

Received 27 October 2004
Accepted 18 January 2005

MOLECULAR CLONING AND CHARACTERIZATION OF A NOVEL HUMAN GENE CONTAINING 4 ANKYRIN REPEAT DOMAINS

JIXI LI, CHAONENG JI, HUARUI ZHENG, XIANGWEI FEI, MEI ZHENG,
JIANFENG DAI, SHAOHUA GU, YI XIE and YUMIN MAO*

State Key Laboratory of Genetic Engineering, Institute of Genetics, School of
Life Sciences, Fudan University, Shanghai 200433, People's Republic of China

Abstract: Ankyrin repeat, one of the most important protein motifs, plays a wide variety of roles in protein-protein interactions and in the signal pathways. Via large-scale sequencing, a novel 941-bp gene was isolated from an 18-week old human fetal brain cDNA library. It encodes a putative protein of 158 amino acid residues with four conserved ankyrin repeat domains. It displays a high degree of homology with rat low-density lipoprotein receptor-related protein 2-binding protein (Lrp2bp), and was therefore named hLrp2bp (human Lrp2bp). The hLrp2bp gene was located in chromosome 4q35 and the conserved ankyrin repeat domains were located between amino acid residues 10 and 116. RT-PCR revealed that hLrp2bp was mainly expressed in the human testis, small intestine, colon and blood leukocytes, and in human pancreatic adenocarcinoma cells. A HEK293 cell was transfected with the ORF of hLrp2bp, and analyses showed that the protein was distributed both in the cytoplasm and nucleus.

Key Words: hLrp2bp, Ankyrin Repeat Domain, RT-PCR, Subcellular Location

INTRODUCTION

The ankyrin repeat is an approximately 33 residue protein motif that was originally found in two yeast cell-cycle regulators, Cdc10 and Swi6, and in the Notch and LIN-12 developmental regulators from *Drosophila melanogaster* and

*Corresponding author; tel: +86-21-65643573, fax: +86-21-65642502, e-mail: ymmao@fudan.edu.cn

Abbreviations used: hLrp2bp – human low-density lipoprotein receptor-related protein 2-binding protein; RT-PCR – reverse transcription PCR; GFP – green fluorescent protein.

Caenorhabditis elegans [1,2], and subsequently in the cytoskeletal protein ankyrin [3]. The ankyrin repeat is one of the most common protein sequence motifs, and has been discovered in 400 proteins, including cyclin-dependent kinase (CDK) inhibitors, transcriptional regulators, cytoskeletal organizers, developmental regulators and toxins in various organisms ranging from viruses to humans [4-6]. The number of ankyrin repeats within any one protein varies considerably. Some ankyrin repeat proteins consist only of ankyrin repeats; others are multi-domain molecules in which ankyrin repeats are combined with other unrelated structural motifs [4]. Most ankyrin repeat proteins contain four to six repeats, while the tumor suppressor protein, p16, may be the minimum folding unit for ankyrin repeats in general [7].

Ankyrin repeat domains not only serve as protein-protein interaction domains [5] providing the linkage between the cytoskeleton and the membranes of intracellular organelles, but they are also involved in the protein sorting, signal transduction and signal regulation pathways [8-11]. Loss of ankyrin has grave functional consequences, including embryonic fatality, developmental aberration, and membrane instability [12, 13]. One ankyrin molecule is able to bind several different membrane proteins facilitating the assembly of integral membrane proteins into specialized regions of the plasma membrane [14]. Ankyrin repeat proteins are also related to Na-channels [15], Na/K-ATPase [16], and the ryanodine receptor [17].

In general, the conserved repeat domains of ankyrin have the following sequence [3, 18]: -GXTPLHXAAXXGHXXXV/AXXLLXXGAXXN/DXXX-X-. Fifteen amino acid residues in these repeats are highly conserved, while the others vary in different repeats. A single ankyrin repeat forms two anti-parallel α -helices connected by a short loop, followed by a β -hairpin, or a long loop in some instances, which protrudes away from the helices at an approximately 90-degree angle [19]. The structure of an ankyrin repeat resembles that of a hand, with the protruding β -hairpin/loop as the fingers and the helix-loop-helix as the palm. Ankyrin repeats pack against each other in a linear fashion with hydrophobic interactions predominating between the helices and hydrogen bonding predominating between the β -hairpin/loop regions. Myotrophin was shown as a model system to study the folding of ankyrin repeat proteins. The protein displays cooperative two-state folding properties despite its modular nature [20]. The X-ray and NMR structures of ankyrin repeat proteins and their complexes have provided invaluable insights into the molecular basis for the extraordinary variety of biological activities of these molecules [19].

Here, we report on the molecular cloning and characterization of a novel hLrp2bp gene. The hLrp2bp gene, which contains four ankyrin repeat domains, is mainly expressed in the human testis, small intestine, colon and blood leukocytes, and in human pancreatic adenocarcinoma cells.

MATERIALS AND METHODS

Cloning and sequencing of hLrp2bp cDNA[#]

A cDNA library was constructed using human fetal brain mRNA purchased from Clontech. Double-strand cDNAs were prepared with a SMARTTM PCR cDNA Synthesis Kit. After *Sfi* I digestion, cDNAs greater than 500 bp were ligated into the *Sfi* I A and *Sfi* I B sites of the modified pBluescript II SK (+) vector, and then transformed into the *Escherichia coli* DH5 α using electroporation (*E. coli* pulser, Bio-Rad). Both 5' and 3' ESTs were generated with either dye primer or dye terminator chemistries on an ABI377 sequencer using M13 consensus primers. Primer walking was performed if necessary. Assembly program (Sanger Center) was used to assemble the full-length cDNA sequences.

Sequence analysis and chromosomal location

The DNA and putative protein sequence comparisons were carried out using the BLAST tool at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST>). Multiple sequence alignment analysis was performed using Clustal W algorithms. Other sequence analyses were done with Genedoc Software. A BLAST-N search against the human genome was performed to identify the chromosomal location of hLrp2bp.

Reverse transcription PCR

One human Multiple Tissue cDNA (MTC) panel and one human tumor MTC panel were used as the PCR template according to the manufacturer's protocol (Clontech). The primer sequences of hLrp2bp were 5' TGCTTCTTCTCTGGCAGCTGCAA 3' (hLrp2bp-f, corresponding to nucleotides 372-395) and 5' TCCCACTGGTATTTTCTACACTTCC 3' (hLrp2bp-r, corresponding to nucleotides 678-702). Twenty-four cycles (for GAPD) or thirty-six cycles (for hLrp2bp) of amplifications (30 s at 94°C, 120 s at 68°C) were performed using Taqplus polymerase (Sangon) in a volume of 50 μ l. The PCR product of hLrp2bp was resolved on 1.5% Metaphor agarose.

The subcellular location of hLrp2bp

The putative ORF of hLrp2bp was cloned into the pEGFP-C1 expression vector (Clontech), allowing the expression of hLrp2bp as a green fluorescent fusion protein (GFP). The generated fusion plasmid was then transfected into HEK293 cells with LipofectamineTM 2000 (Invitrogen). The blank pEGFP-C1 plasmid served as a control.

[#] The nucleotide sequence reported on in this paper has been submitted to the Genbank/EMBL Database with the accession number AY296056.

RESULTS AND DISCUSSION

Cloning and sequence analysis of the hLrp2bp gene

Ankyrin repeats have been identified in hundreds of proteins that are found in prokaryotes and eukaryotes. Like other specific repeats such as WD-repeats, SH2- and SH3-domains, ankyrin repeats may be a universal module mediating protein-protein interactions in the evolution (<http://coot.embl-heidelberg.de/SMART>). We cloned a novel hLrp2bp gene from the human fetal brain cDNA library and deposited it into GenBank with the accession number AY296056. The length of the cDNA is 941bp and it has an open reading frame from 240 to 716, encoding a putative protein with a predicted molecular mass of 16.9 kDa (Fig. 1). Bioinformatic analysis revealed that hLrp2bp contains four

```

1   gcagttttgttctacgtccggccgtggcggcctggacagggccggctcctttaggat+
60  taggctgctcggcctaacccgaactggcggcggcctccaaaggccggccggaaagtc+
120 caecggctcttaaaattctccgtctcagggccacctcggcattcttacctctcggggtg+
180 cggcgagtgtctcacctctctgcacttccaaaggactcttctcatctccttagcgggaa+
240 atgctgttctgctgattgcaaccccgaggatggatgctcgaagcatttctgagacagg+
    M L L L D C N P E V D G L K H L L E T G+
300 gcctcggcacaaccacccccgatccctgcaagcagtcgctctccacttagccgaggaa+
    A S V N A P P D P C K Q S P V H L A A G+
360 agcggccttctcttcttctctctcggcagctgcaaacggcggcctgacctcaaccagcag+
    S G L A C F L L W Q L Q T G A D L N Q Q+
420 gatgttttaggagaactccactacacaaggcagcaaaaagttggaagcctggagtccta+
    D V L G E A P L H K A A K V G S L E C L+
480 agcctcctttagccagtgatgccaaattgatttatgtaataagaacggcacaacagct+
    S L L V A S D A Q I D L C N K N G Q T A+
540 gaagatctccttggctcagtgatttccagactgtgccaaagtttcttacaacaattaaa+
    E D L A W S C G F P D C A K F L T T I K+
600 tgtatgcagacaataaaaagcaagtgaaacccctgacaggaatgattgtttccctgctc+
    C M Q T I K A S E H P D R N D C V A V L+
660 agacagaaacggagctcgggaagttagaaaataccagtgggaaaaaggaagtcctgat+
    R Q K R S L G S V E N T S G K R K C *+
720 cacgtgggttatgaaagctcgaagaacgcttcatttcagcaaatctacaagctcct+
780 gcttttgctttaccatattttgtgtataatctcctctcgaaggacgagaaactttct+
840 tccaagtgaaatccatttaagaacacatgtatttacatgcctataatctgctggttg+
900 tatgctttgtcttttaagttatttaaaaggaacgtctagaaaaa+

```

Fig. 1. The cDNA nucleotide and deduced amino acid sequences of hLrp2bp. The ORF of the 941-bp cDNA extended from nucleotide 240 to 716 and encoding a protein of 158 amino acids. The stop codon (tga) is indicated by '*'. The PKC phosphorylation sites are marked with boxes and the CK2 phosphorylation sites are underlined.

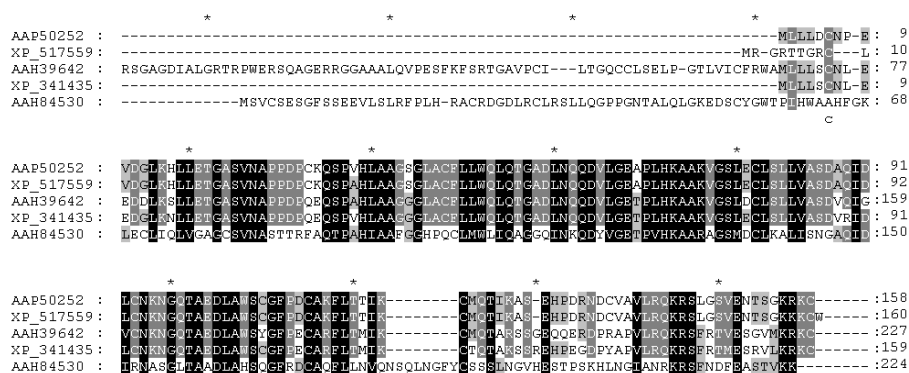


Fig. 2. The protein sequence alignment of AAP50252 (hLrp2bp), XP_517559 (from *Pan troglodytes*), AAH39642 (from *Mus musculus*), XP_341435 (from *Rattus norvegicus*) and AAH84530 (from *Xenopus tropicalis*). Conserved amino acids are shaded.

ankyrin repeat domains located between amino acid residues 10 and 116. The amino acid alignment with four proteins (XP_517559, AAH39642, XP_341435 and AAH84530) is shown in Fig. 2. Protein motif analysis of hLrp2bp revealed that hLrp2bp contains two casein kinase II phosphorylation sites ([ST]X2[DE]) located at residues 53 and 99, and three protein kinase C phosphorylation sites ([ST]X[RK]) located at residues 118, 124 and 153 (Fig. 1) [21]. The conserved ankyrin repeat domains reveal that hLrp2bp may be involved in protein-protein interactions.

Chromosome location of hLrp2bp

By searching the human genome database, we found that the hLrp2bp gene was mapped to 4q35 in a contig NT_022792.15. The comparison result also showed that the hLrp2bp gene has five exons and four introns spanning 3.5kb of human genomic DNA (Fig. 3). All sequences at the exon-intron junction were in consensus with the AG-GT consensus sequence (Tab. 1) [22].

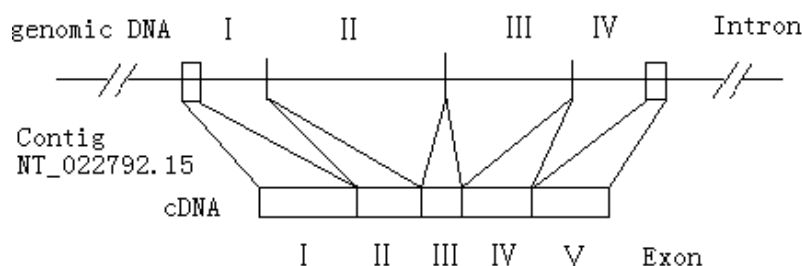


Fig. 3. The genomic structure of the hLrp2bp gene.

Tab. 1. The nucleotide sequence of the exon-intron junctions of the hLrp2bp gene. The intron sequence is shown in lowercase and the exon sequence in uppercase.

3' Splice acceptor	Exon	Size (bp)	5' Splice donor	Intron	Size (bp)
tcttcctgcagGTGGATGGTCT	1	266	GCAACCCCGAGgtgagattcgg		
cttacccttaagGATGTTTTAGG	2	153	TCAACCAGCAGgtaactagta	1	200
cctgttttcagTTTATGTAATA	3	92	GCCCAAATTGAGtgagtatgaa	2	1635
ttataaacagATGTCACGTGG	4	204	AGGAAGTGCTGgtaagtaactc	3	543
	5	226		4	245

Tissue distribution of hLrp2bp

We investigated the expression pattern of hLrp2bp via PCR amplification of human cDNA and human tumor cDNA. The hLrp2bp gene was mainly detected in the testis, small intestine, colon and blood leukocytes, and in poorly differentiated pancreatic adenocarcinoma cells (Fig. 4). The size of the product was the expected 330 bp. We also determined the distribution of hLrp2bp expression using Clontech multiple tissue Northern blot membrane IV, but did not detect the hybridization signal (data not shown). The hLrp2bp gene was cloned from a human fetal brain library. However, it cannot be detected in adult human brain tissue. It is also reported that some ankyrin repeat proteins were expressed in the muscles, macrophages, endothelial cells and embryonic brain [23-26]. This data shows that hLrp2bp may be related to cell development by way of expression level alterations. hLrp2bp was mainly expressed in poorly differentiated pancreatic adenocarcinoma cells and could not be detected in the pancreas.

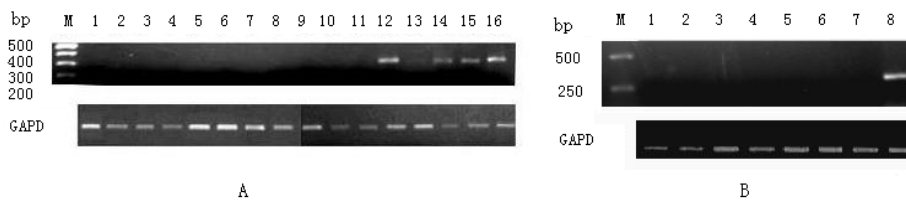


Fig. 4. Reverse transcription PCR analysis of the hLrp2bp cDNA with human tissues and human tumor tissues. GAPD was used as a control. A: M: marker; 1: Heart; 2: Brain; 3: Placenta; 4: Lung; 5: Liver; 6: Skeletal Muscle; 7: Kidney; 8: Pancreas; 9: Spleen; 10: Thymus; 11: Prostate; 12: Testis; 13: Ovary; 14: Small Intestine; 15: Colon; 16: Blood Leukocyte; B: M: marker; 1: human breast carcinoma GI-101; 2: human lung carcinoma LX-1; 3: human colon adenocarcinoma CX-1; 4: human lung carcinoma GI-117; 5: human prostatic adenocarcinoma PC3; 6: human colon adenocarcinoma GI-112; 7: human ovarian carcinoma GI-102; 8: human pancreatic adenocarcinoma GI-103.

The subcellular location of hLrp2bp

The pEGFP-C1-hLrp2bp fusion plasmid was transfected into HEK293 cells. After 48 h of expression, the hLrp2bp GFP was detected in both the cytoplasm

and nucleus (Fig. 5). The control protein is distributed throughout the cell, as expected.

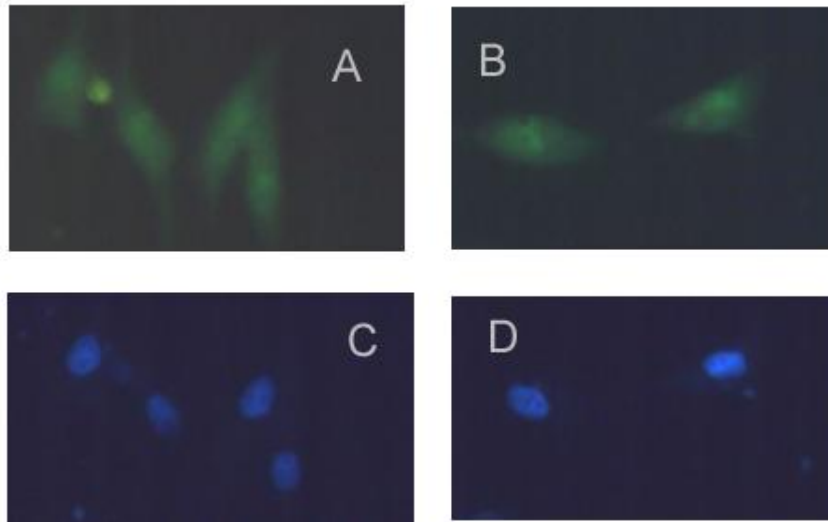


Fig. 5. Subcellular location of hLrp2bp-pEGFP fusion protein in transfected HEK 293 cells. A - pEGFP protein. B - hLrp2bp-pEGFP fusion protein. C - pEGFP protein colored with DAPI. D - hLrp2bp-pEGFP fusion protein colored with DAPI. The hLrp2bp-pEGFP fusion product is distributed both in the cytoplasm and nucleus and the pEGFP control is distributed throughout the whole cell as expected.

In conclusion, a novel cDNA has been cloned which contains the ankyrin repeat domain and may be involved in mediating protein-protein interactions. Although its function is unclear, it may have an important role as indicated by its high conservatism in many species.

REFERENCES

1. Aves, S.J., Durkacz, B.W., Carr, A. and Nurse, P. Cloning, sequencing and transcriptional control of the *Schizosaccharomyces pombe* cdc10 'start' gene. **EMBO J.** 4 (1985) 457-463.
2. Breeden, L. and Nasmyth, K. Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of *Drosophila*. **Nature** 329 (1987) 651-654.
3. Lux, S.E., John, K.M. and Bennett, B. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. **Nature** 344 (1990) 36-43.
4. Sedgwick, S.G. and Smerdon, S.J. The ankyrin repeat: a diversity of interactions on a common structural framework. **Trends Biochem. Sci.** 24 (1999) 311-316.

5. Michaely, P. and Bennett, V. The ANK repeat: a ubiquitous motif involved in macromolecular recognition. **Trends Cell Biol.** 2 (1992) 127-129.
6. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? **Proteins** 17 (1993) 363-374.
7. Zhang, B. and Peng, Z. A minimum folding unit in the ankyrin repeat protein p16INK4. **J. Mol. Biol.** 299 (2000) 1121-1132.
8. Bennett, V. and Chen, L. Ankyrins and cellular targeting of diverse membrane proteins to physiological sites. **Curr. Opin. Cell Biol.** 13 (2001) 61-67.
9. Rubtsov, A.M. and Lopina, O.D. Ankyrins. **FEBS Lett.** 482 (2001) 1-5.
10. De Matteis, M.A. and Morrow J.S. The role of ankyrin and spectrin in membrane transport and domain formation. **Curr. Opin. Cell Biol.** 10 (1998) 542-549.
11. Lahaye, K., Kricha, S. and Bellefroid, E.J. XNAP, a conserved ankyrin repeat-containing protein with a role in the Notch pathway during *Xenopus* primary neurogenesis. **Mech. Dev.** 110 (2002) 113-124.
12. Morrow, J.S., Rimm, D.L., Kennedy, S.P., Cianci, C.D., Sinard, J.H. and Weed, S.A. Of membrane stability and mosaics: the spectrin cytoskeleton. in: **Handbook of Physiology** (Hoffman, J. and Jamieson, J., Eds) London, Oxford University Press (1997) 485-540.
13. Lippincott-Schwartz, J. Cytoskeletal proteins and Golgi dynamics. **Curr. Opin. Cell Biol.** 10 (1998) 52-59.
14. Michaely, P. and Bennett, V. Mechanism for binding site diversity on ankyrin. Comparison of binding sites on ankyrin for neurofascin and the Cl⁻/HCO₃⁻ anion exchanger. **J. Biol. Chem.** 270 (1995) 31298-31302.
15. Wood, S.J. and Slater, C.R. Beta-spectrin is colocalized with both voltage-gated sodium channels and ankyrinG at the adult rat neuromuscular junction. **J. Cell Biol.** 140 (1998) 675-684.
16. Devarajan, P., Scaramuzzino, D.A. and Morrow, J.S. Ankyrin binds to two distinct cytoplasmic domains of Na,K-ATPase alpha subunit. **Proc. Natl. Acad. Sci. USA** 91 (1994) 2965-2969.
17. Bourguignon, L.Y.W., Chu, A., Jin, H. and Brandt, N.R. Ryanodine receptor-ankyrin interaction regulates internal Ca²⁺ release in mouse T-lymphoma cells. **J. Biol. Chem.** 270 (1995) 17917-17922.
18. Palek, J. and Lambert, S. Genetics of the red cell membrane skeleton. **Semin. Hematol.** 27 (1990) 290-332.
19. Sedgwick, S.G. and Smerdon, S.J. The ankyrin repeat: a diversity of interactions on a common structural framework. **Trends Biochem Sci.** 24 (1999) 311-316.
20. Mosavi, L.K., Williams, S. and Peng, Z.Y. Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein. **J. Mol. Biol.** 320 (2002) 165-170.
21. Bairoch, A., Bucher, P. and Hofmann, K. The PROSITE database, its status in 1997. **Nucleic Acids Res.** 25 (1997) 217-221.

22. Shapiro, M.B. and Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. **Nucleic Acids Res.** 15 (1987) 7155-7174.
23. Birkenmeier, C.S., White, R.A., Peters, L.L., Hall, E.J., Lux, S.E. and Barker, J.E. Complex patterns of sequence variation and multiple 5' and 3' ends are found among transcripts of the erythroid ankyrin gene. **J. Biol. Chem.** 268 (1993) 9533-9540.
24. Thevananther, S., Kolli, A.H. and Devarajan, P. Identification of a novel ankyrin isoform (AnkG190) in kidney and lung that associates with the plasma membrane and binds alpha-Na, K-ATPase. **J. Biol. Chem.** 273 (1998) 23952-23958.
25. Hooock, T.C., Peters, L.L. and Lux, S.E. Isoforms of ankyrin-3 that lack the NH2-terminal repeats associate with mouse macrophage lysosomes. **J. Cell Biol.** 136 (1997) 1059-1070.
26. Kunimoto, M., Otto, E. and Bennett, V. A new 440-kD isoform is the major ankyrin in neonatal rat brain. **J. Cell Biol.** 115 (1991) 1319-1331.