GENOMICS **42**, 268–277 (1997)
ARTICLE NO. GE974705

Identification of Two Novel Isoforms of the ZNF162 Gene: A Growing Family of Signal Transduction and Activator of RNA Proteins

Corrado Caslini, Orietta Spinelli, Giovanni Cazzaniga, Josée Golay, Luca De Gioia, Alessandro Pedretti, Ferruccio Breviario, Ricardo Amaru, Tiziano Barbui, Andrea Biondi, Martino Introna, and Alessandro Rambaldi[1]

Divisione di Ematologia, Ospedali Riuniti di Bergamo, 24128 Bergamo; Clinica Pediatrica Università di Milano, Ospedale San Gerardo Monza, Dipartimento di Chimica Inorgânica e Metallorganica Analitica Università di Milano and Istituto di Ricerche Farmacologiche "M.Negri," Milan, Italy

Received October 16, 1996; accepted March 4, 1997

By differential screening of a cDNA library obtained from a GM-CSF-dependent human myeloid leukemia cell line (GF-D8), we identified two novel isoforms of the recently described ZNF162 gene, which is apparently linked to multiple endocrine neoplasia type 1. The shorter of these new isoforms, called B3, presents an open reading frame (ORF) of 1713 bp coding for 571 amino acids. Its nucleotide sequence is homologous to the cDNA coding for the ABCDF isoform of ZNF162, except for a 4-nucleotide insertion that results in a frame shift of the ORF starting from nucleotide 1725 of the ZNF162 sequence. As a consequence, the predicted translation product of B3 contains the consensus sequence of the A motif (G-X-X-X-X-G-K-S) of the "ATP/GTP binding site," which is characteristic of several protein families including protein kinases. Moreover, B3 shows the use of a different stop codon and contains a different tyrosine-rich COOH terminus. The longer isoform, called B4, differs from the ABCDEF isoform of ZNF162 by the insertion, at position 2137, of 383 nucleotides leading to a different, proline-rich COOH terminus. The complex transcription pattern of the ZNF162 gene is characterized by four transcripts, of approximately 3.9, 3.7, 3.2, and 2.9 kb, in GF-D8 cells. The 3.7- and 2.9-kb transcripts are expressed in resting GF-D8 cells. Upon stimulation with GM-CSF the expression of these mRNAs is up-regulated in parallel with the induction of two additional transcripts of 3.9 and 3.2 kb. The same pattern of expression has also been observed in freshly isolated myeloid leukemia cells and normal CD34⁺ stem cells. In light of these data, and since GM-CSF is known to stimulate signal transduction pathways, it becomes relevant that all the different isoforms of ZNF162 contain the KH module, which is a sequence motif present in proteins playing a major role in regulating cellular RNA metabolism. A search for functional domains demonstrates that ZNF162 belongs to a new and growing family of genes dubbed STAR (signal transduction and activator of RNA) proteins that are thought to play a downstream role in cell signaling and also in RNA binding. The mammalian members include Sam68, which is a target of Src, Fyn, and Grb2, and the newly cloned mouse quaking proteins (qkI) necessary in early embryogenesis and myelination. Moreover, since ZNF162 is highly conserved from yeast to humans, it implies that this new pathway has a significant function. © 1997 Academic Press

INTRODUCTION

Development, maintenance, and function of hematopoietic progenitor cells are under the control of a wide spectrum of cytokines (Ogawa, 1993). The interaction of these growth factors with their receptors triggers specific events such as gene transcription and protein expression leading to proliferation and differentiation of bone marrow stem cells as well as activation of mature effector cells circulating in the peripheral blood. Interleukin-3 (IL-3) and granulocyte–macrophage colony-stimulating factor (GM-CSF) are known to exert a series of overlapping effects on hematopoietic progenitor cells (Metcalf, 1989). In recent years signal transduction pathways that involve activation of Janus tyrosine kinases (JAK) and signal transducer and activator of transcription (STAT) proteins have been linked to a number of receptor systems including those hematopoietins that lack intrinsic tyrosine kinase activity but activate cytoplasmic tyrosine kinases (Ihle *et al.*, 1994).

This work was supported in part by grants from the Associazione Italiana per la Ricerca contro il Cancro (AIRC), from the Consiglio Nazionale delle Ricerche (progetto finalizzato ACRO), from the Fondazione Tettamanti, and from the Associazione Paolo Belli, Lotta alla Leucemia. C.C. was supported by a fellowship from AIRC and J.G. was supported by the Istituto Superiore di Sanità (AIDS grants).

Sequence data from this article have been deposited with the EMBL Data Library under Accession Nos. L49345 and L49380.

[1] To whom correspondence should be addressed at the Division of Hematology, Ospedali Riuniti di Bergamo, Largo Barozzi, 1, 24128 Bergamo, Italy. Telephone: 39-35-269491. Fax: 39-35-269667. E-mail: ematologia@cyberg.it.

0888-7543/97 $25.00
Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.

JAK kinases, upon cytokine receptor activation, phosphorylate STAT proteins, which translocate to the nucleus where they activate target genes by binding to specific regulatory sequences. Myeloid leukemic cell lines established from patients with acute myeloid leukemia have been largely used as models to identify and clone genes controlling myeloid differentiation and proliferation (Oval and Taetle, 1990). We established a new growth factor-dependent acute myeloid leukemia cell line named GF-D8 (Rambaldi *et al.,* 1993), whose long-term survival and proliferation *in vitro* are strictly dependent upon the presence of GM-CSF or IL-3. Many other cytokines including G-CSF, IL-6, and Flt ligand are unable to support the proliferation of GF-D8, making this experimental model attractive to search for GM-CSF-regulated genes. A number of growth factor-induced genes were identified by differential hybridization of cDNA libraries using probes obtained from activated compared with quiescent cells (Moscinski and Prystowsky, 1990). Although this method is limited in its sensitivity and detects only those transcripts that are expressed in relatively high abundance, it allowed the identification, in cultured vascular endothelial cells, of a novel IL-1β-inducible pentraxin-related gene called PTX-3 (Breviario *et al.,* 1992). Thus we use this strategy to characterize new genes whose transcription could be induced or regulated by GM-CSF.

MATERIALS AND METHODS

Cells. The establishment and characterization of the human acute myeloid leukemia cell line GF-D8 has been reported previously (Rambaldi *et al.,* 1993). The cells are routinely grown in RPMI 1640 (GIBCO, Grand Island, NY) supplemented with 10% fetal calf serum (Hyclone; Sterile System, Inc, Logan, UT) and 50 ng/ml human recombinant GM-CSF (Genzyme, Boston, MA; sp act 5×10^7 proliferation units/mg of protein).

Upon informed consent, human hematopoietic stem cells ($CD34^+$ cells) were obtained from the peripheral blood of normal donors treated with G-CSF to mobilize progenitor cells for allogeneic transplantation (Rambaldi *et al.,* 1996). A highly purified fraction of $CD34^+$ cells (>95% pure) was obtained by positive selection using CD34 magnetic microbeads, XS+ column, and a SuperMACS separator (Miltenyi Biotec, Germany), which allows collection of up to 1×10^9 positive cells.

Acute myeloid leukemic cells were obtained from bone marrow aspirates from patients at diagnosis. The leukemic blast population was purified by Ficoll gradient sedimentation and was more than 90% pure as assessed by conventional morphology and flow cytometric analysis using a FACScan apparatus (Becton–Dickinson, Mountain View, CA) and stem cell- and myeloid-specific monoclonal antibodies (HPCA-2, anti-CD34; Leu M7, anti-CD13; and Leu M9, anti-CD33) (Becton–Dickinson).

cDNA libraries and differential screening. Total RNA was extracted and purified by guanidinium isothiocyanate and cesium chloride gradient as previously described (Rambaldi *et al.,* 1987) from GF-D8 cells cultured for 2 h in the presence or absence of both GM-CSF (50 ng/ml) and cycloheximide (10 μg/ml). Poly(A)$^+$ RNA was further purified by affinity chromatography on oligo(dT)-cellulose. A cDNA library was constructed in the Lambda ZAPII vector (Stratagene, La Jolla, CA) as described (Breviario *et al.,* 1992; Sambrook *et al.,* 1989). The library was plated at approximately 1 plaque/cm^2 and transferred onto nitrocellulose membranes (Schleicher & Schuell, Dassel, Germany) in duplicates by standard procedures (Breviario *et al.,* 1992). The cDNA probes from resting and GM-CSF-activated GF-D8 cells were obtained as described (Breviario *et al.,* 1992; Berger and Kimmel, 1987), and the membranes were hybridized at 42°C for 48 h in 0.2 *M* Tris–HCl, pH 7.6, 50% deionized formamide, 5× SSC, 1× Denhardt's solution, 10% dextran sulfate, 0.1% SDS, 100 μg/ml denatured salmon sperm DNA, and 1×10^7 cpm/100 cm^2 cDNA probe. Membranes were washed at 50–60°C in 0.2× SSC, 0.1% SDS. Plaques showing a stronger hybridization signal with the cDNA probes obtained from GM-CSF-stimulated relative to untreated GF-D8 were picked. These were rescreened three times differentially to obtain single clones. Phage inserts were rescued in the Bluescript vector and sequenced with the dideoxynucleotide chain termination method (Tabor and Richarson, 1987).

Two additional human cDNA libraries, generated from PHA-stimulated peripheral blood lymphocytes and from the KG1 myeloid leukemia cell line (Clontech Laboratories, Palo Alto, CA), were screened.

Northern blot analysis. Northern blot analysis was performed using either total cellular RNA or poly(A)$^+$ selected mRNA. RNA samples were electrophoresed through a 1% agarose gel with 7% formaldehyde and blotted onto GeneScreen Plus membranes (New England Nuclear, Boston, MA). To control equal RNA loading, filters were hybridized with the cDNA coding for the human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (Tso *et al.,* 1985). Probes were labeled to a specific activity of 10^9 cpm/mg by using hexanucleotide primers and [^{32}P]dCTP (Feinberg and Volgestein, 1983) and hybridization was performed as described previously (Rambaldi *et al.,* 1987).

RT-PCR. RT-PCR reactions were performed using cDNAs synthesized with 1 μg of total denaturated RNA (from GM-CSF-activated GF-D8 cells) in a 20-μl reaction mixture containing 100 μM dNTPs, 1× First-strand buffer, 1 μl Superscript RT RNase H$^-$ virus reverse transcriptase (Gibco BRL, Gent, Belgium), 20 U RNase inhibitor, and random hexamers, using a commercial kit (Perkin–Elmer Cetus). All PCR experiments were performed using an automated heat block (PCR System 9600; Perkin–Elmer Cetus).

cDNA clones containing the whole coding sequences of the new ZNF162 isoforms were obtained *in vitro* by Expand High Fidelity PCR System (Boehringer, Mannheim). The PCR was performed by adding 5 μl of the cDNA product to a PCR mixture containing 1.5 m*M* $MgCl_2$, 1× Expand HF buffer, 400 μM dNTPs, 2.6 U of a mixture of *Taq* and *Pwo* DNA polymerases, 300 n*M* 3′ amplimer common to all ZNF162 isoforms (nucleotides 2501–2520, according to Toda *et al.,* 1994), and 300 n*M* 5′ amplimer corresponding to the 5′ untranslated sequence of ZNF162 (nucleotides 328 to 347, Toda *et al.,* 1994). After an initial denaturation at 95°C for 1 min, 12 cycles of amplification were performed, each consisting of a denaturation step at 95°C for 30 s, an annealing step at 56°C for 30 s, and an extension at 72°C for 3 min; and 16 cycles each consisting of a denaturation step at 95°C for 30 s, an annealing step at 56°C for 30 s, and an extension at 72°C for 3 min + 20 s/cycle. A final extension at 72°C for 10 min (1 cycle) and cooling at 4°C was done at the end. The amplified DNAs were directly subcloned from the PCR into a pMOS plasmid (Amersham, Buckinghamshire, UK).

A specific region of the B74b4 cDNA was subcloned by RT-PCR using as 5′ primer the oligonucleotide CTCCCATGGACCCTTCTA and as 3′ primer the oligonucleotide TCCCAAGCGAATCCTCAG (both derived from the B74b4 sequence). Five microliters of the cDNA product were added to a PCR mixture containing 1.5 m*M* $MgCl_2$, 1× PCR buffer (Perkin–Elmer Cetus), 200 μM dNTPs, and 2.5 U of *Taq* DNA polymerase. After an initial denaturation at 95°C for 1 min, 35 cycles of amplification reactions were performed, each consisting of a denaturation step at 95°C for 30 s, an annealing step at 54°C for 30 s, and an extension at 72°C for 30 s. The 5′ portion of ZNF162 was similarly cloned by PCR using as 5′ primer an oligonucleotide from position 45 to 66 and as 3′ primer an oligonucleotide from position 1024 to 1046 (both derived from the published sequences of ZNF162) (Toda *et al.,* 1994) (using the same amplification protocol described above).

In vitro transcription and translation. The cDNA clones B3OS (subcloned into the pBluescript II SK$^{(+/-)}$ vector downstream of the T3 promoter) and B4OS (cloned into the pMOS vector), containing the complete open reading frame (ORF) of the two new isoforms of the ZNF162 gene, were used for the *in vitro* transcription and translation assay. One microgram of each recombinant plasmid was

used in a rabbit reticulocyte lysate assay with a commercially available kit (Promega) in the presence of ^{35}S-labeled methionine. The final products were run in a standard 8.5% SDS–polyacrylamide gel, fixed in 40% methanol–10% acetic acid, dried, and autoradiographed.

Computer methods. All sequences were obtained from the Swissprot WWW and GenBank servers. Multiple sequence alignment was performed using the software package Camelon (Oxford Molecular, Ltd., Oxford, UK) and ClustalW 1.6. The mutation data matrix (Dayhoff *et al.,* 1979) was used as similarity matrix. Similar results were obtained with other matrices (data not shown). All sequences were scanned against the Prosite database (Bairoch, 1991) to look for sites and patterns.

RESULTS

Identification and Characterization of Two New Isoforms of the ZNF162 Gene

A cDNA library was constructed from poly(A)$^+$ mRNA purified from the GF-D8 cell line stimulated with GM-CSF for 2 h in the presence of cycloheximide. More than 5000 recombinant clones were screened for GM-CSF-inducible genes by differential hybridization using the probe from GM-CSF-stimulated cells and from untreated cells. After the first screening, 76 clones were identified for their stronger signal but only 20 were found positive after the third differential screening. All these clones were partially (approximately 200-bp) sequenced at the 3′ and 5′ ends. Sequences from 17 of 20 clones were found identical to already known genes including the eosinophil basic major protein (3 clones) and c-myc (2 clones) as expected from the activation of the cell cycle progression induced by GM-CSF. Partial nucleotide sequences from 3 of these clones (B74c1, B74b4, and B74b3) revealed a partial identity with the ZNF162 nuclear protein (originally mentioned as ZFM1) recently described by Toda and co-workers (1994). Although the ZNF162 sequence showed a high degree of identity with the sequence of our cDNA clones, we noticed some important differences. As shown in Fig. 1A, the B74b3 cDNA clone differs only in a 4-nucleotide insertion at nucleotide 1724 of the ABCDF isoform of ZNF162 (Toda *et al.,* 1994), resulting in a frame shift of the ORF. Clone B74b4 contains the insertion of 383 additional nucleotides at position 2137 of the nucleotide sequence coding for the ABCDEF isoform of ZNF162 (Toda *et al.,* 1994). Again, this insertion determines a different sequence of the putative protein product toward the COOH terminus. The same differences in the nucleotide sequences were found in several similar clones isolated from two other cDNA libraries generated from mitogen-stimulated T lymphocytes (schematically indicated in Fig. 1A as T7, T9, T1, and T4) and the KG1 myeloid leukemia cell line (schematically indicated in Fig. 1A as KG4 and KG10).

To clone the cDNAs containing the whole coding sequences of these two new isoforms of ZNF162, we performed a series of RT-PCR amplification experiments with the use of RNA from GM-CSF-activated GF-D8 cells, a 3′ primer common to all isoforms, and a 5′ primer corresponding to the 5′ untranslated sequence of ZNF162 (see Fig. 1A for position of primers). Using this approach, we amplified two cDNA clones: B4OS, a 2576-bp-long sequence with an ORF of 1917 bp coding for 639 aa, and B3OS, a 1943-bp-long sequence that presents an ORF of 1713 bp coding for 571 aa. As observed in B74b3, the nucleotide sequence of B3OS confirms the presence of the 4 additional nucleotides (GTAA) at position 1724 of the ZNF162 ABCDF sequence. On the other hand, the complete sequence of B4OS confirmed the absence of the 4 nucleotides present in B3OS, while showing the presence of the insertion of 383 nucleotides already found in several cDNA clones. The additional 383-nucleotide sequence is shown in Fig. 1B. The B3OS and B4OS cDNAs were therefore considered representative of two new isoforms of the ZNF162 gene. They will be hereafter referred to as B3 and B4, respectively.

To verify that the cDNA clones B3OS and B4OS coding for two new isoforms of ZNF162 could effectively initiate protein synthesis, experiments of *in vitro* transcription and translation were performed using the reticulocyte lysate assay. B3OS and B4OS, upon subcloning, were indeed able to generate two protein products of 62 and 68 kDa, respectively, in total agreement with the molecular weight predicted by their ORF (data not shown).

Expression of ZNF162 Is Regulated by GM-CSF

ZNF162 was originally cloned from human brain and fetal liver (Toda *et al.,* 1994). To confirm its expression in myeloid cells and its inducibility by GM-CSF, we used the B74c1 cDNA clone (919 bp long, corresponding to the 3′ end of ZNF162) as a probe on Northern blots of total cell RNA as well as on poly(A)$^+$ RNA isolated from resting and GM-CSF-activated GF-D8 cells. Figure 2A shows low but appreciable levels of ZNF162 mRNAs in resting GF-D8 cells; on the contrary, after stimulation for 2 h with GM-CSF in the presence of cycloheximide, a significant induction of mRNA transcription can be detected, suggesting that ZNF162 may represent an early response gene to GM-CSF. When poly(A)$^+$ RNA was used, it became more evident that the pattern of transcription is highly complex (Fig. 2B). Under basal conditions, two broad bands with an apparent molecular weight of 2.9 and 3.7 kb are evident, whereas upon treatment with GM-CSF the intensity of these two bands is consistently increased and accompanied by the appearance of two new bands of 3.2 and 3.9 kb. Moreover, by RT-PCR, we were able to generate a cDNA probe specific for the 383-nucleotide sequence exclusively present in the B4 isoform (Fig. 1A) and identified as B4xOS. When used as a probe in Northern blot experiments, this B4-specific sequence hybridized only to the high-molecular-weight transcripts (3.9 and 3.7 kb), indicating that this portion of the nucleotide sequence of ZNF162 is retained only in the longer isoforms of this gene (Fig. 2C). Very similar data were obtained when GF-D8 cells were exposed to IL-3, which is another growth factor for this cell line (Rambaldi *et al.,* 1993 and data not shown).

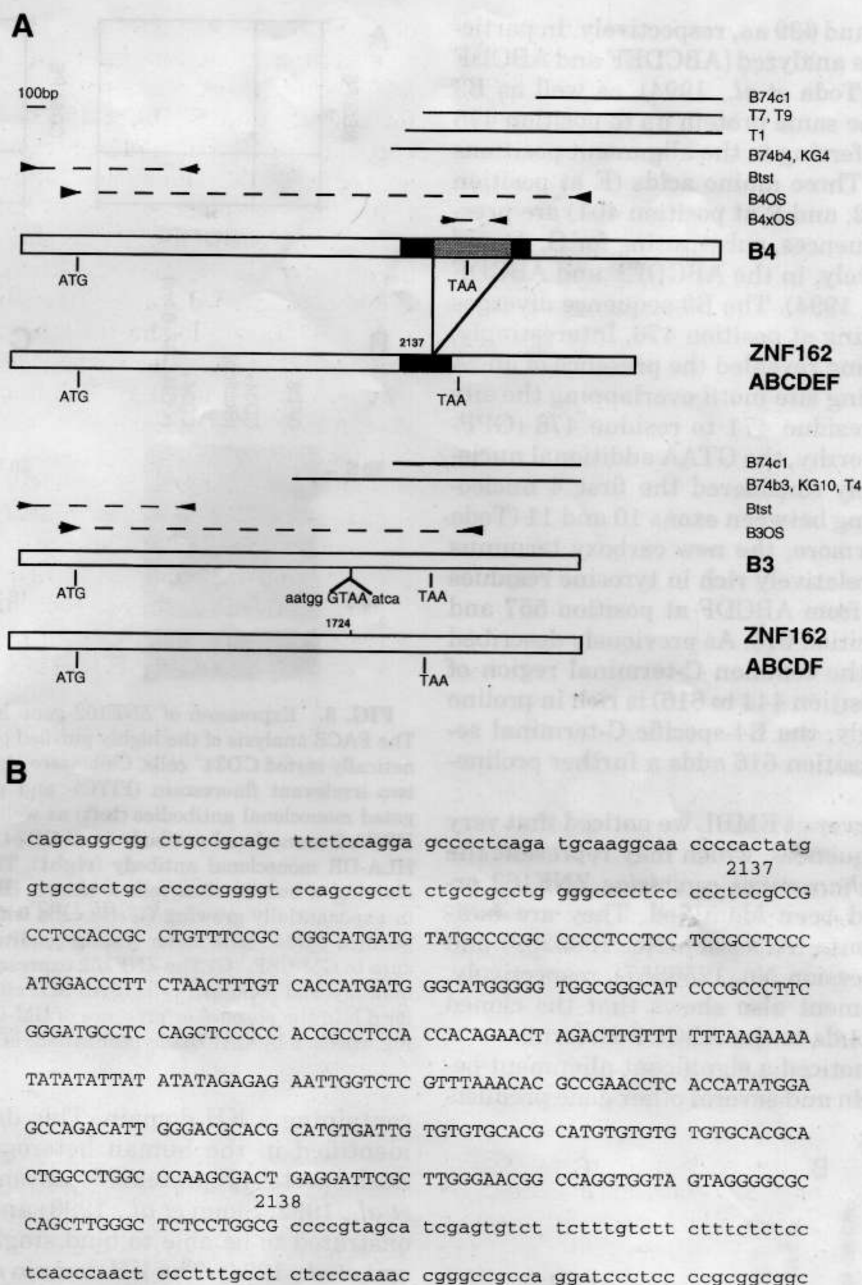


FIG. 1. (A) A schematic diagram of the ZNF162 isoforms (ABCDEF, ABCDF, B3, and B4) with the overlapping cDNA clones that were obtained by the screening of three different cDNA libraries (plain lines). Clones B74c1, B74b4, and B74b3 were obtained from the stimulated GF-D8 library; clones T1, T4, T7, and T9 were obtained from a stimulated T cell library; and clones KG4 and KG10 were obtained from a KG1 cell library. Clones obtained by RT-PCR are shown as dashed lines (with arrows at both ends indicating the position of primers). The positions of the ATG and TAA codons are indicated. The black box indicates the E portion of ZNF162 (Toda *et al.*, 1994). (**B**) The sequence of the additional 383 nucleotides of the B4 isoform (in capital letters), inserted at position 2137 within the E portion of the ABCDEF isoform (the gray box in **A**).

To study whether ZNF162 expression was restricted to the GF-D8 cell line or, rather, was detectable also in other hematopoietic cells, we isolated highly purified populations of human CD34$^+$ stem cells (Fig. 3A). As shown in Fig. 3B, ZNF162 expression was already evident in resting cells and augmented by 6-h exposure to GM-CSF.

Finally, ZNF162 gene expression is detectable in the majority of freshly isolated human myeloid leukemic samples (Fig. 3C and data not shown). Such levels of expression may reflect *in vivo* cytokine stimulation, as suggested by the decline of the message when the leukemic cells from one patient were maintained for 6 h in the absence of GM-CSF (Fig. 3C).

ZNF162 Belongs to the Signal Transduction and Activator of RNA (STAR) Protein Family

As can be seen from the alignment in Fig. 4, the B3 and B4 cDNAs code for two novel isoforms of the

ZNF162 gene, of 571 and 639 aa, respectively. In particular, the four isoforms analyzed [ABCDEF and ABCDF previously reported (Toda *et al.,* 1994), as well as B3 and B4] encode for the same protein up to position 475 (with the numbers referring to the alignment positions as shown in Fig. 4). Three amino acids (E at position 291, R at position 372, and R at position 401) are present in B3 and B4 sequences, substituting for G, A, and W residues, respectively, in the ABCDEF and ABCDF isoforms (Toda *et al.,* 1994). The B3 sequence diverges from the others starting at position 476. Interestingly, the PROSITE screening revealed the presence of an "A type" ATP/GTP-binding site motif overlapping the site of divergence from residue 471 to residue 478 (GPPPMGKS). Also noteworthy, the GTAA additional nucleotides were previously considered the first 4 nucleotides of the intron lying between exons 10 and 11 (Toda *et al.,* 1994). Furthermore, the new carboxy terminus of B3 appears to be relatively rich in tyrosine residues (8/164). B4 diverges from ABCDF at position 557 and from ABCDEF at position 616. As previously described (Toda *et al.,* 1994), the common C-terminal region of these isoforms (aa position 444 to 616) is rich in proline residues. Interestingly, the B4-specific C-terminal sequence starting at position 616 adds a further proline-rich region (23/53).

Using the Blast server at EMBL we noticed that very recently two new sequences, which may represent the mouse and the *Saccharomyces cerevisiae* ZNF162 orthologous genes, had been identified. They are indicated as CW17R Mouse (Accession No. 1083269) and CW17R SacCer (Accession No. 1256857), respectively, in Fig. 4. The alignment also shows that the cloned mouse gene corresponds to the ABCDF isoform.

Furthermore, we noticed a significant alignment between ZNF162 protein and several other gene products

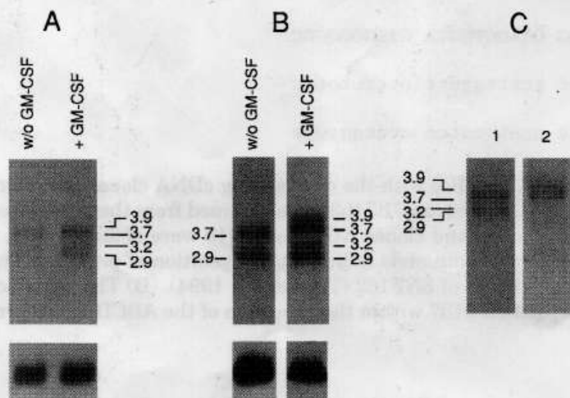


FIG. 2. Expression of the ZNF162 gene in GF-D8 cells. Northern blot analysis of total cellular RNA (**A**) and poly(A)$^+$ RNA (**B**) from resting (w/o GM-CSF) and GM-CSF-stimulated (+GM-CSF) GF-D8 cells, probed with the B74c1 cDNA clone. In **C,** total RNA from GF-D8 cells was sequentially hybridized on the same filter with the B74c1 cDNA clone (lane **1**) and, upon removal of this probe by high stringency washing, with a B4-specific sequence (clone B4xOS, shown in Fig. 1). The latter hybridizes only to the high-molecular-weight transcripts (3.9 and 3.7 kb) (lane **2**). **Bottom** panels in **A** and **B** show by hybridization with GAPDH cDNA that equal amounts of RNA were loaded.

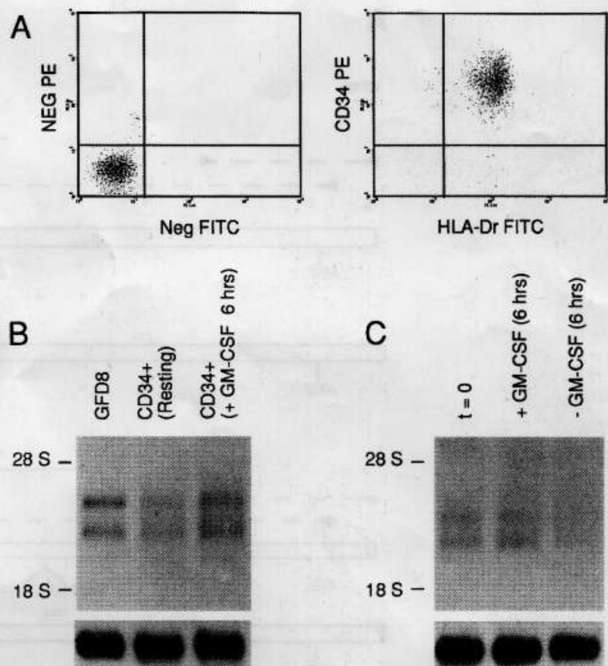


FIG. 3. Expression of ZNF162 gene in hematopoietic cells. (**A**) The FACS analysis of the highly purified population of immunomagnetically sorted CD34$^+$ cells. Cells were simultaneously stained with two irrelevant fluorescein (FITC)- and phycoerythrin (PE)-conjugated monoclonal antibodies (**left**) as well as with FITC-conjugated HPCA-2 monoclonal antibody (anti-CD34) and PE-conjugated anti-HLA-DR monoclonal antibody (**right**). The typical double-positive staining of human stem cells is shown. (**B**) The ZNF162 expression in exponentially growing GF-D8 cells with GM-CSF and in highly purified CD34$^+$ cells under resting conditions and after 6 h of exposure to GM-CSF. (**C**) The ZNF162 expression in freshly isolated human myeloid leukemia cells (FAB M1) either untreated or cultured for 6 h in the absence or presence of GM-CSF. The equal RNA loading, checked by GAPDH hybridization, is shown below.

containing a KH domain. This domain was originally identified in the human heterogeneous nuclear ribonucleoprotein K protein (Swanson *et al.,* 1988; Matunis *et al.,* 1992; Siomi *et al.,* 1993) and has now been demonstrated to be able to bind single-stranded RNA (Siomi *et al.,* 1994). The KH domain extends from position 1 to position 100 in the alignment shown in Fig. 5, according to the exact definition of the boundaries of the KH domain recently obtained with the three-dimensional solution structure (Musco *et al.,* 1996).

Although a very large group of proteins that bind nonspecifically to ssRNA show the KH domain (Musco *et al.,* 1996), a subgroup containing the ZNF162 protein has been identified for structural reasons and shown to include the ZNF162 gene, the Grp33 from *Artemia salina* (brine shrimp) (Cruz-Alvarez and Pellicer, 1987) (Accession No. P13230), the human Sam68 (previously and erroneously identified as p62) (Wong *et al.,* 1992; Lock *et al.,* 1996) (Accession No. 420044), the quaking gene (qkI) from mouse (Ebersole *et al.,* 1996) (Accession No. 1181698), the Gld-1 Cele from *Caenorhabditis elegans* (Jones and Schedl, 1995) (Accession No. 841255), and the YKCA Cele gene also from *C. elegans* (Accession No. P42083). We now update this growing family,

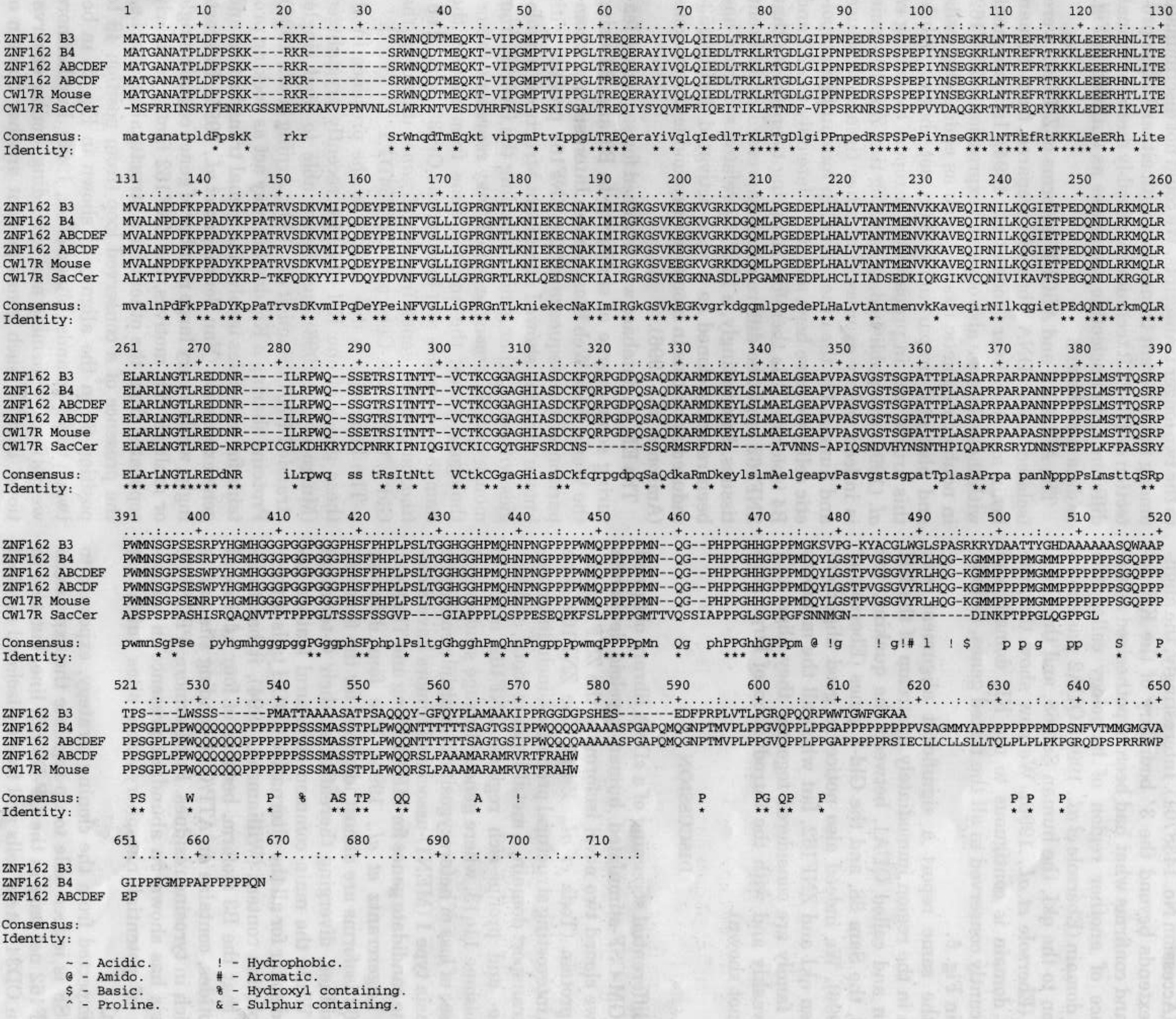


FIG. 4. Multiple sequence alignment of ZNF162 isoforms (ABCDEF, B4, ABCDF, and B3) with the CW17R mouse and CW17R SacCer genes. Asterisks indicate identity.

showing the alignment also with CW17R from mouse (Accession No. 1083269) and CW17R SacCer from *S. cerevisiae* (Accession No. 1256857), the mouse orthologue of Sam 68 (Accession No. 602251), and yet another *C. elegans* gene, F54D1.1 Cele (Wilson *et al.,* 1994) (Accession No. 1438701). Interestingly, the alignment extends beyond the 3′ boundary of the KH domain and confirms what had been hypothesized as the presence of another region of homology called the QUA2 domain (Ebersole *et al.,* 1996). QUA2 was shown common to the qkI, the human Sam68, and the Gld-1 genes (Ebersole *et al.,* 1996). We now show that the QUA2 domain is contiguous to the KH domain and significantly conserved in all the tested genes, as indicated in Fig. 5.

In the same report a significant alignment was shown in the region immediately upstream of the KH domain and called QUA1, between the quaking, the Grp33, the Sam 68, and the Gld-1 genes (Ebersole *et al.,* 1996). We indeed also noticed a good alignment between qkI and ZNF162, but when all the members of the family are considered together they align only very weakly and with the introduction of large gaps (data not shown).

DISCUSSION

By differential screening of a cDNA library obtained from GM-CSF-stimulated human myeloid leukemia cells, we cloned two novel isoforms of the ZNF162 nuclear protein (Toda *et al.,* 1994). The ZNF162 gene shows interesting structural properties, including a nuclear transport domain, a metal-binding motif, and glutamine- and proline-rich regions. ZNF162 maps to chromosome 11q13, where genetic linkage studies and analyses of loss of heterozygosity in multiple endocrine neoplasia type 1 (MEN1)-associated tumors have localized the candidate genes for MEN1 (Pang and Thakker, 1994; Lagercrantz *et al.,* 1995). The four different ZNF162 isoforms are almost identical for the first 447 aa residues, diverging thereafter at their C-terminal ends. While the main common structural feature described so far for all the alternative COOH tails is their high proline content (Williamson, 1994), it has to be noted that the B3 isoform, because of four additional nucleotides, contains an ATP/GTP-binding site and is also rich in tyrosine residues. Furthermore, PROSITE analysis has shown an abundant presence in all isoforms of potential phosphorylation sites (data not shown).

As expected from the cloning strategy, exposure to GM-CSF is indeed able to up-regulate the expression of ZNF162 not only in the GF-D8 cell line, but also in human $CD34^+$ stem cells and freshly isolated myeloid leukemias. More in particular, we noticed the presence of four different transcripts even though we cannot yet determine their relative contribution to the four different isoforms so far characterized. Nonetheless, the B4-specific probe we generated by RT-PCR hybridized only to the two longer transcripts. We could not discriminate among the different transcripts with either RNase protection or RT-PCR techniques, possibly due to marginal differences in nucleotide sequences as outlined in Fig. 1 (data not shown).

The ABCDEF and ABCDF isoforms of ZNF162 were isolated from cDNA libraries prepared from human brain and fetal liver (Toda *et al.,* 1994). In agreement with these data we also observed expression of ZNF162 in nonhematopoietic tissues, such as endothelial cells and placental preparations (data not shown). While this article was in preparation, a report by Arning *et al.* (1996) was published, suggesting that ZNF162 gene is indeed widely expressed in other tissues of mouse and human. On the other hand, in human hematopoietic cells, we have been able to isolate only the B3 and B4 isoforms. It is therefore tempting to speculate that ZNF162 might be differentially expressed in different tissues. Interestingly, a ZNF162-related sequence has been recently cloned as a gene activated during p53-induced apoptosis in mouse M1 myeloid leukemia cells (Amson *et al.,* 1996).

The EMBL database was searched for sequences featuring some degree of similarity to B3 and B4 using the World Wide Web Blast server. Interestingly, a comparison of the sequence of the ZNF162 product with known proteins showed significant structural homologies with nine other genes, including the CW17R from mouse and *S. cerevisiae,* which may represent the orthologous genes; the quaking gene from mouse; the Sam68 from mouse and human; the Grp33 from *A. salina;* and three genes from *C. elegans,* including the Gld-1, the YKCA-Cele, and the F54D1.1.

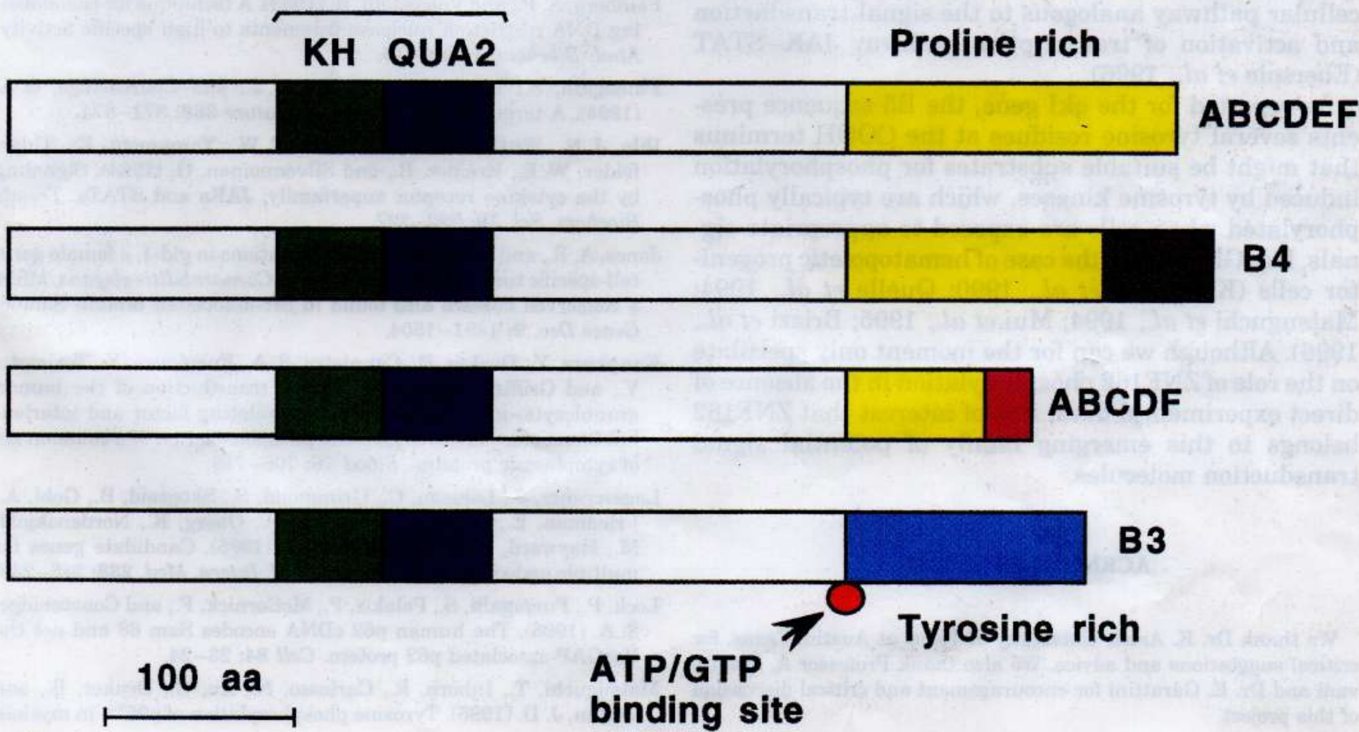
The homology is due mainly to the presence in all these genes of one KH domain, recently crystallized (Musco *et al.,* 1996) and shown able to bind single-stranded RNA (Siomi *et al.,* 1994; Arning *et al.,* 1996). Proteins sharing a KH domain may act as docking proteins for molecules involved in signal transduction and gene expression (Van Seuningen *et al.,* 1995). Interestingly, while a very large number of genes contain single or multiple KH motives, the ZNF162 family is clearly separated from the others for a structural reason, i.e., the presence of a conserved loop (from position 48 to position 68 in the alignment shown in Fig. 5) in between two β strands (Musco *et al.,* 1996). Furthermore we noticed, confirming and extending similar observations, that all such genes present another domain of

FIG. 5. Multiple sequence alignments of KH and QUA2 domains from 10 different members of the STAR family. Asterisks indicate identity.

FIG. 6. Schematic representation of the different ZNF162 isoforms. The KH and QUA2 domains are indicated in green and blue, respectively. The alternative COOH termini with proline- and tyrosine-rich regions are indicated with different colors. The complete sequences of B3 and B4 isoforms have been submitted to the EMBL Database under Accession Nos. L49345 and L49380, respectively.

```
                   1        10        20        30        40        50        60
                   ....:....+....:....+....:....+....:....+....:....+....:....+....:
ZNF162 Human       RVSDKVMIPQDEYPEINFVGLLIGPRGNTLKNIEKECNAKIMIRGKGSVKE-GKVG--RKDG-QM
CW17R Mouse        RVSDKVMIPQDEYPEINFVGLLIGPRGNTLKNIEKECNAKIMIRGKGSVEE-GKVG--RKDG-QM
CW17R SacCer       KFQDKYYIPVDQYPDVNFVGLLLGPRGRTLRKLQEDSNCKIAIRGRGSVKE-GKNA--SDLP-PG
QkI Mouse          QLQEKLYVPVKEYPDFNFVGRILGPRGLTAKQLEAETGCKIMVRGKGSMRDKKKEE--QNRGKPN
Gld-1 Cele         SITEKIYVPKNEYPDYNFVGRILGPRGMTAKQLEQDTGCKIMVRGKGSMRDKSKES--AHRGKAN
Sam68 Human        KLKERVLIPVKQYPKFNFVGKILGPQGNTIKRLQEETGAKISVLGKGSMRDKAKEEELRKGGDPK
Sam68 Mouse        KLKERVLIPVKQYPKFNFVRKILGPQGNTIKRLQEETGAKISVLGKGSMRDKAKEEELRKGGDPK
F54D1.1 Cele       YIPEPPVSIDGKKVKCNYIGRILGPSGMSARMIENQYDVTLLIRGAGSVRNKAMDERVRKRN---
YKCA Cele          -------------NNTNPVGRLIGPRGMTIRQLEKDLGCKLFIRGKGCTKDDAKEE--RLRERVG
GRP3 Artsa         KLVSRCCLPVDQFPKYNFLGKLLGPGGSTMKQLQDETMTKISILGRGSMRDRNKEEELRNSGDVK

Consensus:          ! ~$  !p   yp  Nfvg$!lGP G t!k l  ~   ki !rGkGs $~ !k~   r  g
Identity:                          *    ** *             * *

                   66      75        85        95        105       115       125
                   ....+....:....+....:....+....:....+....:....+....:....+....:....+
ZNF162 Human       LPGEDEPLHALVTANTMEN---VKKAVEQIRNILKQGIETPE---DQNDLRKMQLRELARLNGT
CW17R Mouse        LPGEDEPLHALVTANTMEN---VKKAVEQIRNILKQGIETPE----DQNDLRKMQLRELARLNGT
CW17R SacCer       AMNFEDPLHCLIIADSEDK---IQKGIKVCQNIVIKAVTSPE----GQNDLKRGQLRELAELNGT
QkI Mouse          WEHLNEDLHVLITVEDAQNR-AEIKLKRAVEEVKKLLVPAAE----GEDSLKKMQLMELAILNGT
Gld-1 Cele         WEHLEDDLHVLVQCEDTENR-VHIKLQAALEQVKKLLIPAPE----GTDELKRKQLMELAIINGT
Sam68 Human        YAHLNMDLHVFIEVFGPPCE-AYALMAHAMEEVKKFLVPD------MMDDICQEQFLELSYLNGV
Sam68 Mouse        YAHLNMDLHVFIEVFGPPCE-AYALMAHAMEEVKKFLVPD------MMDDICQEQFLELSYLNGV
F54D1.1 Cele       -EHLEEPLHVLLIARHNDKTKCEEILNKAAEKIESLLTPI-------HDEYKMDQLVSYAKMNGT
YKCA Cele          WEHLKEPIHVMISVRSDSE----EAASEKLSSIKKMLQEFLEH---TDSELKRSQLMQLAVIEGT
GRP3 Artsa         YAHLNEQLHIEIISIASPAE-AHARMAYALTEIKKYITPEEDPNYMMMAGHGAGPMMGMGGMMGG

Consensus:            hl ~ lH! ! !           !  !!  ! k !!  ~          ~!$  ql ela !nGt
Identity:                  *                                                      *

       ~ - Acidic.      ! - Hydrophobic.
       @ - Amido.       # - Aromatic.
       $ - Basic.       % - Hydroxyl containing.
       ^ - Proline.     & - Sulphur containing.
```

5



6

homology immediately adjacent to the 3′ end of the KH domain and called QUA2 (Ebersole *et al.,* 1996), which may take up an α helical conformation (data not shown).

Intriguingly, for three genes of the family genetic lesions have been described; the ZNF162 may be a likely candidate for the genetic lesion responsible for the MEN1 syndrome; the Gld-1 is subject to three different mutations that can give two different phenotypes, one is a loss of function and involves the RNA-binding ability of the KH domain, the other two, more interestingly, give a change in phenotype (MOG for masculinization of germline) (Jones and Schedel, 1995) and are in the loop region, which is conserved and present only in this group of genes. Therefore it has been hypothesized that functions other than the RNA binding may be associated with the loop sequence. Finally, the recently cloned qkI gene is responsible for the quaking mutation in mouse. The qkI gene is expressed in the earliest cells of the embryonic nervous system and in the myelinating tracts of the neonatal brain and can suffer for truncations or ENU-induced point mutations in a region immediately upstream of the KH domain (Ebersole *et al.,* 1996).

The presence of the Sam68 gene, which has been demonstrated to be a target of Src, Fyn, and Grb2 phosphorylation (Fumagalli *et al.,* 1994), and the clear identification of a subclass of KH genes led to the proposal to identify such genes as STAR, indicating the association of QUA and KH domains as a marker for this family (Fig. 6). We now further substantiate this proposal by the demonstration that ZNF162 is a GM-CSF-inducible gene and with the alignment of 10 different genes that clearly indicate the contemporary presence of KH and QUA2 domains in all members. This new group of STAR proteins has been proposed as a general cellular pathway analogous to the signal transduction and activation of transcription pathway JAK–STAT (Ebersole *et al.,* 1996).

As reported for the qkI gene, the B3 sequence presents several tyrosine residues at the COOH terminus that might be suitable substrates for phosphorylation induced by tyrosine kinases, which are typically phosphorylated when cells are exposed to appropriate signals, like GM-CSF in the case of hematopoietic progenitor cells (Kanakura *et al.,* 1990; Quelle *et al.,* 1994; Matsuguchi *et al.,* 1994; Mui *et al.,* 1995; Brizzi *et al.,* 1996). Although we can for the moment only speculate on the role of ZNF162 phosphorylation in the absence of direct experimental data, it is of interest that ZNF162 belongs to this emerging family of potential signal transduction molecules.

ACKNOWLEDGMENT

We thank Dr. K. Artzt, University of Texas at Austin, Texas, for critical suggestions and advice. We also thank Professor A. Mantovani and Dr. E. Garattini for encouragement and critical discussion of this project.

REFERENCES

Amson, R. B., Nemanj, M., Roperch, J. P., Israeli, D., Bougueleret, L., LeGall, I., Medhioub, M., Linares-Cruz, G., Lethrosne, F., Pasturaud, P., Piouffre, L., Prieur, S., Susini, L., Alvaro, V., Millasseau, P., Guidicelli, C., Bui, H., Massart, C., Cazes, L., Dufour, F., Bruzzoni-Giovanelli, H., Owadis, H., Hennion, C., Charpak, G., Dausset, J., Calvo, F., Oren, M., Cohen, D., and Telerman, A. (1996). Isolation of 10 differentially expressed cDNAs in p53-induced apoptosis: Activation of the vertebrate homologue of the Drosophila seven in absentia gene. *Proc. Natl. Acad. Sci. USA* **93:** 3953–3957.

Arning, S., Gruter, P., Bilbe, G., and Kramer, A. (1996). Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* **2:** 794–810.

Bairoch, A. (1991). PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Res.* **19:** 2241–2245.

Berger, S. L., and Kimmel, A. R. (1987). Preparation of cDNA and the generation of cDNA libraries: Overview. *Methods Enzymol.* **152:** 307–325.

Breviario, F., d'Aniello, E. M., Golay, J., Peri, G., Bottazzi, B., Bairoch, A., Saccone, S., Marzella, R., Predazzi, V., Rocchi, M., Della Valle, G., Dejana, E., Mantovani, A., and Introna, M. (1992). Interleukin-1-inducible genes in endothelial cells. *J. Biol. Chem.* **267:** 22190–22197.

Brizzi, M. F., Aronica, M. G., Rosso, A., Dentelli, P., Yarden, Y., and Pegoraro, L. (1996). Granulocyte–macrophage colony stimulating factor (GM-CSF) stimulates JAK2 signaling pathway and rapidly activates p93fes, STAT1 p91 and STAT3 p92 in polymorphonuclear leukocytes. *J. Biol. Chem.* **271:** 3562–3567.

Cruz-Alvarez, M., and Pellicer, A. (1987). Cloning of a full length complementary DNA for an *Artemia salina* glycine rich protein: Structural relationship with RNA binding proteins. *J. Biol. Chem.* **262:** 13377–13380.

Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. (1979). "Atlas of Protein Sequences and Structure" (M. O. Dayhoff, Ed.), Vol. 5, p. 345, NBRF, Washington, DC.

Ebersole, T. A., Chen, Q., Justice, M. J., and Artzt, K. (1996). The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. *Nature Genet.* **12:** 260–265.

Feinberg, A. P., and Volgestein, B. (1983). A technique for radiolabeling DNA restriction nuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.

Fumagalli, S., Totty, N. F., Hsunn, J. J., and Courtneldge, S. A. (1994). A target for Src in mitosis. *Nature* **368:** 871–874.

Ihle, J. N., Witthuhn, B. A., Quelle, F. W., Yamamoto, K., Thierfelder, W. E., Kreider, B., and Silvennoinen, O. (1994). Signaling by the cytokine receptor superfamily: JAKs and STATs. *Trends Biochem. Sci.* **19:** 222–227.

Jones, A. R., and Schedl, T. (1995). Mutations in gld-1, a female germ cell-specific tumor suppressor gene in *Caenorabditis elegans,* affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9:** 1491–1504.

Kanakura, Y., Druker, B., Cannistra, S. A., Furukawa, Y., Torimoto, Y., and Griffin, J. D. (1990). Signal transduction of the human granulocyte–macrophage colony-stimulating factor and interleukin-3 receptors involves tyrosine phosphorylation of a common set of cytoplasmic proteins. *Blood* **76:** 706–715.

Lagercrantz, J., Larsson, C., Grimmond, S., Skogseid, B., Gobl, A., Friedman, E., Carson, E., Phelan, C., Oberg, K., Nordenskjold, M., Hayward, N. K., and Weber, G. (1995). Candidate genes for multiple endocrine neoplasia type 1. *J. Intern. Med.* **238:** 245–248.

Lock, P., Fumagalli, S., Polakis, P., McCornick, F., and Courtneidge, S. A. (1996). The human p62 cDNA encodes Sam 68 and not the RasGAP-associated p62 protein. *Cell* **84:** 23–24.

Matsuguchi, T., Inhorn, R., Carlesso, N., Xu, G., Druker, B., and Griffin, J. D. (1995). Tyrosine phosphorylation of p95^{Vav} in myeloid

cells is regulated by GM-CSF, IL-3, and steel factor and is constitutively increased by $p210^{BCR/ABL}$. *EMBO J.* **14:** 257–265.

Matsuguchi, T., Salgia, R., Hallek, M., Eder, M., Druker, B., Ernst, T., and Griffin, J. D. (1994). Shc phosphorylation in myeloid cells is regulated by granulocyte–macrophage colony stimulating factor, interleukin-3, and steel factor and is constitutively increased by $p210^{BCR/ABL}$. *J. Biol. Chem.* **269:** 5016–5021.

Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1988). Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.* **8:** 2237–2241.

Metcalf, D. (1989). The molecular control of cell division, differentiation commitment and maturation in haematopoietic cells. *Nature* **339:** 27–30.

Moscinski, L. C., and Prystowsky, M. B. (1990). Identification of a series of differentiation-associated gene sequences from GM-CSF stimulated bone marrow. *Oncogene* **5:** 31–37.

Mui, A. L. F., Wakao, H., O'Farrel, A. M., Harada, M., and Miyajima, A. (1995). Interleukin-3, granulocyte-macrophage colony stimulating factor and interleukin-5 transduce signals through two STAT5 homologs. *EMBO J.* **14:** 1166–1175.

Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M. A., Nilges, M., Gibson, T. J., and Pastore, A. (1996). Three-dimensional structure and stability of the KH domain: Molecular insights into the fragile X syndrome. *Cell* **85:** 237–245.

Ogawa, M. (1993). Differentiation and proliferation of hematopoietic stem cells. *Blood* **81:** 2844–2853.

Oval, J., and Taetle, R. (1990). Factor dependent human leukemia cell lines: New models for regulation of acute non lymphocytic leukemia cell growth and differentiation. *Blood Rev.* **4:** 270–279.

Pang, J. T., and Thakker, R. V. (1994). Multiple endocrine neoplasia type 1 (MEN1). *Eur. J. Cancer* **30A:** 1961–1968.

Quelle, F. W., Sato, N., Witthuhn, B. A., Inhorn, R. C., Eder, M., Miyajima, A., Griffin, J. D., and Ihle, J. N. (1994). JAK2 associates with the beta c chain of the receptor for granulocyte–macrophage colony-stimulating factor, and its activation requires the membrane proximal region. *Mol. Cel. Biol.* **14:** 4335–4341.

Rambaldi, A., Bettoni, S., Tosi, S., Giudici, G., Schirò, R., Borleri, G. M., Abbate, M., Chiaffarino, F., Colotta, F., Barbui, T., and Biondi, A. (1993). Establishment and characterization of a new granulocyte–macrophage colony-stimulating factor-dependent and interleukin-3-dependent human acute myeloid leukemia cell line (GF-D8). *Blood* **81:** 1376–1383.

Rambaldi, A., Viero, P., Bassan, R., Buelli, M., Rossi, A., Bellavita, P., Spinelli, O., Amaru, R., Micheletti, M., Borleri, G., Cortelazzo, S., Comotti, B., and Barbui, T. (1996). G-CSF-mobilized peripheral blood progenitor cells for allogeneic transplantation of resistant or relapsing acute leukemias. *Leukemia* **10:** 860–865.

Rambaldi, A., Young, D. C., and Griffin, J. D. (1987). Expression of the M-CSF (CSF-1) gene by human monocytes. *Blood* **69:** 1409–1413.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Siomi, H., Choi, M. C., Siomi, M. C., Nussbaum, R. L., and Dreyfuss, G. (1994). Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell* **77:** 33–39.

Siomi, H., Matunis, M. J., Michael, W. M., and Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* **21:** 1193–1198.

Swanson, M. S., and Dreyfuss, G. (1992). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA binding specificities. *Mol. Cell. Biol.* **12:** 164–171.

Tabor, S., and Richarson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84:** 4767–4771.

Toda, T., Iida, A., Miwa, T., Nakamura, Y., and Imai, T. (1994). Isolation and characterization of a novel gene encoding nuclear protein at locus (D11S636) tightly linked to multiple endocrine neoplasia type 1 (MEN1). *Hum. Mol. Genet.* **3:** 465–470.

Tso, J. Y., Sun, X. H., Kao, T. H., Reece, K. S., and Wu, R. (1985). Isolation and characterization of rat and human glyceraldehyde-3-phosphate dehydrogenase cDNAs: Genomic complexity and molecular evolution of the gene. *Nucleic Acids Res.* **13:** 2485–2502.

Van Seuningen, I., Ostrowski, J., Bustelo, X. R., Sleath, P. R., and Bomsztyk, K. (1995). The K protein domain that recruits the interleukin 1-responsive protein kinase lies adjacent to a cluster of c-Src and Vav SH3-binding sites. *J. Biol. Chem.* **270:** 26976–26985.

Williamson, M. P. (1994). The structure and function of proline-rich regions in proteins. *Biochem. J.* **297:** 249–260.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Dubin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, M., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohldman, P. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368:** 32–38.

Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M. F., Polakis, P., and McCormick, F. (1992). Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* **68:** 551–558.