TCC    810
 I  L  E  S  N  D  L  G  E  V  H  P  L  N  E  N  S  E   288
ATT CTG GAA AGT AAT GAC CTT GGT GAA GTG CAT CCC CTT AAT GAA AAC AGC GAG    864
 A  L  E  C  R  R  L  S  S  F  I  V  K  E  N  E  Q  Q   306
GCC CTT GAA TGC CGC AGG CTC AGC TCC TTC ATT GTT AAG GAG AAT GAA CAG CAG    918
 P  D  H  T  N  R  G  T  T  E  P  L  Q  I  S  Q  V  S   324
CCA GAC CAC ACC AAC CGG GGT ACC ACA GAG CCT TTG CAG ATC AGT CAA GTA TCT    972
 L  I  S  K  D  T  E  P  V  E  L  N  C  N  F  S  F  S   342
TTG ATC TCC AAA GAC ACA GAG CCA GTA GAA TTA AAC TGT AAT TTT TCT TTT TCA   1026
 R  K  R  K  M  S  C  T  I  C  G  H  K  F  P  R  K  S   360
AGG AAA AGA AAA ATG AGC TGT ACC ATC TGT GGT CAT AAA TTC CCT CGA AAG AGC   1080
 Q  L  L  E  H  M  Y  T  H  K  G  K  S  Y  R  Y  N  R   378
CAA TTG TTG GAA CAC ATG TAT ACA CAC AAA GGT AAA TCT TAC AGA TAT AAC CGA   1134
 C  Q  R  F  G  N  A  L  A  Q  R  F  Q  P  Y  C  D  S   396
TGC CAA AGG TTT GGT AAT GCA TTG GCC CAG AGA TTT CAG CCA TAC TGT GAC AGC   1188
 W  S  D  V  S  L  K  S  S  R  L  S  Q  E  H  L  D  L   414
TGG TCT GAT GTC TCC CTG AAA AGT TCT CGC TTG TCA CAA GAA CAC TTA GAC TTG   1242
 P  C  A  L  E  S  E  L  T  Q  E  N  V  D  T  I  L  V   432
CCT TGT GCC TTA GAG TCA GAG CTC ACA CAA GAA AAT GTG GAT ACT ATC CTA GTT   1296
 E  *                                                    434
GAG TAG  cagcttctctttcagaatttttataccaattctgaaaaataaaatttgcgtggcatttttaat 1365

tcataaattattttcctcctcaaaaaaaaa                                        1395
```

Figure 1: Structure and complete sequence of the kup cDNA. **Upper line:** structure of the human kup cDNA, the heavy line indicates the coding region, the dotted line is the polyA tail, major restriction sites are also indicated, the EcoRI sites at each end are derived from the linkers introduced during the construction of the λGT10 library. **Lower part:** Complete sequence of the kup cDNA and translation of the major open reading frame, using the one letter code for amino acids.

sulfonyl fluoride), washed three times in the same buffer, eluted with a second buffer (20 mM Hepes, pH 8.4/1 M NaCl/2 mM $MgCl_2$/10 μM $ZnCl_2$/0.2 mM Phenyl-methylsulfonyl fluoride) and analysed by 12% PAGE.

RESULTS

Molecular cloning of the kup cDNA

In search of human homologs of the yeast *ras* exchange factors SCD25 and CDC25, a cDNA was isolated from a human pheochromocytoma library (37), by very low stringency hybridization with the yeast SCD25 probe. Sequencing revealed no homology with SCD25 or CDC25, but two zinc fingers were present in the encoded protein that we named KUP for KrUppel homologous Protein. Thus we serendipitously isolated the cDNA for a new protein with two zinc fingers. The fact that this cDNA was isolated fortuitously suggests that a large number of zinc finger containing proteins are expressed in human cells. A similar observation has been made for the large family of small G-proteins where several members were found fortuitously (43).

Structure and sequence of the kup cDNA

The structure of the kup cDNA is shown on Figure 1a. The kup cDNA isolated here is 1942 b.p. long (not including the polyA), a size compatible with the 2 k.b. message observed in several human cells (see next paragraph) and suggesting that this cDNA is nearly full length, as most other cDNAs that we isolated from this very good library (37). Complete sequencing of this cDNA shows only one long open reading frame of 1299 nucleotides that starts at position 557 and ends with a TAG stop codon at position 1856 (see Fig. 1b). The best calculated score for a ribosome binding site (44) is found at position 550−559, exactly in place, upstream of the first ATG of this long O.R.F. There is thus a large 5′ non-coding region of at least 556 nucleotides in the kup mRNA. In contrast, the 3′ non-coding region is unusually short (84 nucleotides), with a CATAAA polyA addition signal starting only 64 b.p. after the stop codon. The first A of the polyA tail is found 14 b.p. after this CATAAA sequence that does not fit perfectly the most frequent AATAAA sequence, however this kind of variation has already been observed for several other genes.

Predicted structure and properties of the kup protein

The kup protein is 433 amino acids long with an expected molecular weight of 48.7 kD, and a calculated pI of 6.8.

The two zinc fingers are found in the C-terminal half of the protein, the first zinc finger starts with Cys 238 and ends with His 258, the second one starts at Cys 349 and ends at His 369, the H/C link would therefore be 91 a.a. long (see Fig. 1a and Fig. 2). The N-terminal part of the protein is unusually rich in Ser, Thr, Pro, Gln and His, a region very rich in amino acids with hydroxyl groups (Ser and Thr) is found just upstream of the first zinc finger. However no long stretches of only one repeated amino acid are observed, such as the Ala repeats found in some repressors (45). There are no long hydrophobic regions.

```
KUP  235      IHLCHYCGERFDSRSNLRQHLHTHVSGSLPF
KUP  346      KMSCTICGHKFPRKSQLLEHMYTHKGKSYRY

Consensus     x^Yx^FCxxCxxxFxxxxxLxxHxxxHxxxx
```

Figure 2: Comparison of the two kup zinc fingers with zinc fingers consensus sequence of other C_2H_2 finger proteins. Upper lines: amino-acid sequences of the two kup finger motifs. **Lower line:** the consensus sequence of most zinc fingers of the C_2H_2 type.

Expression of kup

The human kup probe hybridized with rat and mouse genomic DNA efficiently, even at high stringency, indicating that the kup gene is highly conserved in these species and very likely in most mammals. Therefore, we studied kup expression in different tissues from various mammals, including mouse, rat and man.

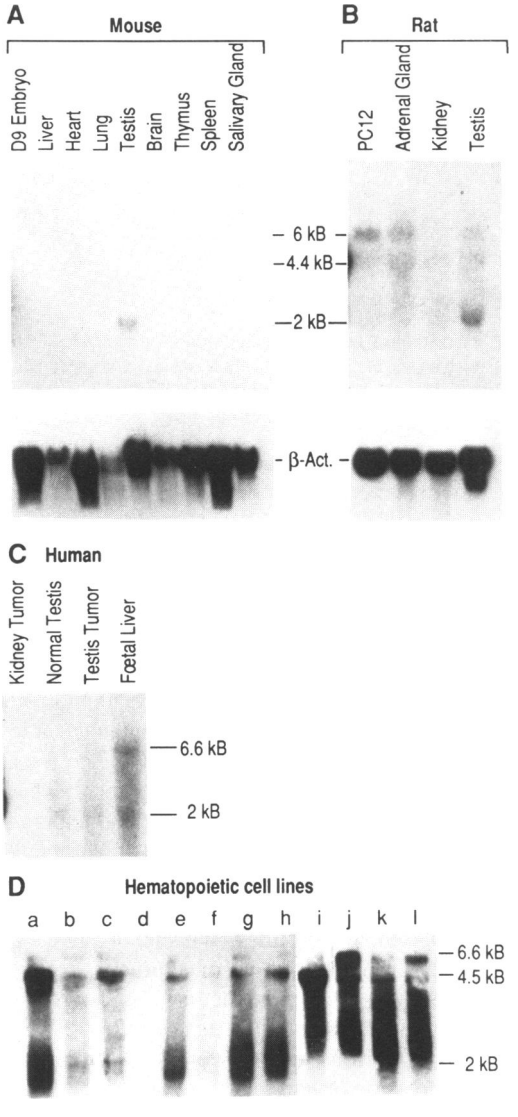


Figure 3: Expression of kup. A: total cytoplasmic RNAs (15 μg) from the indicated mouse tissues were probed with the complete EcoRI insert of the human kup cDNA. **B:** total cytoplasmic RNAs (15 μg) from the indicated rat tissues were probed with the 1.2 k.b. PvuII fragment of the kup cDNA (see Fig.1). **C:** total cytoplasmic RNAs (15 μg) from the indicated human tissues were probed with the same PvuII fragment. **D:** polyA+ RNA (2 μg), from hematopoietic cell lines were also probed with the same PvuII fragment, a: BJAB (Burkitt lymphoma, EBV negative), b: B-cell line, EBV positive, c: K562 erythroleukemia, d: HEL erythroleukemia, e: TPHA cell line, f: KGI, g: B-cell line, EBV positive, h: HL-60 myeloid leukemia, i: U-937 myeloid leukemia, j: Raji (Burkitt lymphoma, EBV positive), k: EL-B (Burkitt lymphoma, EBV positive), l: CCRF T-cell lymphoma. Blots were re-probed with β-Actin to provide an internal standard to quantitate mRNA.

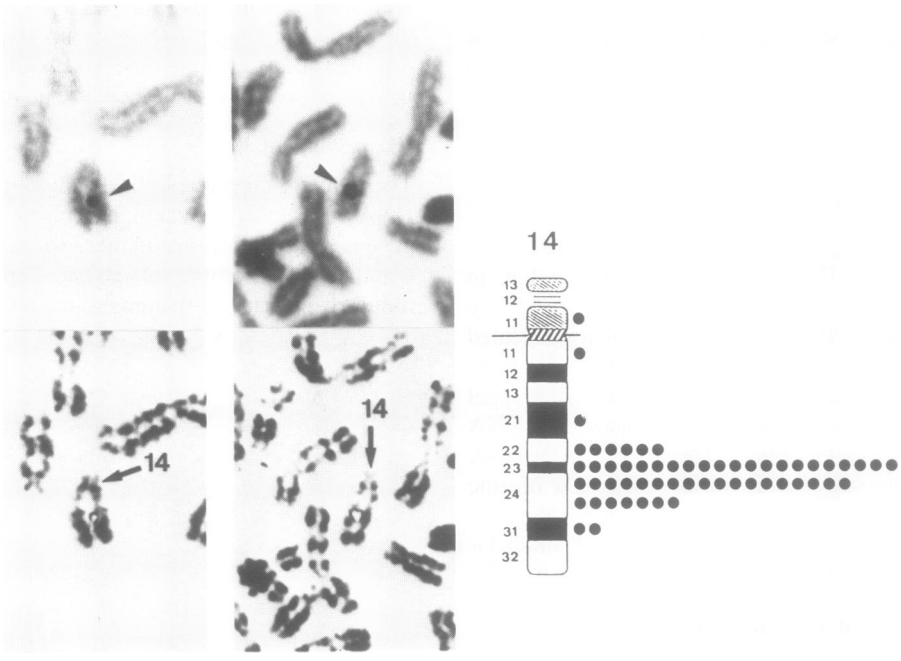


Figure 4: Localization of kup on human chromosomes. **Left:** Two partial human metaphases showing the specific site of hybridization on chromosome 14. Top: arrowheads indicate silver grains on Giemsa stained chromosomes, after autoradiography. Bottom: chromosomes with silver grains were subsequently identified by R-banding (F.P.G. technique). **Right:** Idiogram of the human G-banded chromosome 14 illustrating the distribution of labelled sites for the kup probe.

In mouse, expression was found mainly in testis where an RNA of 2 k.b. was observed, this 2 k.b. RNA was also detectable at very low levels in thymus and spleen (Fig. 3A). In rat, expression was also found in testis where three RNAs of 2, 4.4 and 6.6 k.b. were present, a low level of expression was also detectable in kidney, the adrenal gland and PC12 cells (Fig. 3B). The PC12 pheochromocytoma does not show a higher expression than the normal adrenal gland, indicating that overexpression of kup in pheochromocytomas is not a general feature. In human, two kup transcripts of 6.6 and 2 k.b. were detected, although at low levels, in kidney, testis and foetal liver (Fig. 3C). Since foetal liver is a hematopoietic organ at this stage of development, we also tested several human hematopoietic cell lines for kup expression (Fig. 3D). All cell lines expressed kup, although at variable levels and with RNAs of different size. Transcripts of 2, 2.5, 4.5 and 6 k.b. were observed, the higher molecular weight transcript being predominant in many cell lines. This expression of several different transcripts is suggestive of alternative splicing, as is frequently observed for messengers encoding zinc finger containing proteins, in hematopoietic cells (25, 46). Sixty eight human tumors of various origins were also screened to search for an over-expression of kup related to transformation, without finding a high level of expression in any of these tumors (data not shown).

Chromosomal localization of kup in human

Out of the 100 metaphase cells examined after *in-situ* hybridization, 252 silver grains were associated with chromosomes and 47 of these (18.6%)were located on chromosome 14, the distribution of the grains on this chromosome was not random: 52/57 (91.2%) of them mapped to the [q22-q24] region of chromosome 14 long arm with a maximum in the q23 and q24 bands (Figure 4). These data unequivocally map the kup gene to the 14q23 −14q24 bands of the human genome.

Expression of the kup protein in Bacteria and DNA binding capacity

The kup open reading frame was inserted immediately downstream of a T7 promoter and this vector was introduced in the E.Coli strain pLysS, harboring the T7 RNA polymerase gene, under the control of a lac promoter. This T7 RNA polymerase was induced by adding IPTG, RNA synthesis by bacterial RNA polymerases was subsequently blocked by Rifampicin, and synthesized proteins, mainly those derived from RNAs under the control of T7 promoters, were then labelled with ^{35}S methionine. Figure 5A shows that the major protein synthesized in these conditions is a 52 kD protein, that is neither found in the uninduced control nor in the induced cells harbouring the rho cDNA (encoding a 22 kD protein) in the same vector. This 52 kD protein is therefore the product of the kup cDNA and its observed molecular weight is close to the calculated one of 48.7 kD. Figure 5B shows that, in contrast to most other proteins of the soluble E.Coli fraction, the kup protein remained bound to a DNA-cellulose column and could be eluted at high salt, suggesting a DNA binding capacity.

DISCUSSION

Structure and expression of the kup cDNA

We have isolated a human cDNA encoding a protein with two zinc fingers in its C-terminal part, that we named KUP for KrUppel related Protein. Genomic sequences hybridizing with the human kup probe were found in rat and mouse suggesting a very high phylogenic conservation, at least in mammals, and thus an essential function. The expression of this cDNA was usually very low, the highest levels of expression were found in testis, foetal liver and hematopoietic cells, where at least three RNAs of 2, 4.5 and 6 k.b. were observed. The 2 k.b. RNA is

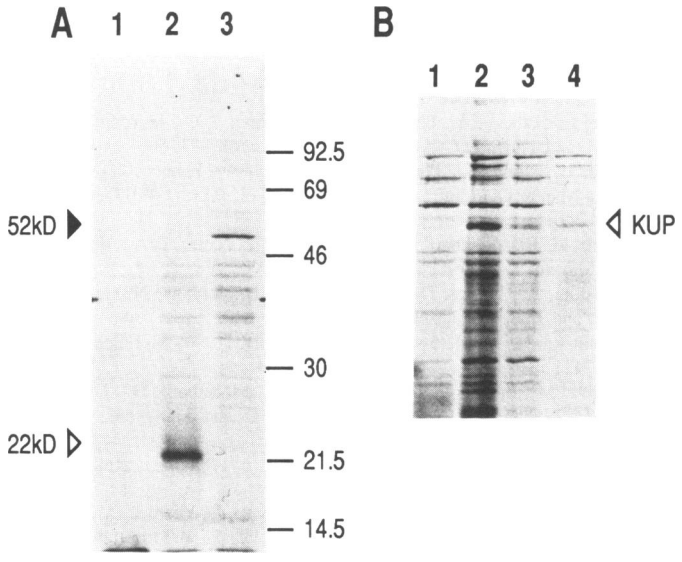


Figure 5: Expression of the kup protein in *E.Coli* and DNA binding capacity. **A.** Proteins synthesized from T7 promoters were preferentially labelled by ^{35}S methionine, boiled in sample buffer, separated by 12% PAGE and autoradiographed, 1: uninduced, 2: rho protein, induced, 3: kup protein, induced, the sizes of molecular weight markers on the right are in kD. **B.** Total ^{35}S labelled soluble proteins (Lane 2) including kup (major band at 52 kD) were incubated with DNA cellulose as described in the methods section, lane 3 is the unbound fraction, in lane 4 are the proteins that remained bound to DNA-cellulose and were only eluted at high salt, lane 1 represents a control of total soluble proteins from uninduced bacteria, all samples were separated by 12% SDS-PAGE and autoradiographed, the molecular weights are aligned with those of the gel in A.

very likely corresponding to the 1950 b.p. cDNA that we isolated, suggesting also that this cDNA is almost full-length. The structures of the larger transcripts are not known, it is likely that they are generated by alternative splicing, as has already been observed for several other finger containing proteins (25, 46), or this might result from the use of a downstream poly-Adenylation signal. It would be interesting to clone the higher molecular weight transcripts from a hematopoietic cell line library to understand the molecular basis of this variability. We also looked for the expression of kup in various transformed cells, without finding a high level of expression in any solid tumor. A high level of expression was found in most hematopoietic cell lines tested, in lymphoid T and B cells, in myeloid as well as erythroid cell lines. The expression of kup in the rat PC12 pheochromocytoma was not higher than in the normal adrenal gland, indicating that a high expression of kup is not a general feature of pheochromocytomas. Taken together, these data indicate that kup is not preferentially expressed in transformed cells, but do not rule out the possibility that an over-expression of kup is associated with transformation in a specific tissue, not tested in this study.

Chromosomal localization of the human kup gene

We unambiguously localized the kup gene on the long arm of human chromosome 14, in the [q23-q24] region. The kup probe will therefore provide a useful genetic marker for linkage analysis and detailed mapping, in this region of chromosome 14. The *fos* oncogene at [q24.3] and the gene for β-Spectrin [q23-q24] have also been mapped to this region. At present we have not identified any human restriction fragment length polymorphism associated with the kup probe.

Structure and properties of the kup protein

The 1950 b.p. cDNA has only one long open reading frame that starts with a typical ribosome binding site according to Kozak (44) and ends close to the 3'end. Furthermore, when this open reading frame was inserted in a bacterial expression vector, a protein with the expected molecular weight was observed. It is therefore very likely that the protein encoded by this cDNA is indeed the one predicted from sequence analysis. The fact that we have found this cDNA fortuitously suggests that many zinc finger containing proteins are expressed in human cells. It has been claimed that at least a hundred of such proteins might be encoded in the human genome (27). The kup protein is 433 amino acids long with an expected molecular weight of 48.7 kD, and an electrophoretic mobility of 52 kD. When screening the NBRF databank (Version 26 of September 1990), the highest homology score was found with a Xenopus protein containing several zinc fingers (14), high scores were also obtained for the finger region of Krüppel and many other proteins with fingers of the C_2/H_2 type, with 30%-40% identity in the fingers themselves. In other finger proteins higher homologies are usually found, but this is mainly due to the conservation of the H/C link, that is not present in kup. Two zinc fingers are found in the C-terminal part of the protein, the first zinc finger starts at Cys 238 and ends at His 258, the second one starts at Cys 349 and ends at His 369, see the comparison with the consensus sequence on figure 2. The two fingers are not much more closely related to each other than to the C_2/H_2 fingers of other proteins. No other finger-like region is present even if one or two differences from the consensus are allowed in the search. The H/C link region between the two fingers is 91 a.a. long. In most proteins the zinc fingers are found as imperfect tandem repeats of several finger motifs, some proteins contain two distant finger domains but each domain includes at least two fingers. The kup protein is unusual from this point of view since it has only two isolated fingers. Several cDNAs for new zinc finger containing proteins were isolated by screening cDNA libraries at low stringency, with probes derived from the finger domains of already isolated genes. However, the cross- hybridization is usually not due to the conservation of the zinc fingers themselves, but rather to the highly conserved H/C link sequence, it is therefore not surprising that the cDNAs isolated by this strategy have a conserved H/C link. We show here that a zinc finger containing protein isolated by another mean has not conserved this H/C link. Some variability in the size of this H/C link has already been observed, it might be as short as two a.a. in the Mel-18 protein (30), while isolated zinc fingers have occasionnaly been observed too (26). It might be that the kup protein folds in a way that the two fingers, separated in the primary structure, are in fact close together in the three dimensional structure. Although separated fingers are very unusual in C_2H_2 finger proteins, two separated fingers of the C_4 type have already been described in EF-1a (or GF-1), that is also mainly expressed in cells of the erythroid lineage (48). The N-terminal part of the kup protein has several short hydrophobic regions and is unusually rich in serine, threonine, proline, glutamine and histidine, as is frequently found in many DNA binding proteins, a region very rich in amino acids with hydroxyl groups (Ser and Thr) is found just upstream of the first zinc finger. However there are no long stretches of one repeated amino acid (such as the Ala repeats found in some repressors, 45). Although the zinc fingers themselves are slighly more closely related to Krüppel, the general structure of the protein is more similar to the one found in the finger protein associated with

Wilm's tumors (33, 34). As all other proteins with similar motifs, kup is therefore expected to be a transcription factor. Preliminary results show that the kup protein, expressed in bacteria, is able to bind DNA. However these results do not demonstrate the specificity of DNA binding and the next step towards understanding the precise function of this putative transcription factor will be to determine the DNA sequence specifically recognized by the kup protein. It has been suggested that the number of fingers might correlate with the length of the DNA sequence recognized by the protein, if this is indeed the case, one would expect that the kup protein recognizes a short motif of DNA or two short motifs, distantly spaced. The preferential expression of kup in hematopoietic cells suggests a specific role in the regulation of genes differentially expressed in this cell type.

ACKNOWLEDGEMENTS

We thank Annie Lamouroux and Jacques Mallet for the gift of their very good cDNA library, Jacques Camonis and Michel Jacquet for the SCD25 plasmid, Dinha N'Diaye for technical assistance, Françoise Vecchini for her help in the construction of the expression vector, Brigitte Sola for the gift of some Northern blots and for stimulating discussions and Armand Tavitian for continuous support and encouragement. P.C. is supported by INSERM.

REFERENCES

1. Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J., **4**, 1609–1614.
2. Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986) Nature, **320**, 283–287.
3. Nehlin, J.O. and Ronne, H. (1990) EMBO J., **9**, 2891–2898.
4. Adams, T.H., Boylan, M.T. and Timberlake, W.E. (1988) Cell, **54**, 353–362.
5. Rosenberg, U.B., Schröder, C., Preiss, A., Kienlin, A., Côté, S., Riede and Jäckle, H. (1986) Nature, **319**, 336–339.
6. Vincent, A., Colot, H.V. and Rosbash, M. (1985) J. molec. Biol., **186**, 146–166.
7. Tautz, D., Lehman, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jäckle, H. (1987) Nature, **327**, 383–389.
8. Boulay, J.L., Dennefeld, C. and Alberga, A. (1987) Nature, **330**, 395–398.
9. Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) Cell, **40**, 37–43.
10. Coulter, D.E., Swaykus, E.A., Beran-Koehn, A., Goldberg, D., Wieschaus, E. and Scheld, P. (1990) EMBO J., **8**, 3795–3804.
11. Moses, K., Ellis, M.C. and Rubin, G.M. (1989) Nature, **340**, 531–536.
12. Baldarelli, R.M., Mahoney, P.A., Salas, F., Gustavson, E., Boyer, P.D., Chang, M.-F., Roark, M. and Lengyel, J.A. (1988) Dev. Biol., **125**, 85–95.
13. Ruiz i Altaba, A., Perry-O'Keefe, H. and Melton, D.A. (1987) EMBO J., **10**, 3065–3070.
14. Nietfeld, W., El-Baradi, T., Mentzel, H., Pieler, T., Koester, M., Poeting, A. and Knoechel, W. (1989) J. molec. Biol., **208**, 639–659.
15. Chowdhury, K., Deutsch,U. and Gruss, P. (1987) Cell, **48**, 771–778.
16. Chavrier, P., Zerial, M., Lemaire, P., Almendral, J., Bravo, R. and Charnay, P. (1988) EMBO J., **7**, 29–35.
17. Lemaire, P., Relevant, O., Bravo, R. and Charnay, P. (1988) Proc.Natl.Acad.Sci. USA, **85**, 4691–4695.
18. Passananti, C., Felsani, A., Caruso, M. and Amati, P. (1989) Proc.Natl.Acad.Sci. USA, **88**, 9417–9421.
19. Joseph, L.J., LeBeau, M.M., Jamieson, Jr., G.A., Acharya, S., Shows, T.B., Rowley, J.D. and Sukhatme, V.K. (1988) Proc.Natl.Acad.Sci. USA, **85**, 7164–7168.
20. Kadonaga, J.T., Carner, K.R., Masiarz, F.R. and Tjian, R. (1987), Cell, **51**, 1079–1090.
21. Pannuti, A., Lanfrancone, L., Pascucci, A., Pellici, P.-G., La Mantia, G. and Lania, L. (1988) Nucleic Acids Res., **16**, 4227–4237.
22. Ruppert, J.M., Kinzler, K.W., Wong, A.J., Bigner, S.H., Kao, F.-T., Law, M.L., Seuanez, H.N., O'Brien, S.J. and Vogelstein, B. (1988) Mol. Cell. Biol., **8**, 3104–3113.
23. Kato, N., Shimotono, K., VanLeeuwen, D. and Cohen, M. (1990) Mol. Cell. Biol., **10**, 4401–4405.
24. Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L.G. (1987) Cell, **51**, 1091–1104.
25. Schneider-Gädicke, A., Beer-Romero, P., Brown, L.G., Mardon, G., Luoh, S.-W. and Page, D.C. (1989) Nature, **342**, 708–711.
26. Lania, L., Donti, E., Pannuti, A., Pascucci, A., Pengue, G., Feliciello, I., La Manta, G., Lanfrancone, L. and Pelicci, P.G. (1990) Genomics, **6**, 333–340.
27. Bellefroid, E.J., Lecocq, P.J., Benhida, A., Poncelet, D.A., Belayem, A. and Martial, J.A. (1989) DNA, **8**, 377–387.
28. Morishita, K., Parker, D.S., Mucenski, M.L., Jenkins, N.A., Copeland, N.G. and Ihle, J.N. (1988) Cell, **54**, 831–840.
29. Ernoult-Lange, M., Kress, M. and Hamer, D. (1990) Mol.Cell.Biol., **10**, 418–421.
30. Tagawa, M., Sakamoto, T., Shigemoto, K., Matsubara, H., Tamura, Y., Ito, T., Nakamura, I., Okitsu, A., Imai, K. and Taniguchi, M. (1990) J.Biol.Chem., **265**, 20021–20026.
31. Wright, J.J., Gunter, K.C., Mitsuya, H., Irving, S.G., Kelly, K. and Siebenlist, U. (1990) Science, **248**, 588–591.
32. Tanaka, K., Miura, N., Satokata, I., Miyamoto, I., Yoshida, M.C., Satoh, Y., Kondo, S., Yasui, A., Okayama, H. and Okada, Y. (1990) Nature, **348**, 73–76.
33. Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., Jones, C. and Housman, D.E. (1990) Cell, **60**, 509–520.
34. Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H. and Bruns, G.A.P. (1990) Nature, **343**, 774–778.
35. Rauscher III, F.J., Morris, J.F., Tournay, O.E., Cook, D.M. and Curran, T. (1990) Science, **250**, 1259–1262.
36. Boy-Marcotte, E., Damak, F., Camonis, J., Garreau, H. and Jacquet,M. (1989) Gene, **77**, 21–30.
37. Grima, B., Lamouroux, A., Boni, C., Julien, J.-F., Javoy-Agid, F. and Mallet, J. (1987) Nature, **326**, 707–711.
38. Sambrook,J., Fritsch, E.F. and Maniatis,T. (1989) Molecular Cloning, a laboratory manual 2nd Ed., C.S.H. Laboratory Press.
39. Mattei, M.-G., Philip,N., Passage, E., Moisan, J.-P., Mandel, J.-L. and Mattei, J.-F. (1985) Human Genetics, **69**, 268–271.
40. Rosenberg, A.H., Lade, B.N., Chui, D.-S., Lin, S.-W., Dunn, J.J. and Studier, F.W. (1987) Gene, **56**, 125–135.
41. Frech, M., Schlichting, I., Wittinghofer,A. and Chardin, P.(1990) J.Biol.Chem., **265**, 6353–6359.
42. Ollo, R. and Maniatis,T. (1987) Proc.Natl.Acad.Sci. USA, **84**, 5700–5704.
43. Chardin, P. (1988) Biochimie, **70**, 865–868.
44. Kozak, M. (1987) Nucleic Acids Res., **15**, 8125–8133.
45. Licht, J.D., Grossel, M.J., Figge, J. and Hansen, U.M. (1990) Nature, **346**, 76–79.
46. Bordereaux, D., Fichelson, S., Tambourin, P. and Gisselbrecht, S. (1990) Oncogene, **5**, 925–927.
47. Gisselbrecht,S., Fichelson, S., Sola, B., Bordereaux, D., Hampe, A., André, C., Galibert, F. and Tambourin, P. (1987) Nature, **329**, 259–261.
48. Tsai, S.-F., Martin, D.I.K., Zon, L.I., D'Andrea, A., Wong, G.G. and Orkin, S.H. (1989) Nature, **339**, 446–451.