Genomics **63,** 123–132 (2000)
doi:10.1006/geno.1999.6027, available online at http://www.idealibrary.com on IDEAL®

DXS6673E Encodes a Predominantly Nuclear Protein, and Its Mouse Ortholog *DXHXS6673E* Is Alternatively Spliced in a Developmental- and Tissue-Specific Manner

Maurice P. Scheer,* S. van der Maarel,† S. Kübart,* A. Schulz,* J. Wirth,* S. Schweiger,* H.-H. Ropers,*,‡ and H. G. Nothwang*,[1]

*Max Planck Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin–Dahlem, Germany; †Department of Human Genetics, University of Leiden, Leiden, The Netherlands; and ‡Department of Human Genetics, University Hospital, Nijmegen, The Netherlands

Received June 28, 1999; accepted October 14, 1999

DXS6673E **is a candidate gene for nonspecific X-linked mental retardation and encodes a novel Zn-finger protein. The ortholog murine gene *DXHXS6673E* in XC-D was isolated and characterized. It is ubiquitously expressed in all embryonic stages and adult tissues. Two different transcription start sites exist that result in two major transcripts of 6055 and 5352 nucleotides, each composed of 25 exons. Exon 1A is tissue specific, whereas exon 1B is transcribed constitutively. Both variants are translated into the same 1370-amino-acid protein. Transcripts are subject to alternative splicing at the 5′-end. Some of the isoforms are developmental stage and tissue specific. Among them, one was present only in embryos and adult brain. Sequence analysis demonstrated evolutionary conservation down to the arthropods and defined several conserved protein motifs. Subcellular localization studies with green fluorescent protein as a reporter showed that DXS6673E is predominantly located in the nucleus due to several functional nuclear localization signals. Three distinct protein distribution patterns in COS-7 cells could be identified.** © 2000 Academic Press

INTRODUCTION

Mental retardation (MR) is a frequent and very heterogeneous disorder that affects about 1–3% of individuals. Population studies indicate that X-linked genes account for approximately 20–50% of all cases (XMR). Recently, several genes for nonsyndromic XMR (NSXMR), i.e., MR without additional clinical features, have been isolated, *FMR2, oligophrenin-1, GDIα* and *PAK3,* where mutations have been found in independent patients (Gecz *et al.,* 1996; Billuart *et al.,* 1998; d'Adamo *et al.,* 1998; Allen *et al.,* 1998). Another candidate gene, *DXS6673E* in Xq13, was found to be disrupted by a balanced X;13 translocation in a mentally retarded female (van der Maarel *et al.,* 1996). This gene is composed of 26 exons with two different untranslated exons 1A and 1B. The X-chromosomal breakpoint is located in exon 1A, and sequences downstream of the breakpoint are still expressed from the derivative chromosome X. The gene is highly conserved among vertebrates and most abundantly expressed in brain. It encodes a 1358-amino-acid protein of unknown function. Sequence analysis revealed that it represents the first member of a new gene family, defined by a five-times repeated, novel, conserved Zn-finger-related motif of the general form $CX_2CX_{19-22}CX_3CX_{13-19}CX_2CX_{19-25}FCX_3CX_3F/Y$. It is unknown whether it mediates protein–DNA or protein–protein interactions (Xiao *et al.,* 1998; Smedley *et al.,* 1998; Reiter *et al.,* 1998).

Interestingly, the second member, *ZNF198,* was isolated by positional cloning in the study of another genetic disorder, the t(8;13) myeloproliferative syndrome (Xiao *et al.,* 1998; Smedley *et al.,* 1998; Popovici *et al.,* 1998). The reciprocal translocation fuses *ZNF198* to the fibroblast growth factor receptor-1 and thereby causes lymphoblastic lymphoma. The disorder generally progresses to full-blown acute myelogeneous leukemia within a year of diagnosis. Another member, *KIAA0425,* was identified by a database search (Reiter *et al.,* 1998).

To gain further insight into the function of this gene family, we isolated and characterized *DXHXS6673E* (approved MGD nomenclature), the murine ortholog of *DXS6673E*. This allowed expression studies in various developmental stages and adult tissues. Evolutionary conservation was assessed by searching publicly available databases. To test putative nuclear addresses, subcellular localization studies of DXS6673E–green fluorescent protein (GFP) constructs were performed.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF156605.

[1] To whom correspondence should be addressed. Telephone: +49 (30) 8413 1250. Fax: +49 (30) 8413 13831. E-mail: nothwang@molgen.mpg.de.

0888-7543/00 $35.00
Copyright © 2000 by Academic Press
All rights of reproduction in any form reserved.

MATERIALS AND METHODS

Library screening and sequence analysis. A Lambda ZAP II mouse brain cDNA library (Stratagene) was screened, and Bluescript KS(+) plasmids were isolated by *in vivo* excision according to the manufacturer's instructions. To obtain genomic clones, the P1 library 703 (German Resource Centre) containing genomic mouse DNA inserts was screened. Double-stranded sequencing of cDNA clones was performed with labeled universal M13 primers and the Thermo sequenase fluorescence-labeled primer cycle sequencing kit (ABI) on a Licor sequencer (MWG-Biotech). To generate the genomic sequence at the exon–intron boundaries, P1 clones were digested with *Hin*dIII and subsequently shotgun-cloned into pBluescript. Selected subclones were sequenced either with M13 or with labeled exon-specific oligonucleotides. Some of the boundaries were characterized by genomic long-range PCR. Sequence analysis was performed with the GCG package (Devereux *et al.,* 1984).

RNA isolation and RT-PCR analysis. Total mouse embryo RNA was isolated by the guanidinium–thiocyanate method (Chomczynski and Sacchi, 1987). Day 1 was defined as day of vaginal plug, and embryos were prepared at 9, 12, 14, 16, and 18 days of gestation. RNA from tissues was isolated using the RNAeasy kit (Qiagen), and poly(A)$^+$ RNA was obtained by subsequent purification on magnetic beads using a mRNA isolation kit (Boehringer Mannheim). For synthesis of cDNA, approximately 40–200 ng of purified mRNA or 100–500 ng of total RNA was randomly primed by using SuperScript II RNase H- reverse transcriptase (Gibco BRL). Reverse transcription was carried out for 40 min at 42°C, 10 min at 45°C, and 10 min at 50°C. Negative controls without reverse transcriptase were run under identical conditions. PCR amplification with *DXHXS6673E* gene-specific primers was performed on 1/50 of the original RT reaction.

Nested PCR was carried out (96°C for 30 s, 60°C for 30 s, and 72°C for 2 min) in a final volume of 50 μl containing 40 pmol of each primer and 1 U of *Taq* polymerase. Nested primer pairs were as follows: first PCR, forward primer, TTGAATGAGAGGAGGGGATC, and reverse primer, GCTGGTCTTGCCATTACGAT; nested PCR, CTTGAGCCAGGTAAGGAAAG and ACACTGGTCACAACAGTTGG (exon 1A to exon 7); first PCR, CCG TTG CTG TCT GCA GTT CT and GCTGGTCTTGCCATTACGAT; nested PCR, CTT GTA GCGGTT ACT CTA AC and ACACTGGTCACAACAGTTGG (exon 1B to exon 7); first PCR, ACTCCAAGAAGCCCTTGGGC and CACGCTGGCAG-GTTTGGGA; nested PCR, GAA AGA CCT GTA CCT TCT GC and TTG CTT GTA CAG CAG TAC AC (exon 5 to exon 12); first PCR, CTGCCGGCAGGAAAAGCTCC and AAT CAG CCT CCA AGT CTT GC; nested PCR, CAC GAG AAG CTT CGA TTC AG and CAT CTA AGA CAT TAG CCA TC (exon 11 to exon 18); and first PCR, GCA GAG GGA CTC CTG GAA GA and CAATGCAGGAGCGCTCAGG; nested PCR, GGA CAG CTC GGG ATG ATG TC and AAC ACA TCA TTT CGA GTT CG (exon 18 to exon 25). One-fifth of the PCR product was analyzed on 1.5% agarose gels with 0.5 μg/ml ethidium bromide in 1× TAE.

Northern blot analysis. A Northern blot of RNA from embryos of stage 9, 12, 14, 16, and 18 days of gestation was prepared according to standard protocols (Sambrook *et al.,* 1989), and an adult multiple tissue Northern blot was purchased from Clontech. Both blots were hybridized in express hybridization solution (Clontech) with a ^{32}P-labeled cDNA insert for 90 min at 65°C. Hybridized filters were washed for 2× 30 min in 2× SSC, 0.1% SDS, and 30 min in 0.1× SSC, 0.1% SDS at 65°C. Autoradiography took 16–40 h at −70°C using two intensifying screens.

GFP constructs and site-directed mutagenesis. The ORF of DXS6673E was amplified in two overlapping fragments from cDNA clones 50:17 and 50:33 (van der Maarel *et al.,* 1996) by using the proofreading *Pfu* polymerase (Stratagene). To analyze both N- and C-terminal fusion of DXS6673E to GFP, fragments were ligated into the green fluorescent protein vectors peGFP-N3 and peGFP-C1 (Clontech) using primers with the respective restriction site for amplification. Site-directed mutagenesis was performed using the QuickChange kit (Stratagene). Individual clones were sequenced to verify their insert, and Western blot analyses were used to assess full-length translation of fusion constructs. DNA from two independent clones for each construct was purified using the Qiagen plasmid purification kit (Qiagen). To construct C-terminal truncated DXS6673E fusion proteins, fragments were amplified using primers with additional restriction enzyme sites. Constructs Δ655–1358 and Δ474–1358 contain amino acid residues (aa) 1–654 and 1–473, respectively.

Cell culture and transfection experiments. COS-7 cells were maintained at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium. For fluorescence microscopy analysis, cells were grown directly on untreated glass coverslips. In a typical transfection experiment, 3 μg of plasmid DNA was delivered into COS-7 cells using 3 μl of Lipofectace transfection reagent (Gibco BRL). Cells were grown to about 60% confluence, exposed to lipid–DNA complexes for 5 h, and grown for an additional 24 to 72 h until analyzed. Two independent clones per construct were used.

Fluoresence microscopy. Expression of GFP was analyzed either immediately after incubation in living cells or in formaldehyde-fixed cells. For formaldehyde fixation, cells were washed once in phosphate-buffered saline (PBS), once in PBS with 2% paraformaldehyde, and three times again in PBS, 0.1% Triton, dried, and mounted onto a microscope slide in DAPI (4,6-diamidino-2-phenylindol dilactate; Boehringer Mannheim) antifade mounting solution (0.4 μg/ml DAPI in a solution consisting of 9 parts of glycerol containing 2% 1,4-diazobicyclo-(2-2-2)-octane and one part 0.2 M Tris–HCl, pH 7.5). Alternatively, the cells were briefly washed in PBS and mounted onto a microscope slide in PBS. To prevent cells from drying, the coverslips were sealed to the microscopic slides with rubber cement. Green fluorescence staining and DAPI nuclear staining were analyzed with an epifluorescence microscope (Axioskop 50; Zeiss). Photos were taken with a CCD camera. For each clone, 200–300 cells were visually scored.

Subcellular fractionation and Western blot analysis. After transfection, cells were transferred to 150-ml plastic bottles and grown for 24–72 h. Thereafter, cells were washed with ice-cold PBS, scraped from the dishes and pelleted. The pelleted cells were washed twice with isotonic buffer (40 mM Hepes, pH 7.4, 0.32 M sucrose, 1% (v/v) β-mercaptoethanol, and protease inhibitors (Boehringer Mannheim) and then resuspended in the same buffer. The cells were disrupted by dounce homogenization (10 strokes) in 2 ml of the buffer. The nuclei were separated from the cytoplasm by low-speed centrifugation (1000*g* for 10 min), and both fractions were collected. Nuclei were then resuspended in 2 ml.

For Western blot analysis, 150 μl of the nuclear fraction and 150 μl of the cytoplasmic fraction were acetone-precipitated and separated on an 8% SDS–polyacrylamide gel (Bio-Rad) and electroblotted (Bio-Rad) onto Hybond-N nitrocellulose membranes (Amersham Pharmacia). After being blocked with 5% defatted dried milk solution, the filters were incubated with a 1:1000 dilution of polyclonal antisera against GFP (Clontech) for 1 h at room temperature. Subsequently, the blots were incubated with horseradish peroxidase-conjugated anti-rabbit Fab fragments (Amersham Pharmacia) at a 1:2000 dilution and developed with enhanced chemiluminescence reagents (Amersham Pharmacia) as specified by the manufacturer.

RESULTS

Isolation of DXHXS6673E

To isolate the murine ortholog of *DXS6673E,* a mouse fetal brain cDNA library was screened with several human subclones. Seventeen cDNA clones that showed significant sequence similarity were isolated. The composite cDNA corresponds to exons 1b-25 of the human gene (data not shown). No clone similar to the human alternative first exon 1A could be isolated by

library screening or by RT-PCR using primers derived from the human sequence. Partial sequencing of P1 clone ICRFP703J20233, which contained a large part of *DXHXS6673E* (see below), however, resulted in a sequence similar to human exon 1A, and by RT-PCR, a 703-bp product could be amplified from various mouse tissues (see below). To check whether exon 1A and 1B represent two alternative transcription initiation sites, RNA populations from various developmental stages (12, 14, 16, and 18 days of gestation) and adult tissues (brain, kidney, liver, placenta, testis, and lung) were screened by nested RT-PCR for the presence of a transcript containing exons 1A and 1B. No product was obtained using forward primers in exon 1A and reverse primers in exon 1B. Thus, two alternative untranslated first exons exist in mouse that give rise to two different transcripts of 6055 (variant 1A) and 5352 (variant 1B) nucleotides. Both transcripts encode a protein of 1370 amino acid residues (Fig. 1A).

To study the genomic organization of *DXHXS6673E*, RT-PCR products were hybridized to a P1 mouse genomic library. Two positive clones were identified: ICRF703I2291 and ICRF703J20233. Sequence information derived from these clones together with long-range PCR allowed the determination of all exon–intron boundaries. *DXHXS6673E* consists of 26 exons with the initiation codon AUG in exon 2 as its human counterpart. All splice donor and splice acceptor sites conform to the AG/GT rule. Most exons are in the range between 50 and 300 nucleotides. Exceptions are exons 1A, 2, and 25 with 703, 680, and 1419 nucleotides. Cytogenetic analysis of PAC clone ICRF703J20233 mapped the gene to chromosome XC-D (data not shown). This assignment corresponds well with the localization of human *DXS6673E* in Xq13.

Sequence Analysis

A detailed comparative sequence analysis at both the nucleotide and the amino acid levels between *DXS6673E* and its murine ortholog revealed sequence identities higher than 90%. Interestingly, both genes contain a dinucleotide repeat in exon 1A, but it is composed of (GA) in human and (GT) in mouse (data not shown). At the amino acid level, only two significant differences were observed at positions 238–239 and 797–807, where the published human protein lacks 2 (VQ) and 11 (RTPDENGNLGK) amino acid residues, respectively. RT-PCR experiments and a database search, however, showed that these differences are due to alternative splicing (see below).

Analysis of the amino acid composition revealed that the DXHXS6673E protein contains the Zn-finger motifs characteristic for members of this novel family, too (aa 313–387, 411–486, 500–576, 593–662, 681–744) (Smedley *et al.*, 1998; Reiter *et al.*, 1998) (Fig. 1A). In addition, five nuclear localization signals (NLSs) of the SV40 large T antigen type at positions 234–241 (PPERKRSE), 283–289 (PFRPRRS), 827–833 (PRKNKAA), 1261–1267 (PVRQRKG), and 1272–1278 (PGKRKRE), one bipartite NLS between amino acid residues 479 and 493 (KRFCNTTCLGAYKKKNT), and three putative tyrosine phosphorylation sites at positions 543–551 (RSLSDPCYY), 1130–1138 (RYEPDSIYY), and 1312–1319 (RTRNDVFY) are conserved between mouse and human (Fig. 1A). The phosphorylation sites are close to some NLSs and may modulate the nuclear import as described for several proteins (Dingwall and Laskey, 1998). Finally, there is a proline-rich sequence at position 814–828 that contains two putative SH3-binding motifs PxxP (PSVPTPPPPPPPATP) (Fig. 1A).

A search for further family members yielded, in addition to the known members *ZNF198* and *KIAA0425*, the human gene *ZNF258*, which displays sequence similarity over a long range (39% identity) (Fig. 1A). However, the encoded protein lacks some cysteines in the C-terminal part of the second, and the N-terminal part of the third, Zn-finger motif (Fig. 1A). This defines the novel Zn-fingers as alternating motifs of types $CX_2CX_{19-22}CX_3C$ and $CX_2CX_{19-25}FCX_3CX_3F/Y$. Sequence similarity in the entire gene family is highest at the C-terminal part with several conserved amino acid sequences, such as two regions with interspersed prolines (PI/vPVPI/VF/yV/iPV/iP (aa 869–879) and PVPF/tXXPV/IPV/mPVPM/vF/lL/iP (aa 888–903)).

Several ESTs from rat (GenBank Accession Nos. AI111670, AI412415, AI409797, AI112706, AI502308), rabbit (C83185, C83052), turkey (AF062408), zebrafish (AI496701, AI397030), AI397051), and fruitfly (AA951966, AA567252, AA543240, AA949953) shared significant sequence similarity with *DXHXS6673E* (Blastp *E* values between 3×10^{-39} and 4×10^{-6}; data not shown) and likely represent homologous genes in the respective species. Although these ESTs provided limited sequence information, a detailed analysis revealed the conservation of the putative tyrosine phosphorylation site at position 1312–1319 down to the arthropod *Drosophila melanogaster*, indicating its functional importance (Fig. 1B). No similarity was found to the almost completely sequenced genome of the nematode *Caenorhabditis elegans*.

Expression Analysis

To study the expression of *DXHXS6673E*, Northern blot analyses were performed. The gene is expressed in all developmental stages examined, i.e., 9, 12, 14 16, and 18 days of gestation (Fig. 2). It is also ubiquitously, but to a variable degree, expressed in all adult tissues analyzed. Strongest expression was observed in adult brain and testis (Fig. 2). One major transcript of about 5600 nucleotides was detected. This transcript size corresponds well with the length of variant 1B (5352 nt), taking into account a poly(A)$^+$ tail of about 200–250 nt. Some weaker signals were also obtained in the lower and higher molecular weight ranges. To analyze whether some of these fainter signals represented al-

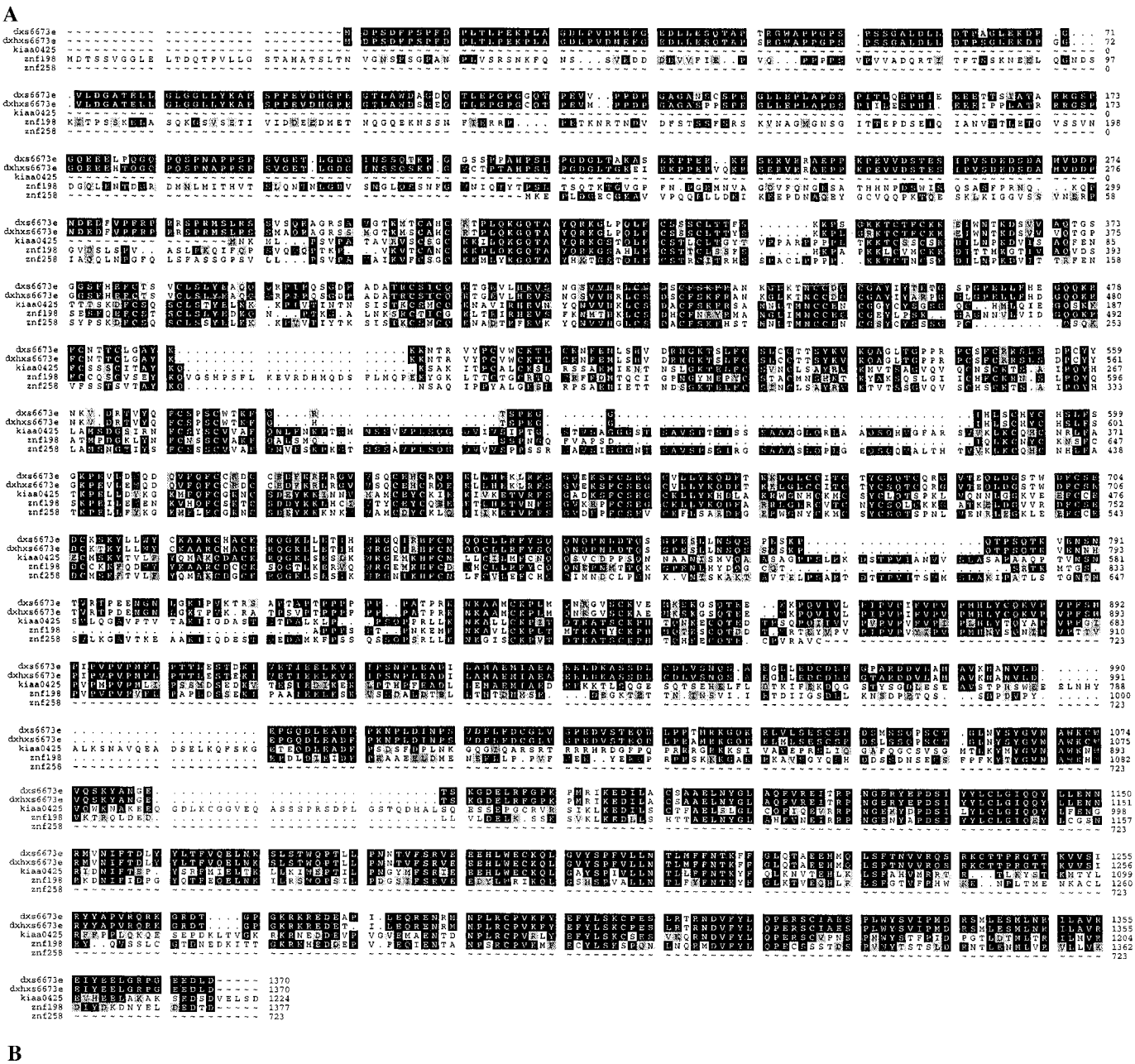

FIG. 1. Sequence alignment of members of the novel gene family. Amino acid sequences were compared using the Pileup program. Conserved amino acid residues are shaded. Numbers at the right indicate amino acid position (**A**) Significant sequence similarity over the entire length in mammals. DXS6673E (GenBank Accession No. X95808, a splice variant with 11 additional amino acids at position 797–807, was used), DXHXS6673E (GenBank Accession No. AF156605), KIAA0425 (GenBank Accession No. AB007885), ZNF198 (GenBank Accession No. AJ224901), and ZNF258 (GenBank Accession No. AF055470). The percentage identities between DXS6673E and other related proteins are 96% (DXHXS6673E), 47% (KIAA0425), 43% (ZNF198), and 37% (ZNF258). (**B**) Evolutionary conservation of a putative tyrosine phosphorylation site. Rn, *Rattus norvegicus* (GenBank Accession No. AI111670); Dr, *Danio rerio* (GenBank Accession No. AI397030); and Dm, *Drosophila melanogaster* (GenBank Accession No. AA567252). The percentage identities between the tyrosine phosphorylation site in DXS6673E and other related proteins are 100% (DXHXS6673E), 65% (KIAA0425), 68% (ZNF198), 65% (Rntps), 52% (Drtps), and 58% (Dmtps).

ternative splice variants or correspond to the 1A variant, RT-PCR experiments were performed on RNA populations from various developmental stages (12, 14, 16, and 18 days of gestation) and adult tissues (brain, kidney, liver, placenta, testis, and lung). To detect minor size variations, overlapping fragments between 0.9

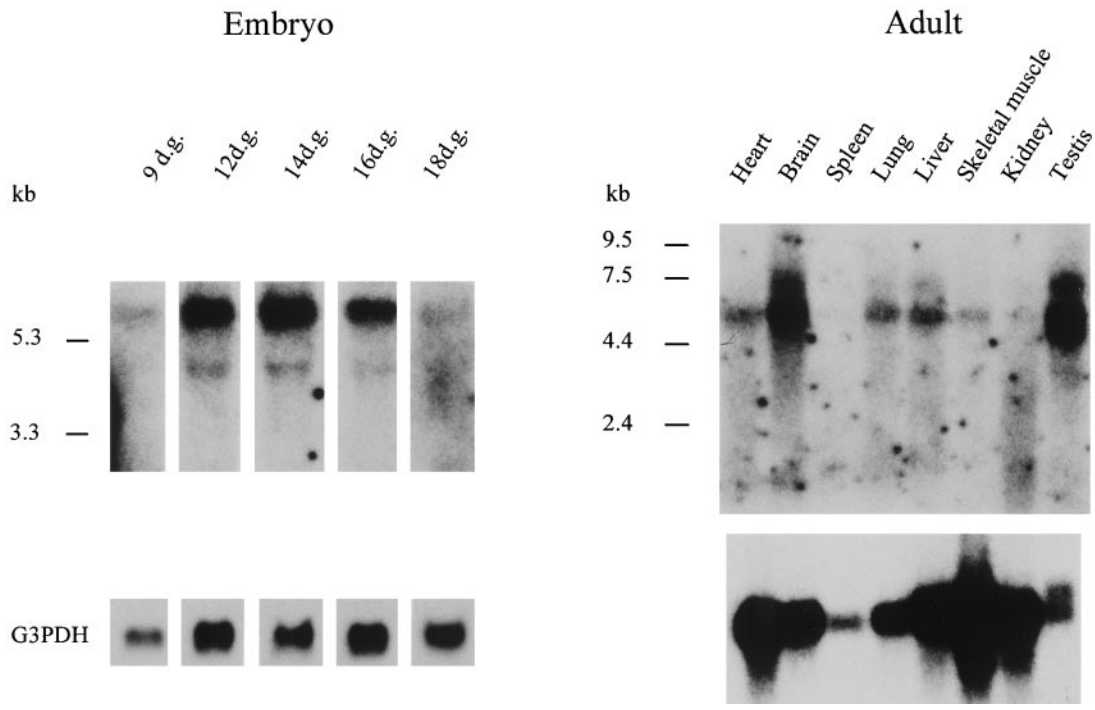

FIG. 2. Northern blot analysis. Hybridization of nucleotides 3385–6173 of *DXHXS6673E* to a fetal (total RNA) and adult (poly(A)$^+$ RNA) Northern blot from several developmental stages or tissues. Age of embryo (days of gestation (d.g.) with day 1 as day of vaginal plug) and tissues are indicated. As a control, glyceraldehyde 3-phosphate dehydrogenase (G3PDH) cDNA was hybridized to the Northern blots.

and 1.4 kb, starting from exon 1A to exon 25, were amplified. Products of unpredicted size were sequenced. Extensive alternative splicing was observed at the 5′-end of both 1A and 1B variants (Fig. 3). Exons 2 and 5 could be spliced out completely. In addition, exons 2 and 3 showed alternative internal donor splice site usage that removed 522 and 6 nt, respectively. The alternative splice site usage in exon 3 reconstitutes a NLS (aa 234–241) (Fig. 1). For 1B variants, the observed isoforms occurred, with the exception of 1BΔ2Δ3iΔ5 and 1BΔ2iΔ5, in all stages of development and tissues analyzed (Fig. 3B). The dominant isoform corresponded to the full-length 1B transcript. Isoforms lacking exon 5 (Δ5), or 522 nt of exon 2 (Δ2i), are three- to fourfold less abundant, and exon 2 is completely absent in few transcripts. In contrast, 1A transcripts are more often alternatively spliced in a developmental stage- and tissue-specific manner (Fig. 3A). In most RNA populations, the 5′ full-length 1A variant represented the most prominent isoform. However, exon 1A was not transcribed in liver, and in lung and placenta, only the 5′ full-length 1A variant or an isoform without exon 2 was present, respectively. Isoforms 1AΔ2i and 1AΔ2iΔ3i, derived from the internal donor splice site in exons 2 and 3, were detected only in embryos and in adult brain tissue, and splice variant 1AΔ2Δ5 was detected in embryos at 14 days of gestation.

Between exons 6 and 25, only exon 23 was found to be alternatively spliced. This isoform was observed as a minor transcript in all RNA populations tested (data not shown).

Open reading frame (ORF) analysis of the isoforms revealed that they fell into three classes: (i) In isoforms due to internal donor site usage in exon 2 and/or exon 3, as represented by brain-specific transcripts 1AΔ2i and 1AΔ2iΔ3i, the ORF is maintained. These variants lack amino acid residues 50 to 223 and/or 238 and 239. The ORF is also maintained in exon 23 isoforms. (ii) In splice variants without exon 5 (Δ5), a shift in the ORF occurs and introduces a stop codon in exon 7 at position 86–88. These isoforms are translated into a shorter polypeptide containing an alternative carboxyl terminus of 87 novel amino acid residues. (iii) Splice variants without the entire exon 2 lack the original initiation codon AUG at position 717. The next start codon that fits the Kozak consensus sequence is situated in exon 7 in a different ORF and would give rise to a short polypeptide of 22 amino acid residues (MAAWYTDSAAILASPNSEPTRD). This peptide is identical to the C-terminus of Δ5 splice variants with the initiation codon in exon 2. However, the possibility is not excluded that Δ2 isoforms encode a truncated form of *DXHXS6673E* by a leaky ribosome scanning mechanism or by a non-AUG translational initiation as reported for the c-myc gene (Hann *et al.*, 1988).

Subcellular Localization of the DXS6673E Encoded Protein

The presence of several putative nuclear targeting sequences in DXS6673E and its mouse ortholog suggested a nuclear localization of the proteins. To study their functional significance, fusion constructs with the GFP as reporter were analyzed. Using fluorescence

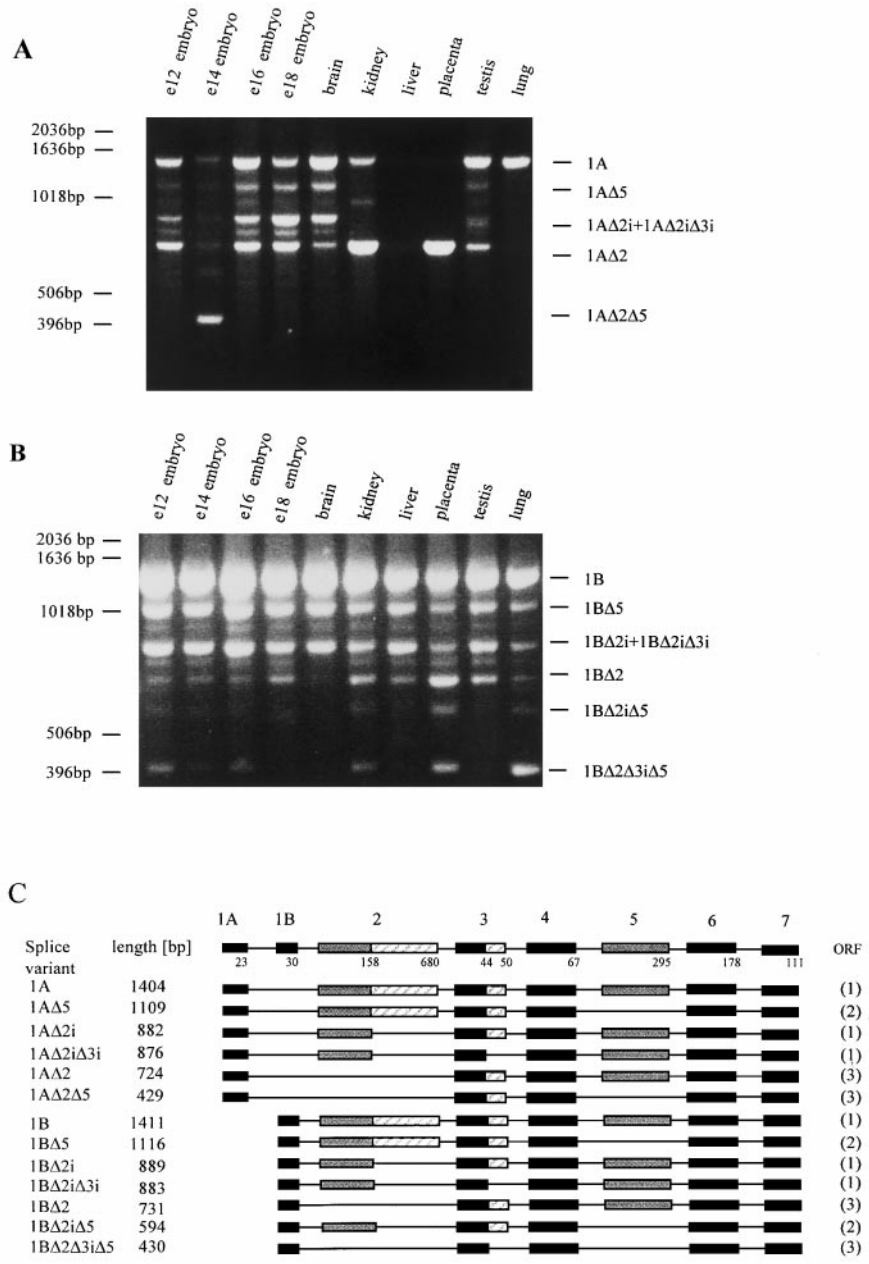

FIG. 3. Alternative splicing in mouse embryos and adult tissues. Nested RT-PCR using forward primers in exon 1A and reverse primers in exon 7 (**A**) or exon 1B and exon 7 (**B**) was performed on various RNA populations. Cloned and sequenced products are indicated and correspond to splice variants listed in (**C**). Age of embryo (day of gestation) and adult tissues, used to isolate the respective RNA population, are shown above the lanes. (**C**) Schematic representation of splice variants at the 5′-end of *DXHXS6673E*. Names of splice variants and their sizes are given to the left. Constitutive exons are in black, alternatively spliced exons are in gray, and alternative donor splice sites are striped. Numbers above the exons indicate exon number; numbers below the exons indicate nucleotides within an exon. Exons 1A, 1B, and 7 are not full-length due to internal PCR primers. (1) ORF maintained, (2) novel C-terminus of 87 amino acid residues, (3) alternative initiation codon.

microscopy, transient expression of fusion proteins in COS-7 cells revealed a dual localization in the cytoplasm and in the nucleus, with exclusion of the nucleoli (Figs. 4A–4C). The distribution between these two compartments, however, varied within the cell population. In the majority of cells (47%), fluorescence was strongest in the nucleus (Figs. 4A and 5). In one quarter, even distribution of the fluoresence was observed in the cell with a layer surrounding the nuclear envelope (Figs. 4B and 5), and in the remaining quarter, fluorescence was stronger in the cytoplasm, again mostly with a layer surrounding the nuclear envelope (Figs. 4C and 5). This distribution pattern did not vary over time (24, 48, and 72 h). Furthermore, in the DXS6673E constructs used (see also below), it was independent of the position of GFP. The only difference was a stronger fluorescence signal with N-terminal GFP.

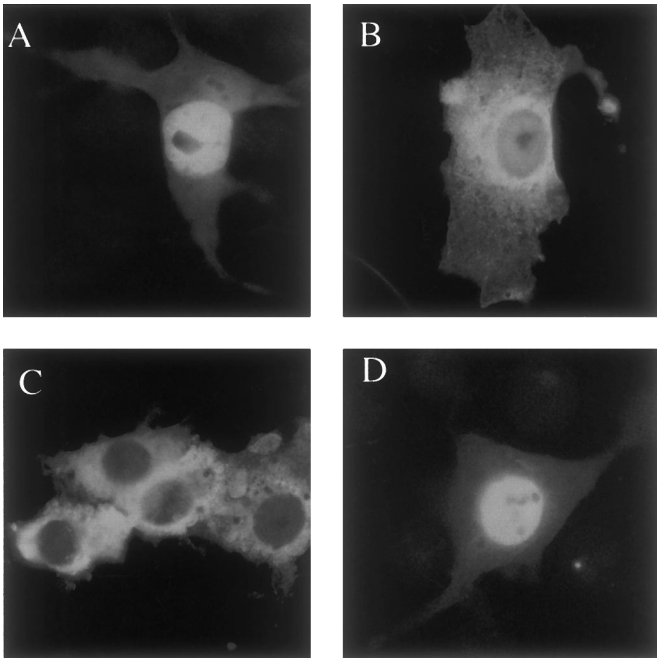

FIG. 4. Subcellular localization of GFP–DXS6673E fusion proteins in COS-7 cells. Different parts of the DXS6673E protein were fused to peGFP-C1 and analyzed 24 h after transfection into COS-7 cells for their subcellular localization. Cell type with predominant nuclear localization of the full-length fusion protein (**A**), cell type with even distribution of the full-length fusion protein (**B**), and cell type with predominant cytoplasmic localization and a strong perinuclear staining of the full-length fusion protein (**C**). Nuclear localization of truncated fusion protein Δ474–1358 (**D**). Subcellular localization was inspected 24 h after transfection in formaldehyde-fixed cells.

By comparison, GFP alone distributed in all cells analyzed throughout the nucleus and cytoplasm, since GFP (M_r 28,000) is able to diffuse passively into and out of the nucleus in mammalian cells (Ohno *et al.*, 1998). As previously shown, staining was slightly stronger in the nucleus than in the cytoplasm (Lightfoot *et al.*, 1999). This was also confirmed by subcellular fractionation and Western blot analysis (data not shown).

To quantify these results, subcellular fractionation was performed with subsequent Western blot analysis using an anti-GFP antibody. Most fusion protein was isolated with the nuclear fraction, whereas in the cytoplasmic supernatant, only a weak signal was obtained (Fig. 6). This result confirmed the predominant nuclear localization of the protein. It also supported the conceptual translation of *DXS6673E* into 1358 amino acid residues, since the observed molecular mass of the fusion protein is about 140–150 kDa higher than the molecular mass of the green fluorescent protein (M_r 28,000) alone.

To analyze which of the NLSs are important for nuclear localization, we substituted the basic amino acid residues lysine and arginine in the first part of the bipartite NLS by two glycines, since they were shown to be important for nuclear targeting (Robbins *et al.*, 1991). Fluorescence and subcellular fractionation studies in COS-7 cells, however, yielded no qualitative or quantitative differences in subcellular localization compared to wildtype (Fig. 5). To confirm that the bipartite NLS is indeed not necessary for nuclear transport, a truncated fusion protein was constructed containing the first 473 aa with two NLSs of the SV40 large T antigen type (construct Δ474–1358) (Fig. 5). This fusion protein was predominantly nuclear in all cells (Figs. 4D and 5). Hence, these NLSs are functional and can direct the transport of the DXS6673E protein into the nucleus. Interestingly, no cell types with a predominant cytoplasmic or even localization were observed (Fig. 5). To find out which of the amino acid residues are important for cytoplasmic localization, a further construct was analyzed containing the first 654 amino acids (construct Δ655–1358). Transfection yielded again three cell types with distinct protein distribution patterns. The ratio between these cell types, however, was altered (68:18:14) compared to the full-length protein (47:25:28) (Fig. 5). Nevertheless, this indicates that next to the NLSs, amino acid resi-

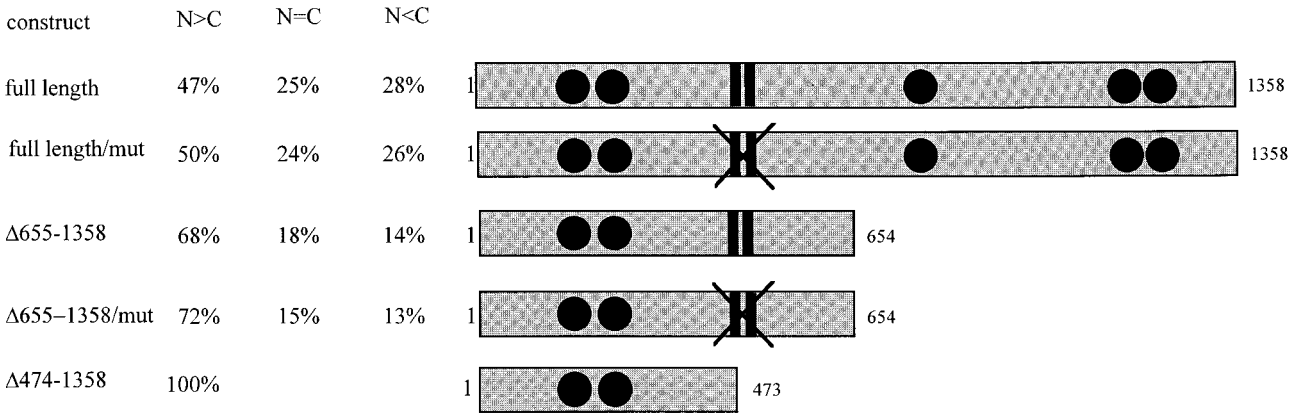

FIG. 5. Subcellular distribution types of DXS6673E fusion protein isoforms. Cell types were quantified by visually scoring 200–300 clones/construct. N > C, predominantly nuclear localization; N = C, even staining in both cell compartments; N < C, predominantly cytoplasmic localization. Black circles represent the NLS of the SV40 large T antigen type, black bars represent the bipartite NLS, and crossed lines represent the mutated bipartite NLS.

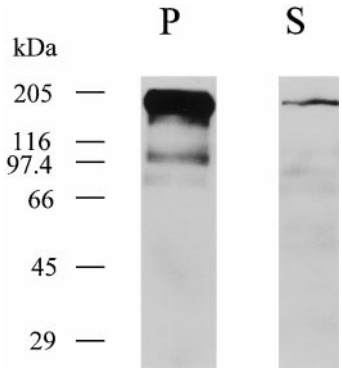

FIG. 6. Subcellular fractionation and immunoblot analysis. COS-7 cells were transfected with the pEGFPC1–DXS6673E fusion construct and harvested after 24 h. Cells were lysed, and nuclei were separated from the cytosol by differential centrifugation. Equal volumes of both nuclear (**P**) and cytosolic (**S**) fractions were separated on a SDS–polyacrylamide gel, and proteins were electroblotted onto nitrocellulose. Blots were incubated first with anti-GFP antiserum and subsequently with horseradish peroxidase-conjugated anti-rabbit Fab fragments.

dues 474–654 contain important information to guide subcellular distribution of the DXS6673E protein.

DISCUSSION

The mouse gene *DXHXS6673E,* isolated and characterized in this study, belongs to a novel, evolutionarily conserved gene family defined by a characteristic zinc-finger motif. Two of the members, *DXS6673E* and *ZNF198,* are likely involved in human diseases, but nothing is known about the function of the encoded proteins (van der Maarel *et al.*, 1996; Xiao *et al.*, 1998). In this regard, the cloning of the first mouse member will pave the way for functional analysis in an animal model, more amenable to genetic and molecular biology studies than human. Our identification of at least one family member in other model organisms like zebrafish and fruitfly will, in addition, help to assess basic gene functions. Surprisingly, no sequence similarity was detected in the almost completely sequenced genome of the nematode *C. elegans.* Thus, this novel gene family may represent another family not present in nematodes, like genes of the hedgehog signaling pathway (Ruvkun and Hobert, 1998).

Expression studies showed the presence of two alternative 5′UTRs that give rise to 1A and 1B *DSHXS6673E* variants. Whereas exon 1B was more prominent and constitutively expressed in all developmental stages and tissues examined, transcription of exon 1A showed tissue specificity and occurred at a low level according to Northern blot analysis. These results, therefore, provide experimental proof for previous speculations based on a CpG island extending from human intron 1A through exon 1B into intron 1B (van der Maarel *et al.*, 1996). From the presence of this CpG island, a ubiquitous expression of variant 1B and a tissue-specific expression of variant 1A were hypothesized. Two alternative promoters might be, in general, characteristic for this gene family, since two independent transcription start sites were also reported for *ZNF198* (Reiter *et al.*, 1998).

DSHXS6673E is also subject to highly alternative splicing at its 5′-end. The observed isoforms lead either to constitution of a NLS or to truncated proteins. Western blot analysis with DXS6673E-specific antibodies will show whether protein isoforms indeed exist or whether some of the characterized transcripts represent incorrectly spliced products. Several observations, however, argue against erroneously processed transcripts. (i) With the exception of exon 23, no further splice variants were observed in the other part of *DXHXS6673E,* but incorrect processing should *a priori* be random. (ii) RT-PCR of both total and poly(A)$^+$ RNA yielded the same proportion of splice variants (data not shown). (iii) Some of the splice variants were developmental stage or tissue specific. For example, alternative use of the exon 2 internal donor splice site in 1A transcripts occurred only in adult brain tissue, whereas the same splice site was alternatively used in 1B variants of all tissues examined. (iv) The ratio of alternative splicing of a given exon varied between 1A and 1B transcripts. Isoform Δ2 was a minor 1B, but a prominent 1A variant. (v) A splice variant of *ZNF198* was also described that disrupts the ORF between the N-terminus and the fourth Zn-finger motif by introducing a stop codon before Met 621 (Reiter *et al.*, 1998). Furthermore, exon 2 and exon 3 of *ZNF198* are variably incorporated into the mature transcript, and exon 4 contains an internal splice site (Kulkarni *et al.*, 1999). Thus, highly alternative splicing might be a general feature of this novel gene family.

Interestingly, two prominent splice variant classes are complementary: they either lack (Δ2i) or translate only into the N-terminus (Δ5). If *DXHXS6673E* functions as a transcription factor, this alternative processing might modulate its activity as reported for some proteins. A C-terminal truncated form of FosB inhibits Fos/Jun transcriptional activity, and a liver-enriched protein acts as either a transcriptional activator (LAP) or an inhibitor (LIP), depending on its alternative N-terminus (Nakabeppu and Nathans, 1991; Descombes and Schibler, 1991). LIP attenuates the transcriptional stimulation by LAP at substoichiometric amounts and is likely involved in fine-tuning of transcription factor activity (Descombes and Schibler, 1991). The alternatively spliced N-terminus of DXHXS6673E contains one putative NLS and several short proline-rich sequences, one of which occurs twice in the orientation PPSP and once in the opposite orientation, PSPP. Furthermore, the internal splicing in exon 2 disrupts another proline-rich sequence, WAPPGPSP, similar to a peptide previously shown to bind the SH3 domain (Pawson, 1995). It will, therefore, be interesting to see whether the N-terminus is involved in protein–protein interactions and is important for the function of DXHXS6673E.

A splice variant without the initiation codon in exon

2 (1AΔ2Δ5) represented the dominant 1A transcript in the mouse embryo at 14 days of gestation. At this stage, comparable to a 6-week-old human embryo, most of the major organs have been laid down, and future development is largely a matter of refinement and cellular differentiation. The nervous system is composed largely of differentiating neuroblasts. A detailed study of the diencephalon showed that the second stage of nuclear development in the diencephalic wall occurs from 13 to 15 days of gestation with differentiation of the various layers: germinal, mantle, and marginal (Rugh, 1990). Two other isoforms, 1AΔ2i and 1AΔ2iΔ3i, were observed only during development and in adult brain. This tissue-specific transcription and alternative splicing of variant 1A might be important with respect to the human DXS6673E. In this gene, disruption of exon 1A by a balanced translocation is associated with NSXMR, even though sequences 3′ of the breakpoint, and thus the ORF, are still expressed. Lack of the X-chromosomal promoter and part of exon 1A might lead to altered expression pattern, splicing, stability, or transport of 1A variants and hence influence brain development and functioning.

Since it is not known whether the Zn-finger motifs found in this novel family mediate DNA–protein or protein–protein interactions, it was important to study the subcellular localization of one of its members. The analysis of GFP–DXS6673E fusion proteins demonstrated a predominant nuclear localization. Due to the molecular weight of the various fusion proteins examined ($M_r > 75{,}000$), an active transport must occur, induced by nuclear targeting sequences (Ohno *et al.*, 1998). Our analysis of DXS6673E revealed five NLSs of the SV40 type and one bipartite NLS that may direct the nuclear transport. The most reliable indicator of nuclear localization, the bipartite motif, is present in more than 50% of nuclear proteins, while it is present in less than 5% of nonnuclear proteins (Dingwall and Laskey, 1991). The fact that its mutation did not alter the subcellular distribution of the fusion protein is likely due to the presence of additional functional nuclear targeting signals. That the bipartite NLS is dispensable for nuclear localization is supported by sequence analysis, since it is not present in other family members (ZNF198 and ZNF258). This signal might, therefore, be involved in fine-tuning of the nuclear transport of DXS6673E and DXSHXS6673E proteins.

Three different subtypes of cells could be identified according to the nuclear–cytoplasmic distribution of the fusion protein. Whether this observation reflects any physiological function like cell cycle dependency, and whether the elevated cytoplasmic concentration in some cells represents rather a retention of the synthesized protein in the cytoplasm or a nuclear export of the protein, is the object of current studies.

In summary, the data presented here show that DXS6673E and its mouse ortholog DXHXS6673E are members of an evolutionarily conserved family of zinc-finger proteins. The cloning of *DXHXS6673E* will facilitate genetic and functional analysis in an animal model to elucidate the precise role of this novel family further.

ACKNOWLEDGMENTS

We thank S. Schneider and D. Vogt for expert technical assistance. H. Madle and S. Freier are gratefully acknowledged for help with the cell culture. We are grateful to M. Hemberger, R. Kirschner, and S. Lenzner for providing mouse embryos and RNA from various mouse tissues and to the German Ressource Center for providing genomic P1 clones. This work was supported by the Bundesministerium für Bildung und Forschung (Grant 4763 to H. H. Ropers).

REFERENCES

Allen, K. M., Gleeson, J. G., Bagrodia, S., Partington, M. W., MacMillan, J. C., Cerione, R. A., Mulley, J. C., and Walsh, C. A. (1998). PAK3 mutation in nonsyndromic X-linked mental retardation. *Nat. Genet.* **20:** 25–30.

Billuart. P., Bienvenu. T., Ronce. N., des Portes, V., Vinet, M. C., Zemmi, R., Crollius, H. R., Carrie, A., Fauchereau, F., Cherry, M., Briault, S., Hamel, B., Fryns, J. P., Beldjord, C., Kahn, A., Moraine, C., and Chelly, J. (1998). Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. *Nature* **392:** 823–826.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162:** 156–159.

d'Adamo, P., Menegon, A., Lo Jigro, C., Grasso, M., Gulisano, M., Tamanini, F., Bienvenu, T., Gedeon, A. K., Oostra, B., Wu, S. K., Tandon, A., Valtorta, F., Balch, W. E., Chelly, J., and Toniolo, D. (1998). Mutations in GDI1 are responsible for X-linked non-specific mental retardation. *Nat. Genet.* **19:** 134–139.

Descombes, P., and Schibler, U. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell* **67:** 569–579.

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

Dingwall, C., and Laskey, R. A. (1991). Nuclear targeting sequences—A consensus? *Trends Biochem. Sci.* **16:** 478–481.

Dingwall, C., and Laskey, R. A. (1998). Nuclear import: A tale of two sites. *Curr. Biol.* **8:** 922–924.

Gecz, J., Gedeon, A. K., Sutherland, G. R., and Mulley, J. C. (1996). Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat. Genet.* **13:** 105–108.

Hann, S. R., King, M. W., Bentley D. L., Anderson C. W., and Eisenman, R. N. (1988). A non-AUG translational initiation in c-myc exon-1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell* **52:** 185–195.

Kulkarni, S., Reiter, A., Smedley, D., Goldmanm J. M., and Cross, N. C. (1999). The genomic structure of ZNF198 and location of breakpoints in the t(8;13) myeloproliferative syndrome. *Genomics* **55:** 118–121.

Lightfoot, J., Alon, N., Bosnoyan-Collins, L., and Buchwald, M. (1999). Characterization of regions functional in the nuclear localization of the Fanconi anemia group A protein. *Hum. Mol. Genet.* **8:** 1007–1015.

Nakabeppu, Y., and Nathans, D. (1991). A naturally occurring truncated form of FosB that inhibits Fos/Jun transcriptional activity. *Cell* **64:** 751–759.

Ohno, M., Fornerod, M., and Mattaj, I. W. (1998). Nucleocytoplasmic transport: The last 200 nanometers. *Cell* **92:** 327–336.

Pawson, T. (1995). Protein modules and signalling networks. *Nature* **373:** 573–580.

Popovici, C., Adelaide, J., Ollendorff, V., Chaffanet, M., Guasch, G., Jacrot, M., Leroux, D., Birnbaum, D., and Pebusque, M. J. (1998). Fibroblast growth factor receptor 1 is fused to FIM in stem-cell myeloproliferative disorder with t(8;13). *Proc. Natl. Acad. Sci. USA* **95:** 5712–5717.

Reiter, A., Sohal, J., Kulkarni, S., Chase, A., Macdonald, D. H., Aguiar, R. C., Goncalves, C., Hernandez, J. M., Jennings, B. A., Goldman, J. M., and Cross, N. C. (1998). Consistent fusion of ZNF198 to the fibroblast growth factor receptor-1 in the t(8; 13)(p11;q12) myeloproliferative syndrome. *Blood* **92:** 1735–1742.

Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: Identification of a class of bipartite nuclear targeting sequence. *Cell* **64:** 615–623.

Rugh, R. (1990). "The Mouse: Its Reproduction and Development," Oxford Univ. Press, Oxford.

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in *Caenorhabditis elegans*. *Science* **282:** 2033–2040.

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

Smedley, D., Hamoudi, R., Clark, J., Warren, W., Abdul-Rauf, M., Somers, G., Venter, D., Fagan, K., Cooper, C., and Shiple, J. (1998). The t(8;13)(p11;q11-12) rearrangement associated with an atypical myeloproliferative disorder fuses the fibroblast growth factor receptor 1 gene to a novel gene RAMP. *Hum. Mol. Genet.* **7:** 637–642.

van der Maarel, S. M., Scholten, I. H., Huber, I., Philippe, C., Suijkerbuijk, R. F., Gilgenkrantz, S., Kere, J., Cremers, F. P., and Ropers, H. H. (1996). Cloning and characterization of DXS6673E, a candidate gene for X-linked mental retardation in Xq13.1. *Hum. Mol. Genet.* **5:** 887–897.

Xiao, S., Nalabolu, S. R., Aster, J. C., Ma, J., Abruzzo, L., Jaffe, E. S., Stone, R., Weissman, S. M., Hudson, T. J., and Fletcher, J. A. (1998). FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13) leukemia/lymphoma syndrome. *Nat. Genet.* **18:** 84–87.