Genomic Structure of a Novel LIM Domain Gene (ZNF185) in Xq28 and Comparisons with the Orthologous Murine Transcript

Nina S. Heiss,* Gernot Gloeckner,† Dietmar Bächtner,† Petra Kioschis,* Sabine M. Klauck,* Bernd Hinzmann,† André Rosenthal,† Gail E. Herman,§ and Annemarie Poustka*.1

*Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; †Department of Molecular Genome Analysis, Institut für Molekulare Biotechnologie, Beutenbergstraße 11, 07745 Jena, Germany; ‡Department of Gene Regulation and Differentiation, Gesellschaft für Biotechnologische Forschung, Mascheroder Weg 1, 38124 Braunschweig, Germany; §Department of Molecular and Human Genetics and Pediatrics, Baylor College of Medicine, Houston, Texas 77030; and *Children’s Hospital Research Foundation, Department of Pediatrics, The Ohio State University, Columbus, Ohio 43210

Received December 23, 1996; accepted May 7, 1997

INTRODUCTION

As part of a long-range positional cloning approach in Xq28, transcript maps were established in the 300-kb region around the G6PD gene (Sedlacek et al., 1993), the 900-kb candidate region for myotubular myopathy (MTM-1; Kioschis et al., 1996), the 1200-kb candidate region for the hereditary form of incontinentia pigmenti (IP2; Rogner et al., 1996), and the 700 kb around the DXS52 loci (Heiss et al., 1996). Although the identification of novel genes linked to Xq28 is ongoing, an association with a particular disease is lacking for most. Undoubtedly disease-causing genes include the genes for Waisman syndrome (Gregg et al., 1991), otopalato-digital syndrome (Biancalana et al., 1991), Goe- minne syndrome (Zuffardi et al., 1982), X-linked mental retardation (MRX3; Gedeon et al., 1991; Nordström et al., 1992), the X-linked dominant form of chondrodysplasia punctata (CDPX2 or Happle syndrome; Traupe et al., 1992), dyskeratosis congenita (DKC; Connor et al., 1986; Arnglimsson et al., 1993), neuronal chronic idiopathic X-linked intestinal pseudodobstruction (CIIP; Auricchio et al., 1996), and periventricular heterotopia (Eksioglu et al., 1996). The candidate regions for most of these diseases cover several megabases and span the polymorphic DXS52 loci. The transcript map of the DXS52 region (Heiss et al., 1996) provided scope for the further characterization of single genes and allowed first associations to be made among tissue expression, homologies to known protein domains, possible function, and disease. Of the novel cDNAs located in the DXS52 region, XAP105 was interesting because of its homology to a murine cDNA that maps to the syntenic region on the mouse X chromosome. Both the human and the mouse cDNAs contain a LIM zinc finger domain in the carboxyl terminus and have been designated ZNF185 and Zfp185, respectively.

LIM domain proteins represent a diverse family of regulatory proteins occurring in mammals, birds, amphibians, fish, yeast, and Drosophila (Dawid et al., 1995). The term “LIM” stems from the first letters of the Caenorhabditis elegans lin11 gene (Freyd et al., 1990), the rat isl1 gene (Karlsson et al., 1990), and the C. elegans mec3 gene (Way and Chalfie, 1988), in which LIM domains were originally identified. The conserved cysteine and histidine residues that make up a LIM domain form a C2HC and a C4 motif, each of which...
coordinated binds one zinc ion (Sanchez-Garcia and Rabbits, 1994). Strengthened by findings that LIM proteins are distributed in the cytoplasm or nucleus of cells, the assumption that LIM proteins carry out regulatory functions by protein–protein interactions rather than by interacting directly with DNA is generally favored (Perez-Alvarado et al., 1994; Schmeichel and Beckerle, 1994). Although at present no consistent LIM protein dimerization partners are known, accumulating evidence suggests that these proteins regulate cellular proliferation and differentiation through diverse mechanisms (Dawid et al., 1995). This is reflected by their involvement in a wide range of cellular functions including transcriptional activation (isl-1; Karlsson et al., 1990), somatic patterning (lin-11, mec3; Freyd et al., 1990), focal cell adhesion (zyxin; Sadler et al., 1992), and immediate-early response to serum stimulation (CRP; Wang et al., 1992; Feuerstein et al., 1994). Conceivably, when mutated, LIM proteins may play a role in the etiology of hereditary and somatic diseases.

Currently there is no direct association between the novel ZNF185 LIM domain gene and a specific disease linked to Xq28 is apparent. Although it remains open to speculate about its potential role in the causation of disease, more information was gained by comparing the human and the orthologous mouse cDNAs at the sequence and expression levels. Furthermore, the genomic structure of the ZNF185 gene was determined.

**MATERIALS AND METHODS**

Sequencing and database comparisons. For the genomic sequencing of cosmids Qc11F9, Qc10G10, LA1733, and Qc17B8 (Heiss et al., 1996), the DNA was isolated by using the standard alkaline lysis method and purified on CsCl gradients (Radloff et al., 1967). A total of 10 μg of cosmid DNA was sonicated, and protruding ends were subsequently removed by treating the DNA with 30 U mung bean nuclease for 10 min at 30°C in 1X nuclease buffer (New England Biolabs). After phenol extraction and ethanol precipitation the DNA was size-fractionated on a 0.8% agarose gel. DNA fragments ranging in size from 1.2 to 1.6 kb were excised from the gel, purified with Qiaex II (Qiagen), and subcloned into the Smal site of the M13mp18 vector. Prior to carrying out the sequencing reaction, the DNA templates were prepared by using the TriZol preparation method. For the sequencing of cloned cDNAs, the plasmid DNA was extracted using the HUSAR were labeled to a specific activity of 10^6 to 10^7 cpm for random primed reactions and then subjected to at least 2 h of competition with 25 μg human Cot-1 DNA (Gibco, BRL) at 65°C. Hybridizations were carried out at 65°C for 16 h in Church solution, and filters were washed twice at 65°C for 20 min in 2X SSC containing 0.2% SDS. For Northern hybridizations, blots from Clontech were used.

RNA in situ hybridizations. Embryos and tissue samples from adult animals were fixed overnight in 4% paraformaldehyde in PBS at 4°C. Riboprobes were generated from the cloned 2-kb cDNA by RNA in vitro transcription. Prior to transcription, the vector was linearized with XbaI and MluI, and antisense and sense probes were then transcribed from the T3 and T7 promoters using T3 and T7 polymerases, respectively. In situ hybridization was performed as described previously (Bächner et al., 1993). Briefly, the probes were labeled to a specific activity of >10^6 dpm/μg with [α-35S]UTP. Probe length was reduced to 150–200 nucleotides by alkaline hydrolysis. The slides were prehybridized at 54°C in a solution containing 50% formamide, 10% dextran sulfate, 0.3 M NaCl, 10 mM Tris, 10 mM sodium phosphate, pH 6.8, 20 mM dithiothreitol, 0.2% Denhardt's reagent, 0.1% Triton X-100, 1.25 mg/ml yeast RNA, and "cold" 0.1 mM [α-35S]UTP. For hybridization, 80,000 dpm/μl [α-35S]UTP-labeled riboprobe was added to the hybridization mix, and hybridization was continued at 54°C for 16 h in a humid chamber. The slides were washed in hybridization salt solution to which dithiothreitol was added, competition, and filter hybridizations. Fifty to 100 ng of DNA was labeled to a specific activity of 5 × 10^6 to 10 × 10^6 cpm for random primed reactions and then subjected to at least 2 h of competition with 25 μg human Cot-1 DNA (Gibco, BRL) at 65°C. Hybridizations were carried out at 65°C for 16 h in Church solution, and filters were washed twice at 65°C for 20 min in 2x SSC containing 0.2% SDS. For Northern hybridizations, blots from Clontech were used.

**PCR analyses.** Exons were PCR amplified in 50-μl volumes. The cDNA templates were initially denatured for 2 min at 94°C. This was followed by 35 cycles at 94°C, 30 s, 60°C, 30 s, 72°C, 30 s and terminated with a final extension of 10 min at 72°C. To exclude the possibility that failure to amplify 5 exons was owing to incompatible primers, these were tested at the genomic level. To obtain fragments of adequate length, most exons were amplified by combining forward and reverse primers of adjacent exons. Primer sequences were:
FIG. 1. Schematic diagram showing the genomic structure of the ZNF185 gene. Exons 10–18 were derived from assembled cDNA clones isolated by cDNA selection and sequenced ESTs (open boxes). Exons 13–16 were simultaneously isolated by exon trapping (open boxes). Exons 5–9 (diagonally hatched boxes in solid lines) and exons 1–4 (horizontally hatched boxes in broken lines) were predicted based on homology of the mouse Zfp185 cDNA to the human genomic sequence. Exons 5–9 were verified at the peptide level, while prediction of exons 1–4 was solely based on nucleotide sequence homology. The positions of exons predicted by GRAIL are shown in small open boxes and are numbered A–F. The exons and introns are not drawn to scale.

added. After RNase A digestion (40 μg/ml) the slides were washed consecutively for 30 min at 37°C with 2 × SSC and 0.1 × SSC containing 0.1% SDS. The tissues were subsequently dehydrated by passage through increasing concentrations of ethanol. The slides were coated with Ilford K5 photoemulsion for autoradiography. After 3 weeks of exposure at 4°C, the slides were developed and stained with Giemsa solution. The embryos and sections were analyzed using bright- and dark-field illumination with a Zeiss SV11 stereo micro-

<table>
<thead>
<tr>
<th>Exon</th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>ZNF185 (human)</td>
</tr>
<tr>
<td>5</td>
<td>Zfp185 (mouse)</td>
</tr>
<tr>
<td>6</td>
<td>Exon 7</td>
</tr>
<tr>
<td>7</td>
<td>Exon 8</td>
</tr>
<tr>
<td>8</td>
<td>Exon 9</td>
</tr>
<tr>
<td>9</td>
<td>Exon 10</td>
</tr>
<tr>
<td>10</td>
<td>Exon 11</td>
</tr>
<tr>
<td>11</td>
<td>Exon 12</td>
</tr>
<tr>
<td>12</td>
<td>Exon 13</td>
</tr>
<tr>
<td>13</td>
<td>Exon 14</td>
</tr>
<tr>
<td>14</td>
<td>Exon 15</td>
</tr>
<tr>
<td>15</td>
<td>Exon 16</td>
</tr>
<tr>
<td>16</td>
<td>Exon 17</td>
</tr>
</tbody>
</table>

FIG. 2. CLUSTAL alignment of the ZNF185 and Zfp185 peptide sequences. The positions of the exons demarcated by arrows refer to the structure of the human gene. The LIM domain is shaded and conserved cysteine residues (C) and the histidine residue (H) are shown in boldface type. (**) Positions of identical amino acids; (.) positions of similar but conserved amino acids; spaces, different amino acid positions that may alter the protein structure; gaps, regions of dyshomology.
FIG. 3. CLUSTAL alignments of the ZNF185 and Zfp185 LIM domains with those of other LIM domain proteins. LIMKI and LIMKII represent the first and second LIM domains of the protein kinase double LIM domain protein (Mizuno et al., 1994), CRPLIMI and CRPLIMII represent the two LIM domains of the CRP protein (human cysteine-rich protein) (Liebhaber et al., 1990). The conserved cysteine and histidine residues are surrounded by boxes. Conserved hydrophobic residues are circled.

scope and an Axioplan microscope and photographed using Kodak Ektachrome 320T tungsten film.

RESULTS

Structure of the ZNF185 Gene

The ZNF185 cDNA (EST XAP105; Accession No. X96622) had been assembled from overlapping cDNA selection clones isolated by using complex exon trap products as probes on enriched cDNA libraries (Heiss et al., 1996). By sequencing 1 of the 19 EST clones identified by BLASTN searches (clone 26810; Accession No. R14025), a final cDNA sequence with a length of 3140 bp was obtained. Although the 5’ end of the cDNA was missing, a continuous open reading frame of 307 amino acids terminating in a TAG stop codon was present. The 3’ UTR consisted of 2220 bp bearing a AATAAA poly(A) signal and a poly(A) tail.

To determine the genomic structure of the ZNF185 gene, cosmids to which ZNF185 mapped were sequenced. Alignment of the cDNA sequence with the genomic sequence revealed the genomic structure of the 3’ end of the gene and consisted of exons 10–18 and introns X–XVII (Fig. 1). Exons 13–16 had previously been defined by four adjacent exon trap products (Heiss et al., 1996), and their authenticity was verified by the availability of the genomic sequence (Fig. 1). Attempts to extend the 5’ end of the ZNF185 cDNA by performing 5’ RACE and screening conventional cDNA libraries were unsuccessful. Furthermore, GRAIL did not reliably predict the 5’ exons. Subsequently, the 5’ end of the human cDNA was constructed by comparing the human genomic sequence with the full-length murine Zfp185 cDNA sequence (clone P1A; Accession No. U46687; Levin et al., 1996). The homology of the mouse cDNA and the human genomic sequence was sufficiently high to permit an accurate prediction of the size and location of the 5’ ZNF185 exons. In this way, exons 4–9 and introns IV–IX were identified (Fig. 1). Except for the splice junctions flanking exon 4, all other predicted exon–intron boundaries represented conserved splice junctions and maintained the open reading frame. Although homology between the human and the mouse first coding exons was low at the nucleotide level, a higher degree of conservation at the peptide level facilitated determination of the position and sequence of human exon 4. Alignment of the mouse cDNA with the human genomic sequence indicated that the 5’ UTR probably contains three additional noncoding exons (exons 1–3) with their respective introns (introns I–III; Fig. 1).

While exons 5, 6, 10, 11, 12, and 14 present in the human and mouse cDNA sequences were also recognized by GRAIL, the remaining exons were not predicted (Fig. 1). GRAIL, however, predicted six additional exons, numbered A–F (Fig. 1). The predicted exons were flanked by classical splice junctions, but because they were not present in the cDNA it was necessary to verify their authenticity. Failure to amplify GRAIL exons A–F by RT-PCR implied that they do not represent true exons. It cannot be excluded, however, that the predicted exons represent alternatively spliced exons. By adding up the length of the cDNA with the predicted 5’ end and by taking into account an average length of 200 bp for the poly(A) tail, a length of ~4.2 kb was calculated for the full-length transcript. From the combined sequence data, the ZNF185 gene was found to span ~59 kb of genomic DNA and to exhibit a centromere to telomere direction of transcription.

A number of diseases are associated with regions of expanded triplet repeats that may occur in the coding or noncoding sequence of the candidate gene (Mandel, 1997). On searching the cDNA sequence for such repeats, although atypical, a 7× GAG repeat was detected in exon 5 of the ZNF185 cDNA. This repeat was absent in the murine Zfp185 cDNA. In addition, a 5× and 3× GGA repeat interrupted by TGA in intron XI was identified in the genomic sequence of ZNF185. The
A NOVEL LIM DOMAIN GENE (ZNF185) IN Xq28

The cDNA and the genomic sequences have been given Accession Nos. Y09538 and U82671, respectively.

Comparisons of the ZNF185 and Zfp185 cDNA and Peptide Sequences

FASTA alignments demonstrated that the homology of the human and mouse cDNAs was unevenly distributed, with an average identity of only 71.6 and 54% at the nucleotide and peptide levels, respectively. Regions of dyshomology and regions of very high homology were confined to certain exons (Fig. 2). While at the nucleotide level, the first coding exon, exon 4, was not very conserved (53.8% identity); exons 5–9 showed higher levels of homology (71.4–89% identity). The degree of homology dropped within exons 10–12 and was largely due to the presence of a duplicated exon, exons 11 and 12, found only in the human cDNA (Fig. 2). The duplicated exons exhibited 92% identity at the nucleotide level, while at the peptide level, 81.8% of the amino acids were identical, and 7.6% were different but conserved. Owing to the high homology of human exons 11 and 12, it was impossible to determine which one of the two exons is missing in the murine counterpart. The duplicated exons were also present in the independently derived genomic sequence and confirmed the authenticity of the result. Although the intron sequences surrounding exons 11 and 12 were not wholly conserved, short blocks of repetitive sequences bordering the exons were identified. The sequence TTCATAAGG-GTGCTATTTCTTCTTCTT(C or CCT)CAG immediately flanked the 3' intron splice site of both exons. The only difference was the lack of the G (underlined) and a CCT (in parentheses) instead of CTT in the sequence preceding exon 12. The sequence flanking the 5' intron splice site was less well conserved, but contained the sequence TGTTTTTGT 40 and 42 nucleotides into the intron sequence following exons 11 and 12, respectively. Downstream of exon 12, the homology to the murine cDNA rose dramatically, reaching 93.9% in the last coding exons 16 and 17. This coincided with the presence of a conserved LIM zinc finger domain in both cDNAs.

In the segment encoding the LIM zinc finger domain only four diverging residues were identified, of which two were conserved and two were unique (Figs. 2 and 3). The motif consisted of a stretch of 56 amino acids containing the conserved sequence CysX2CysX19HisX2. CysX2CysX19CysX19CysX2CysX2CysX2CysX2Cys. The first “finger” demarcated by the C2HC motif was encoded by exon 16, while the second “finger” defined by the C4 motif was encoded separately by exon 17 (Figs. 2 and 3). In most LIM domains the amino acids X17–19 forming the loop regions, although less rigidly conserved, contain hydrophobic isoleucine (I), valine (V), leucine (L), phenylalanine (F), and/or methionine (M) residues at conserved positions (Dawid et al., 1995). In agreement with this, both the ZNF185 and the Zfp185 peptides contained such conserved hydrophobic residues (Fig. 3). In addition, a conserved glycine (G) residue at the fourth amino acid position following the histidine (H) residue found in many but not all LIM domains was present (Fig. 3). Hydrophobicity profiles showed that the peptide sequences 5' of the LIM domain do not
contain hydrophobic regions of sufficient length as is
typical of signal peptides or membrane-spanning pro-
teins. In accordance with the low hydrophobicity pro-
files and a probable nuclear localization, both the hu-
man and the mouse peptides have a low pl of 5.29
and 4.86, respectively. Although no further regulatory
homeodomains or kinase domains were identified, a
proline (P) content of 8.850 and 8.239% for the ZNF185
and Zfp185 peptides, respectively, was evident. Consid-
ering that the proline content of most eukaryotic pro-
teins typically lies at 4.5–5.5%, this is remarkably
high.

Expression of the ZNF185 and Zfp185 Genes

Northern blot hybridizations. Northern blots pre-
viously showed that ZNF185 is expressed as a single
~5-kb transcript in placenta, pancreas, and kidney
(Heiss et al., 1996). On acquiring more cDNA sequence and extending the Northern blot analyses, however, additional ~3.8- and ~4.4-kb transcripts expressed variably in different tissues were identified (Fig. 4). The cDNA discussed in the present report, including the predicted 5’ end, probably represents the ~4.4-kb transcript. To investigate whether the presence of more than one transcript was owing to the existence of a related gene, panels with total genomic DNA, genomic DNA from the Q1Z cell line carrying Xq28 on a hamster background, and micro21d bearing the remainder of the X chromosome on a hamster background were hy-
bridized. One extra band was observed in the lanes
containing total genomic DNA, suggesting that
ZNF185 belongs to a gene family (Fig. 4).

RNA in situ hybridizations. RNA in situ hybridiza-
tions were performed on mouse embryo sections from
10.5 to 18.5 days post coitum (dpc) and on different
adult tissue sections. No expression was evident in
10.5- and 12.5-dpc mice and was only visible in mid to late gestation embryos. Expression was first detected in 14.5-dpc embryos and was restricted to distinct
condensing mesenchymal cells, for example, adjacent to
the hyoid bone (Figs. 5d and 5e). In midgestation de-
velopment, enhanced expression was also observed as an
elongated stripe of connective tissue adjacent to the
vertebrae (Figs. 5f and 5g), in condensing mesenchymal
cells of the limbs (not shown), and in the proximal tail
region (arrows in Fig. 5i). Possibly, these cells repres-
sent differentiating tendons. This assumption was sup-
pported by the presence of high levels of expression in
more differentiated tendons (arrowheads in Fig. 5i).

While expression in tendons of the limbs was main-
tained in late embryogenesis (arrows in Fig. 5i), a lower
level of expression was additionally observed in the
connective tissues sheaths (epimysium) surrounding the
skelatal muscles (Figs. 5k and 5l). Expression was also
observed in mesenchymal cells adjacent to the de-
veloping distal skeletal elements of the limbs, for example,
in the distal tibia and the calcaneum (Figs. 5a and 5b)
and in the epithelia of the epididymis of the testis (not
shown). Hybridization with a sense control probe
showed only background grain distribution (Fig. 5c).

DISCUSSION

The candidate regions for many diseases linked to
Xq28 are in the megabase range and span the DXS52
loci. The chromosomal region surrounding these marks
appears to contain a hot spot of recombination owing
to the highly polymorphic and repetitive nature of
the DXS52 loci (Oberlé et al., 1985; Bell et al., 1989; Fei
et al., 1990). Furthermore, the presence of expressed
replicative sequence elements consisting of the MAGE
gene cluster in humans (Rogner et al., 1995; Heiss et
al., 1996) and the Xlr3 gene cluster in the syntenic
region on the mouse X chromosome (Levin et al., 1996)
suggests a gene and “conserved” chromosomal insta-
bility. As part of a long-range positional cloning ap-
proach in Xq28, we previously established a transcript
map in a 700-kb region surrounding the DXS52 mark-
ers (Heiss et al., 1996). The ZNF185 cDNA (EST
XAP105) contains a LIM zinc finger domain, suggesting
a function in the regulation of cellular proliferation
and/or differentiation and a possible association with
disease. For this reason, we characterized this gene
further.

The genomic structure of the 3’ part of the ZNF185
gene became apparent by comparing the 3140-bp cDNA
with the cosmid-derived genomic sequence. The cDNA
did not represent a full-length transcript, and attempts
to isolate its 5’ end were unsuccessful. By using the
genomic sequence and the 5’ end of the orthologous
mouse cDNA as a guide (Levin et al., 1996), an accurate
prediction of the remaining transcript was accom-
plished. Since GRAIL predicted six false and six true
exons, but failed to predict the remaining eight coding
and four noncoding exons, this indicated that although
GRAIL is useful in predicting the genomic structure of
some genes (Chen et al., 1996), it is unsuccessful in
predicting others (Lopez et al., 1994). This justifies the

![FIG. 5.](image-url) RNA in situ analysis of Zfp185 expression in mouse embryogenesis. Bright-field (a, d, f, h, k) and corresponding dark-field (b, c, e, g, i, l) images of transversal sections through 18.5-dpc (a–c, k, l) and sagittal sections through 14.5-dpc (d, e) and 16.5-dpc (f–l) embryos hybridized with a Zfp185 antisense (b, e, g, i, l) and a sense (c) riboprobe are shown. At Day 14.5 dpc enhanced expression was visible at
condensed mesenchymal cells adjacent to the hyoid bone (d, e). At Day 16.5 enhanced expression was further visible in cells lining the
vertebrae (f, g) and tendons of the proximal tail (arrowheads and arrows in i). In late embryogenesis enhanced expression was detected in
mesenchymal cells adjacent to distal limb bones, for example tibia and calcaneum (b), tendons (arrows in l), and the connective tissue
sheaths surrounding the skeletal muscles (l). Only unspecific grain distribution is obtained with the sense control hybridization (c). bl, bladder; bv, blood vessel; cal, calcaneum, cms, condensed mesenchyme; fem, femur; fib, fibula; g, gut; hyb, hyoid bone; ki, kidney; ms, mesenchyme; sc, spinal cord; sk, skin; sm, skeletal muscle; ta, tail; ten, tendons; tib, tibia; ve, vertebra. Bar a–c, 1 mm; d–l, 100 μm.
importance of using cDNA clones and not relying solely on the genomic sequence for gene identification. In view of the ultimate goal aimed at establishing gene maps, our data also outline the value of integrating the physical maps with the transcript and sequence maps.

Northern blots revealed three transcripts expressed selectively in different tissues. Because Southern blots detected a second locus mapping outside of Xq28, it currently remains uncertain which of the transcripts are derived from the ZNF185 gene and which correspond to the related gene. Furthermore, alternative splicing or alternative polyadenylation may give rise to the third transcript. In agreement with the Northern blots, in situ hybridizations revealed a nonubiquitous and stage-specific expression of Zfp185. In contrast to the Northern blots, however (Fig. 4a; Levin et al., 1996), in situ hybridization detected no expression in organs at any stage of development, but was rather confined to differentiating connective tissue.

The high degree of conservation between genes on the human and the mouse X chromosomes implies a conservation in function (Davison, 1987; Lundin, 1993). The homologous regions are preserved in the form of subchromosomal blocks and include Xq28. Although exceptions do exist (Faust et al., 1992), gene order is largely conserved within the blocks (Davison, 1987; Blair et al., 1995; Rivella et al., 1995). This was again demonstrated by a comparative analysis of the human DXS52 region and the syntenic region on the mouse X chromosome (Levin et al., 1996). Although ZNF185 and Zfp185 represented orthologous genes, they exhibited regions of dyshomology at the nucleotide and peptide sequence level. The largest segment of dyshomology extended from exons 10 to 12 and was largely owing to the presence of a duplicated exon contained only in the ZNF185 cDNA. This points to a sequence divergence after separation of the ancestral human and mouse lineages and is indicative of a recent evolutionary event. It remains open to speculation whether the human exon duplicated or the mouse exon became deleted or rearranged. One may presume with considerable certainty, however, that the function imposed by the LIM domain remained conserved.

According to the LIM protein classification system, ZNF185 and Zfp185 belong to the third group of LIM proteins. Group 3 LIM proteins contain between one and five LIM domains in the carboxyl-terminus, usually in the absence of other regulatory domains (Taira et al., 1995), and include proteins such as the cytosine-rich intestinal protein (CRIp; Birkenmeier and Gordon, 1986; Liebhaber et al., 1990), the ESP1/CRP2 protein (Karim et al., 1996), the cell-adhesion molecule zyxin (Sadler et al., 1992), and the rhombotin gene family (RBTN1-3; Foroni et al., 1992). In contrast to the RBTN3 (Foroni et al., 1992), the mouse Lhx3 (Zhadanov et al., 1995), and the LPP (Petit et al., 1996), proteins whose LIM domains are encoded by one exon, the two motifs composing the LIM domain of ZNF185 are encoded by two separate exons. Several LIM proteins contain additional kinase domains (Mizuno et al., 1994; Bernard et al., 1996), homeodomains (Zhadanov et al., 1995), and serine threonine kinase domains (Cheng and Robertson, 1995). Although ZNF185 and Zfp185 did not contain such extra regulatory domains, both peptide sequences preceding the LIM domain were proline-rich. Proline-rich regions are typically associated with transcriptional activation domains and interact specifically with other factors to initiate transcription (Mermod et al., 1989; Freyd et al., 1990; Cheng and Robertson, 1995). Furthermore, proline-rich regions are frequently interdigitated with negatively charged amino acids (Struhl, 1987) and have been proposed to mediate contacts with positively charged regions of other transcription factors (Ptashne, 1988). The proline-rich domains following and preceding the amino- and carboxyl-terminal LIM domains of the rat isl-1 (Karlsson et al., 1990) and LPP LIM (Petit et al., 1996) proteins, respectively, appear to exhibit such functional activity. Conceivably, the proline-richness and the acidic nature of the ZNF185 and Zfp185 peptides may be of similar significance.

The composition of the ZNF185 and Zfp185 peptide sequences permits prediction of a function in regulating cellular proliferation and/or differentiation, thus implicating a role in the etiology of disease. Levin et al. (1996) previously suggested that Zfp185 may be a candidate for the Bpa mutant, which is the murine counterpart of the human dominant form of chondrodysplasia punctata (CDPX2) linked to Xq28. In the present report, however, expression analyses did not directly support this supposition. It is nonetheless of interest to note that candidate triplet repeat regions were found in exon 5 and intron XI of the ZNF185 gene. Although no anticipation effects have been observed in the mouse mutant (G. Herman, unpublished data) and the question of whether CDPX2 is a triplet repeat disorder remains open, increasing severity of the disease in later generations has been observed in some families. ZNF185 thus remains a good candidate for CDPX2. Knowledge of the genomic structure of ZNF185 will permit extensive mutation analyses. Because numerous diseases are linked to the DXS52 region, but a direct association with a specific disease is lacking, ZNF185 also remains a candidate for all of these diseases.

ACKNOWLEDGMENTS

We thank Petra Wilgenbus and Alexandre Ochs for carrying out the cDNA sequencing work. We thank Zdenek Sedlacek for critically reading the manuscript. This research was supported by grants from the BMBF (Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie) and the Human Genome Analpy Program of the European Community and in part by the National Institutes of Health R01 NS34953 to G.E.H.

REFERENCES


...
A novel LIM domain gene (ZNF185) in Xq28


