Published in final edited form as:
Genomics. 2002 January ; 79(1): 124–136. doi:10.1006/geno.2001.6679.

Sequence Analysis of LRPPRC and Its SEC1 Domain Interaction Partners Suggests Roles in Cytoskeletal Organization, Vesicular Trafficking, Nucleocytosolic Shuttling and Chromosome Activity

Leyuan Liu and **Wallace L. McKeehan**
Center for Cancer Biology and Nutrition, Institute of Biosciences and Technology, Texas A&M University System Health Science Center, 2121 W. Holcombe Blvd., Houston, TX 77030, USA

Abstract

LRPPRC (originally called LRP130) is an intracellular 130-kDa leucine-rich protein that co-purifies with the FGF receptor from liver cell extracts and has been detected in diverse multi-protein complexes from the cell membrane, cytoskeleton and nucleus. Here we report results of a sequence homology analysis of LRPPRC and its SEC1 domain interactive partners. Twenty-three copies of tandem repeats that are similar to PPR, TPR and HEAT repeats characterize the LRPPRC sequence. The N-terminus exhibits multiple copies of leucine-rich nuclear transport signals followed by ENTH, DUF28 and SEC1 homology domains. We used the SEC1 domain to trap interactive partners expressed from a human liver cDNA library. Interactive C19ORF5 (XP_038600) exhibited a strong homology to microtubule-associated proteins (MAP) and a potential arginine-rich mRNA binding motif. UXT (XP_033860) exhibited α-helical properties homologous to the actin-associated spectrin repeat and L/I heptad repeats in mobile transcription factors. C6ORF34 (XP_004305) was homologous to the non-DNA binding C-terminus of the *E. coli* Rob transcription factor. CECR2 (AAK15343) exhibited a transcription factor AT-hook motif next to two bromodomains and a homology to guanylate-binding protein 1. Taken together these features suggest a regulatory role of LRPPRC and its SEC1 domain-interactive partners in integration of cytoskeletal networks with vesicular trafficking, nucleocytosolic shuttling, chromosome remodeling and transcription.

Keywords

FGF receptor; LRP repeat; microtubule-associated proteins; CECR2; chromosome remodeling; exocytosis; secretion; mobile transcription factors

INTRODUCTION

Leucine-rich protein 130 (LRP130, approved HGN name LRPPRC) was first identified in detergent extracts of human hepatoma cells (HepG2) enriched for complexes of radiolabeled FGF and the FGF receptor (FGFR) [1]. It was first cloned as a 4782-bp cDNA that codes for a protein of 130 kDa containing more than 13 percent leucine residues [2], reported to GenBank (M92439) and the gene subsequently located on human chromosome 2 [3]. LRPPRC co-migrated with radiolabeled FGFR on wheat germ lectin affinity columns and appeared together with radiolabeled FGFR in the 125 to 150 kDa molecular weight range on

Corresponding author: Wallace L. McKeehan; telephone number: 713-677-7522; fax number: 713-677-7512; wmckeeha@ibt.tamu.edu.

* GenBank accession numbers: AF411610 for *C6ORF34B*; AF411609 for *CECR2B*.

polyacrylamide gel electrophoresis. Although sequence analysis indicated multiple potential translation initiation sites there was no evidence of a secretory signal sequence. Expression studies in mammalian cells confirmed that the translation product was an intracellular protein and exhibited no apparent carbohydrate to mediate direct binding to wheat germ lectin [2]. This suggested that the appearance in wheat germ lectin bound fractions was likely due to association with lectin-binding glycoproteins similar to FGFR. The mouse homologue of LRPPRC was subsequently reported among retinoic acid-responsive transcripts that were differentially expressed within 48 hours after treatment of P19 embryonic carcinoma cells to induce neuronal differentiation [4].

With advances in interaction cloning strategies and sequence detection methods, diverse investigators have hit GenBank sequence M92439 and reported by personal communication the presence of matching sequences in multi-protein complexes from diverse sources generated by diverse trapping and pull-down strategies. These include complexes formed with mitochondrial DNA binding proteins, sterol regulatory element binding protein SREBP1, nucleic acid-binding Replication Protein A, heterogenous nuclear ribonucleoprotein K (hnRNPK), nuclear mRNA, regulatory elements of the multi-drug resistance (MDR) gene, the cytoplasmic domain of C-cadherin and protein tyrosine phosphatases, several endocytosis-competent transmembrane receptors, the Arp2/3 complex of the actin cytoskeleton, and phosphoproteins that bound to immobilized vitamin E-succinate. This suggests a multi-functional role of LRPPRC that spans membrane, nuclear and possibly mitochondrial compartments.

Here we report results of a sequence analysis of LRPPRC and four interactive partners of unknown function derived from a human liver cDNA library in the yeast two-hybrid complementation system. Utilization of recent informatic search programs and genome sequence banks revealed that LRPPRC consists of 23 copies of pentatricopeptide (PPR)-, tetratricopeptide (TPR)- and HEAT (huntingtin-elongation A subunit-TOR)-like tandem repeat sequences. LRPPRC exhibits an N-terminal domain rich in nuclear import/export motifs, exhibits four copies of the transcriptional co-regulator signature LXXLL and distinct non-overlapping ENTH (Epsin1 N-Terminal Homology), DUF (Domain of Unknown Function) 28 and SEC (SECretory) 1 homology domains extending through its C-terminus. ENTH domains are most commonly associated with proteins involved in endocytosis and cytoskeletal organization, while SEC1 domains are usually involved in vesicular transport processes among cell compartments as synaptic transmission, vesicular exocytosis and general secretion.

A systematic sequence analysis of four LRPPRC-interactive proteins trapped by the SEC1 domain revealed features of proteins that interact with the microtubular and actin elements of the cytoskeleton, subcellular fractions packaged for transport or rearrangement, and chromatin-associated nucleic acid-binding proteins involved in regulation of transcription and mRNA processing. Taken together these results suggest that LRPPRC may be involved in integration of cytoskeletal actin and microtubule networks with vesicular trafficking, nucleocytosolic shuttling and nuclear chromosome remodeling and control of transcription.

RESULTS AND DISCUSSION

Origin and Tissue Distribution of LRPPRC mRNA

LRPPRC was first cloned as a candidate for the FGF receptor kinase (FGFR) from well-differentiated human hepatoma cells (HepG2) which exhibited about 5 to 10 times the number of radiolabeled FGF binding sites than most mammalian cells [1,2,5]. Northern hybridization analysis of mRNA detected three transcripts with sizes of 4.8, 5.2 and 7.0 kb in the HepG2 cells and confirmed that coincident with overexpression of the FGFR,

LRPPRC was expressed in the relatively non-malignant HepG2 cells at least ten times higher than in other more malignant human hepatoma cell lines. Expression in a highly malignant human melanoma cell line was undetectable [2]. Using the human cDNA probe, LRPPRC was undetectable in mouse tissues that was subsequently explained by the considerable difference in sequence between the human and mouse LRPPRC cDNAs [4]. Northern hybridization analysis confirmed the presence of the three species of mRNA in several other human tissues (Fig. 1). Expression was highest in heart, skeletal muscle, kidney and liver, with lesser amounts in brain, colon (none mucosal), spleen and placenta with the least expression in small intestine, thymus, lung and peripheral blood leukocytes.

Definition of the LRP Tandem Repeat

Previously only the high leucine content, of which one or more appears every 15 to 19 amino acid residues, a 30% homology to the consensus sequence for the ATP-binding site of kinases, and a potential protein kinase C site was noted by simple linear sequence homology analysis [2]. We subjected the LRPPRC sequence to a more rigorous analysis with the MEME motif discovery tool. Using the sequence of LRPPRC as a training set, the MEME output revealed 16 copies of motifs with a concensus sequence LLxAYxxxGNVExAxxI. Using these motifs as a base to do a manual alignment across the sequence with lower stringency homology requirement, we found 23 copies of a tandem repeat which consists of about 37 residues that have the consensus as its core. Analysis of the LRPPRC sequence with PHD program to predict secondary structure suggested that 70 percent of the protein might be α-helical. Examination of the PHD predictions in detail showed that two α-helices per repeat were distributed in 37-residue tandem repeats (Fig. 2A). Four residues that most commonly contain a glycine next to an acidic residue split the two helices in the repeat. From these results, we define the LRP repeat as the degenerate consensus sequence shown in Fig. 2A. The LRP repeats are similar to, but distinct from the tetratricopeptide (TPR), pentatricopeptide (PPR) and HEAT repeats (Fig. 2A). The LRP repeat was most similar to the PPR, which is a variant of TPR [6]. Tandem repeats of distinct helical repeats like TPR, PPR, ARM (armadillo) and HEAT that combine to create extended superhelical structures are thought to represent primitive structures which have been duplicated and evolved for diverse and specialized protein-protein recognition interfaces among diverse cellular compartments [6,7]. The most extensively studied example of the ARM repeat is the functionally pleiotropic β-catenin that includes membrane to nucleus signaling while interacting with a series of partners along the way [7]. The PPR repeat, distinct from the TPR repeat, constitutes over 200 *Arabidopsis* gene products, two thirds of which are targeted to mitochondria and chloroplasts [6]. Known RNA binding proteins are common among the products, the PPR repeat is well-suited to bind diverse RNA molecules, thus it was suggested that the diverse PPR repeat proteins may play a role in the peculiarities of RNA processing in plants. A gene product, BSF, has been recently identified in *Drosophila* that exhibits the PPR motif and significant homology to the specific repeats in LRPPRC. BSF binds to the 3'-untranslated region of *bicoid* mRNA and contributes to its stabilization [8]. HEAT motifs in particular have been refined for roles in specialized nuclear targeting apparatuses that respond to extranuclear events [7] and in chromosome dynamics during mitosis [9].

Homology Domain Structure of LRPPRC

Analysis and comparison of the LRPPRC sequence against multiple databases suggested four homology domains (Fig. 2B).

N-terminal Nuclear Export Signals (NES and LXXLL Repeats)—The first 250 residues of LRPPRC exhibit five of the 8 copies of the leucine-rich sequence LXXLXL. This and similar sequences have been identified as nuclear export signals (NES) [10]. The

sequence LXXLL of which there are two copies in the N-terminus and another two copies in the SEC1 domain of LRPPRC has also been identified as a nuclear co-activator/co-repressor signature [11].

Epsin N-terminal Homology (ENTH) Homology Domain—The Pfam HMM search revealed homology within residues 298 to 449 of LRPPRC to the consensus sequence of the 140-residue ENTH domain (Fig. 3A). Currently the ENTH domain is shared by 63 proteins in the Pfam database from different organisms with the most common functional feature that they are involved in endocytosis, vesicular trafficking and changes in cytoskeletal organization [12]. The ENTH domain in the N-terminus of the family's namesake, epsin (Fig. 3A), is followed by structural domains that bind the β-adaptin subunit of the clathrin adaptor protein AP2 and Eps15 which cooperate in its role in clathrin-mediated endocytosis [13]. Epsin 1 via its ENTH domain shuttles between cytosol and nucleus to interact with transcription factor promyelocytic leukemia Zn-finger protein (PLZF) [14]. The *Drosophila* homologue of the ENTH domain-containing AF10 interacts with heterochromatin protein HP1 and is thought to play a role in remodeling of heterochromatin and gene expression [15]. The ENTH domain of A180 (Fig. 3A) which is a component of the clathrin coat assembly binds phosphatidylinositol-4,5-bisphosphate [PtdIns(4,5)P2] [13]. The mammalian homologue, Huntingtin interacting protein 1 (Hip1), of SLA2p (Fig. 3A) that is essential for assembly of the cytoskeleton and endocytosis in yeast, is an actin-binding protein that colocalizes with clathrin, AP2, and endocytosed transferrin associated with clathrin-coated vesicles [13].

DUF28 Homology Domain—The region of LRPPRC from amino acid 676 to 841 was found similar to the DUF28 homology domain common to bacterial and yeast proteins (Fig. 3B). The average length of the predicted structural homology domain is about 230 residues and comprises the central core or in some cases the entire length of most of the proteins in the family. To date the function of proteins exhibiting the DUF28 (Domain of Unknown Function 28) homology domain of which there are currently 31 proteins in the Pfam database is unknown.

SEC1 Homology Domain—The C-terminal portion of LRPPRC from residue 779 to 1223 was identified as a SEC1 homology domain (Fig. 3C). The SEC1 domain defines a family of currently 54 proteins in the Pfam database from yeast, nematode, *Drosophila* and mammals. The homology of the LRPPRC SEC1 domain to several examples is shown in Fig. 3C. Both HUNC18B2 and STB1_BOVIN are mammalian homologues of the *Caenorhabditis elegans* syntaxin-binding protein UNC–18, which are involved in vesicular transport and neurotransmitter release in neurons. The *Drosophila* protein Rop, a homologue of the yeast Sec1p protein, is essential for general secretion and synaptic transmission. RVPS45, is a mammalian homologue of yeast VPS45 which is involved in vesicular transport from the Golgi complex to synaptic vesicles, while R–SLY1 is a homologue of yeast SLY1 involved in vesicle trafficking between endoplasmic reticulum and the Golgi. Both genetic and biochemical approaches generally indicate a role of the SEC1 superfamily in vesicular trafficking in the processes of exocytosis, synaptic transmission and general secretion. Rat brain Sec1 not only binds to the pre-synaptic membrane through a strong interaction with syntaxin, an integral component of the neural membrane required for vesicular fusion, but also to Cdk5, a neural cyclin-dependent kinase [16].

LRPPRC Interaction Partners

To derive further clues concerning the function of LRPPRC, cDNAs coding for segments of the different putative domain structures of LRPPRC as well as the full length cDNA were used as interaction traps for interactive substrates expressed from a human liver cDNA

library using the yeast two-hybrid complementation system (Materials and Methods). A portion of the putative SEC1 homology domain spanning amino acid residues 832 to 1018 yielded the largest number of strongly positive signals, and the positive signals were subjected to rigorous analysis and selection criteria. Of a total of 37 initial colonies that grew out in selective media, 28 survived serial culture. Comparison of *Alu*I digestion patterns of PCR amplified inserts of 23 colonies revealed 8 unique plasmids. Five products of cDNAs from the liver library that survived the foregoing selection process were judged significant of rigorous characterization based on reproducibility and strength of the signal to noise ratio of β-galactosidase activity which reports the interaction with the SEC1 domain of LRPPRC with cDNA product (Fig. 4A). A BLAST search of GenBank revealed that DNA and deduced amino acid sequence of part or all of the human genes coding for the significant interactive partners were present. The function of four interaction partners, C19ORF5 (XP_038600), C6ORF34 (XP_004305), UXT (XP_033860), and CECR2 (AAK15343), is largely unknown, while the fifth was fibronectin. All Trap B-interacting proteins except C6ORF34 passed the two-hybrid tests for ability to interact with the full-length LRPPRC protein (Fig 4B).

We probed the expression of those genes in several tissues using the commercially available array shown in Fig. 1 (Fig. 5). The *C19ORF5* gene displayed a major transcript of about 3 kb in all twelve tissues tested with additional less abundant transcripts apparent in some tissues. Consistent with its name and a previous report [17], the *UXT* gene was expressed as a single transcript of about 0.75 kb at near equal levels in all 12 tissues. The *C6ORF34* gene exhibited an mRNA at about 1 kb, which was most abundant in heart, skeletal muscle, kidney and liver, and least abundant in thymus and spleen. Analysis of the CECR2 gene product was problematic. Hybridization with three different probes from different sections of the RNA predicted from sequence resulted in a smear beginning at about 10 kb from all tissues that varied in intensity among tissues. From the noise, four distinct transcripts with sizes of about 2, 4, 6.5 and 9 kb were distinguishable in samples from placenta which exhibited the best distinct band to noise ratio in the lanes. This may suggest a particular lability of the CECR2 mRNA relative to the other mRNAs from the same membrane and particularly that in liver and small intestine. The CECR2B cDNA cloned in the current study was from an amplified liver cDNA expression library.

To derive clues about familial relationships and potential function of the unknown four interactive partners, each were analyzed by MAXHOM multiple sequence alignment and a 3D-PSSM search for theoretical structure-function patterns.

C19ORF5 (XP_038600)—The two-hybrid screen with the SEC1 domain of LRPPRC yielded a cDNA complementary to chromosome 19 open reading frame 5 (*C19ORF5*), encoding the last 393 residues of a 672-residue proline-rich (12.8%) protein reported to GenBank among transcripts expressed in NT2 neuronal precursor cells after a 2-week retinoic acid treatment. A MAXHOM search (Materials and Methods) revealed that the complete N- and C-terminal halves of C19ORF5 were homologous to stretches of sequence in both the N- and C-terminus of much longer microtubule-associated proteins (MAPs) (Fig. 6A). The N-terminal half (residues 1-322) of the protein shared 31% identity and 25% weighted similarity with the N-terminal region of rat MAP1A (Fig. 6B). The first 114 residues display 41% identity and 42% weighted similarity with the region covering residues 121 to 399 of rat MAP1A, and region 188-331 of electromotor neuron-associated protein 1 from *Torpedo californi*a which is colocalized with a subpopulation of post-Golgi synaptic vesicles [18] (Fig. 6B). The C-terminal half (residues 323-672) of the protein shares a 37 to 40% identity and 42 to 51% weighted similarity with the C-terminus of neuronal microtubule-associated proteins [19] (Fig. 6C). MAP1B, which is immunologically crossreactive with dystrophin, is a flexible filamentous molecule and bridges microtubules

and other components of the neuronal cytoskeleton [20]. Among the homologues are the tubulin-binding protein, neuraxin (amino acids 450-800) (Fig. 6C), whose C-terminus is immunologically related to a C-terminal fragment of MAP1B (also called MAP5) which exhibits low affinity binding sites for microtubules [21,22]. Cytoskeleton-associated protein MAP1A appears in RNA-binding protein complexes, and it has been suggested that some MAP family members interface with the microtubule system to guide mRNAs over large distances to various destination sites [23]. An arginine-rich sequence in MAP1 light chain 3 has been proposed to underlie its interaction with fibronectin mRNA [24]. We noted a similar arginine-rich, potential RNA-binding motif [25] at the N-terminus of the MAP homology in C19ORF5 (Fig. 6B). These observations suggest that C19ORF5 is a member of the MAP family that bridges LRPPRC to the microtubuler cytoskeleton and potentially mRNA binding and processing across cellular compartments.

UXT (XP_033860)—The two-hybrid screen yielded a cDNA encoding the full-length 157-residue UXT (Ubiquitously Expressed Transcript) gene product from Xp11 that is widely expressed, most abundantly in heart and tumor tissues [17]. A 3D-PSSM analysis showed that UXT is characterized by a three α-helix bundle with similarities to the spectrin repeat [26], which was confirmed by MAXHOM multiple sequence alignment (Fig. 7). Examples of the current 1426 spectrin repeat proteins in the Pfam database for which UXT showed highest similarity were in the region of the first spectrin repeat of dystrophin-related protein 2 (DRP2) [27], the N-terminal domain of syntaxin 1a [28] and the coiled-coil dimerization domain of cortexillin I [29]. DRP2 belongs to the broader family of actin-binding proteins that include spectrin, α-actinin and dystrophin [26]. Syntaxin 1a, a 288-residue plasma membrane protein and through its 3-α helix bundle, which is similar to that of UXT, interacts with multiple exocytic proteins including munc18 (yeast Sec1) that regulates vesicular neurotransmitter release [28]. Cortexillin I is a 444-residue actin-bundling protein whose actin-binding domains are located in the N-terminal portion which is then followed by a dimerization domain and phosphatidylinositol 4,5-bisphosphate (PIP2)-binding domains which are in the C-terminal portion [29]. The Cortexillin-UXT homology is in the dimerization domain.

The 3D-PSSM analyses also pointed to a homology of the helical region of UXT to the basic DNA-binding helix-loop-helix domain (bHLH-ZIP) of SREBP1 (sterol response element-binding protein 1 (NP_004167) [30,31]. UXT exhibited the tyr-335 signature in the basic DNA binding domain of SREBP1 that confers specificity for the sterol regulatory element (StRE) over the general E-box DNA motif [31]. The comparison revealed four L/I heptad repeats in the C-terminus of UXT, more perfectly spaced than those in SREBP1 (Fig. 7B). The search strategy also indicated that the helical pattern of UXT was homologous to the last three helices of the 4-helix bundle of transcription factor STAT (signal transducers and activators of transcription) 3b (P42227, residues 136-321) [32] (Fig. 7C). The helical bundle is required for the interaction of STAT3b to the DNA binding p48 subunit of the interferon-alpha-stimulated transcription factor ISGF3 [33]. Both SREBP and the STAT family are transcription factors that traffic from membrane locations to the nucleus. These similarities point to a potential role of UXT as a bridge between LRPPRC, actin-based cellular structures and potentially transcription complexes.

C6ORF34 (XP_004305)—The two-hybrid screen yielded a cDNA to chromosome 6 open reading frame 34 (*C6ORF34*), encoding the last 151 amino acids of the 205-residue human SOUL protein (from pineal gland, Decartes' proposed location of the SOUL) [34]. Characterization of the full-length transcript from liver revealed the presence of a variant of the *C6ORF34* (*C6ORF34B*, GenBank AF411610) that apparently resulted from exon skipping. The result is a deletion of amino acids 12 to 36 (Fig. 8). A 3D-PSSM search revealed with a greater than 90% confidence level that C6ORF34 adopts a similar folding

pattern with the non-DNA binding C-terminus of *E. coli* Rob transcription factor (Fig. 8). Rob is a member of the AraC/XylS family of bacterial transcription factors that regulate genes involved in resistance to antibiotics, organic solvents, heavy metals and superoxide-generating agents [35]. The non-DNA binding C-terminus of the AraC/XylS family members is thought to be involved in regulation of activity and specificity as dimerization domains or interaction domains with regulatory subunits of the transcription complex [36]. The expression of C6ORF34 is abundant in kidney, liver, heart and skeletal muscle, a pattern similar to LRPPRC (Figs. 1, 5). A homologue of C6ORF34 (HBP) was first isolated from liver on the basis of its heme-binding property that can be competed by non-tetrapyrrole ligands as palmitic acid and all-*trans*-retinoic acid [37]. These properties suggest a potential link of the LRPPCR-C6ORF34 to heme metabolism, redox status and oxygen sensing that may link to transcription complexes.

CECR2 (AAK15343)—The SEC1 domain of LRPPRC trapped the product of a liver cDNA, CECR2B (GenBank accession number AF411609), an alternatively spliced form of the deduced 1484-residue full length *CECR2* (Cat Eye Syndrome Chromosome Region, Candidate 2) protein (Fig. 9A). This isoform results from skipping of an exon encoding 29 putative amino acid residues resulting in a deletion of glycine-291 to glutamate-319. An alternative exon beginning after serine-518 encoded 8 novel C-terminal residues prior to truncation caused by a stop codon in the alternate exon. The fused CECR2B protein expressed from the cloned liver cDNA expression library starts at aspartate-169.

CECR2 is one of the few expressed sequences that maps within the region near the human chromosome 22 pericentromere that is linked to the multi-phenotype cat eye syndrome [38]. Combined HMM and 3D-PSSM searches suggested a homology domain structure for *CECR2* summarized in Fig. 9A. The 3D-PSSM search returned a predicted structural homology of N-terminal residues 6 to 372 of CECR2 to the C-terminal portion of human guanylate binding protein 1 (hGBP1) (Fig. 9B). GBP1 is among a family of large GTP-binding proteins whose prototypes are the microtubule-associated dynamins that coordinate microtubule organization and mediate vesicle trafficking [39]. Members of the family include the interferon-induced GBPs [40], yeast VP and yeast mitochondrial MGM1 thought to be involved in partitioning of mitochondrial DNA [39]. The N-terminal residues (6-372) of CECR2 (Fig. 9B) show a predicted structural homology with the extended α-helical C-terminal portion of human GBP1 that does not exhibit the correspondent GTP-binding domain, but corresponds to the Middle and GED (GTPase effector domain) domains of dynamin that are involved in assembly of the multi-subunit complexes involved in endocytosis, synaptic vesicle recycling, caveolae internalization and vesicular trafficking in general [40,41,42]. The Pfam search indicated a well-conserved AT-Hook motif within the GBP homology domain (CECR2 residues 198-210) (Fig. 9B, solid overline). The AT-Hook motif was first described as the DNA binding sequence in the mammalian high mobility group I (HMG1) chromosomal proteins. It forms a rigid basic surface with a conserved gly-arg-pro tripeptide (conserved also in CECR2) at its center and is thought to be a tether to grooves formed by AT-rich regions of DNA [43]. Following the AT-Hook motif spanning amino acid residues 221 to 564 of CECR2 are two bromodomains. The bromodomain is comprised of four helical bundles of about 110 residues each and interacts with chromatin-associated proteins involved in remodeling of chromosome structure and regulation of gene expression [44]. Proteins exhibiting the bromodomain frequently exhibit histone acetyltransferase activity (HAT) and in HAT co-activator P/CAF (p300/CBP-associated factor) the bromodomain appears to interact with acetyllysine that results from acetylation by the transferase [45]. An example of the two-bromodomain pattern in CECR2 was found in the 250 kDa subunit of the two-bromodomain TATA-binding-protein-associated factor (TFIID) (Fig. 9C). In TFIID, the two bromodomains interact selectively with diacetylated

histone H4 tails near the core transcriptional promoter and enhance formation of pre-initiation transcription complexes [46].

The 3D-PSSM indicated that a section of CECR2 (residues 764 to 1126) displayed a predicted structural homology of notably low E-value to the “all beta protein” class protein, *Arabidopsis thaliana* PAP-specific phosphatase. PAP-specific phosphatase is a prototype of an evolutionary conserved superfamily of more than 25 lithium-sensitive phosphatases whose role in sulfate metabolism indirectly impacts RNA processing, sulfation of sugars and phospholipid metabolism [47,48]. However, the homology based on predicted structure of the CECR2 protein from its sequence was species-specific and unique to the PAP-specific phosphatase from *Arabidopsis*. No homology could be detected to the core domain that is conserved across species and defines the broader phosphatase family. It should be noted here that the N-terminal 322 residues of C19ORF5 described earlier (Fig. 6) also exhibited a similar predicted structural homology to the *Arabidopsis* enzyme. Whether this suggests simply a similar structural fold, or is suggestive of a similar functional domain in the two LRPPRC SEC1 domain substrates is unclear.

The Pfam analysis also indicated that the C-terminal sequence 1177 to 1324 of CECR2 showed structural homology predicted from sequence with the C-terminus of poly (A)-binding protein 1 (PABP-C) [49]. This region does not contain the RNA binding domains (RBD), but is involved in oligomerization and interaction with other proteins to form both nuclear and cytosolic RNP complexes that are essential to mRNA function and metabolism [49,50]. On the one hand, the homology domain structure of CECR2 suggests an interface with cellular structures and vesicular trafficking akin to the dynamin family. On the other hand, roles associated with nucleic acid binding, chromosome remodeling and gene expression are suggested. This dual theme is common to both the homology domain structure of LRPPRC and its other SEC1 domain interactive partners.

Conclusions and Predictions

The leucine-rich repeat sequences in LRPPRC similar to the TPR, PPR and HEAT repeat structural modules, the LRPPRC homology domain structure and that of its SEC1 domain interactive substrates suggest that LRPPRC is well-equipped to play an integrative regulatory function through specific, but diverse, protein-protein interactions across cellular compartments. The leucine-rich sequence, particularly in the N-terminal domain, gives rise to potential nuclear transport signals and several LXXLL repeats important in transcription factor cofactor function. Homologies to the ENTH and SEC1 domains suggest functions associated with endocytosis/exocytosis, cytoskeletal organization and nucleocytosolic shuttles. Four of the five interactive proteins trapped from a human liver cDNA expression library are consistent with this theme, but also point to a role in chromosome remodeling, transcription, and the processing and transport of RNA. A fundamental role in assembly and remodeling of actin and microtubule elements of the basic cytoskeleton that underlies cell structure and controls shape and motility cannot be excluded. Although these are multiple possibilities to which the major function of LRPPRC may be limited to one or more, they are not mutually exclusive and the results may indicate a multi-functional LRPPRC that integrates the different functional domains. It is noteworthy that the greatest majority of the homology of LRPPRC and its interactive partners was to modular segments of proteins of known function that are known or predicted to be protein-protein interaction domains rather than the active site that describes the specific function of the homologue. Conceivably, LRPPRC and one or more of its partners in association with both actin and microtubular elements of the cytoskeleton traffics from membrane locations to affect transcription while picking up nuclear cargo for export and participating in vesicular distribution of both membrane and nuclear organelles. In addition to messengers released from intracellular membrane stores, there is an increasing body of evidence that endocytosis and endosomes

bearing internalized plasma membrane components play an active signaling role in addition to a mechanism of down regulating surface signals [51]. A number of extracellular signal polypeptides and their receptors internalized by endocytosis appear in the nucleus, although their precise function there remains obscure [52]. This includes members of the FGF family and the FGFR complex [53,54] with which LRPPRC appears in cell extracts [1,2, unpublished results]. Whether LRPPRC and its interactive partners are involved in FGFR signaling or metabolism of the oligomeric complex of FGFR kinase/heparan sulfate/FGF is under investigation.

The diverse roles implicated by sequence homology of LRPPRC and interaction partners are consistent with the presence of the LRPPRC sequence in the multi-protein complexes reported to us by over 16 personal communications based on GenBank hits of LRPPRC sequence M92439. These include complexes clustered around components involved in transcription and DNA and RNA binding, complexes clustered around the cytoplasmic domain of several different membrane protein complexes both prior to and during endocytosis, and multi-subunit complexes of cytoskeletal components. The secreted extracellular matrix protein fibronectin was a fifth significant interaction partner trapped in the two-hybrid screen with the SEC1 domain of LRPPRC (Fig. 4). Whether this suggests a role of LRPPRC in secretion, endocytosis or intracellular trafficking of fibronectin, or a fortuitous interaction between structural modules is unclear.

Lastly, preliminary screens for interactors with the four primary LRPPRC substrates of unknown function reveal proteins whose functions are known in part and would bridge LRPPRC through its primary interaction substrates to elements of the actin and microtubular cytoskeleton, chromatin and the kinetochore, mRNA maturation and transport, redox pathways and apoptotic signaling. These results to be published elsewhere together with those presented here form a conceptual framework for experimental testing of the multiple roles of LRPPRC and its interactive partners at molecular and cellular levels.

MATERIALS AND METHODS

Tissue Distribution of LRPPRC and Its Interactive Partners

Poly (A^+) blots with 2 μg per lane RNA from 12 different human tissues were obtained from CLONTECH. The cDNA probes were radiolabeled with [α-^{32}P]dCTP (3000 Ci/mmol; ICN Pharmaceuticals, Costa Mesa, CA) using the Prime-It random primer labeling kit (Stratagene, La Jolla, CA) according to the manufacturer's directions. We performed hybridization as described in the manual provided with ExpressHyb hybridization solution (CLONTECH, Palo Alto, CA). We used an β-actin probe to compensate for variable mRNA load among lanes. The cDNA probe for *LRPPRC* in Fig. 1 encoded LRPPRC amino acid residues 846 to 1030 (2539-3091 bp of the 4782 bp M92439, trap B in the two-hybrid screen). Probes used in Fig. 5 covered bp 1192-2454 of the 2454 bp *C19ORF5* (AK023118), bp 55-525 of the 574 bp *UXT* (NM_004182), and bp 274-888 of the 1137 bp *C6ORF34* (NM_014320), respectively. Three *CECR2* probes covered bp 916-1261 and bp 1370-1850 of the 6178 bp *CECR2* (AF336133) and bp 1-1453 of the *CECR2B* (AF411609).

Yeast Two-hybrid Screening System

The sequence coding for Trap B (residues 832-1018) within the SEC1 homology domain of LRPPRC cDNA was first amplified by PCR and subcloned into yeast plasmid pGBKT7 (CLONTECH, Palo Alto, CA). *Bam*HI and *Nco*I restriction sites were incorporated into primers to create an in-frame fusion with the GAL4 DNA-binding domain and c-myc-tag. We transformed the pGBKT7 plasmid with the Trap B insert into yeast strain AH109 using the lithium acetate procedure and plated onto minimal, synthetic dropout media lacking

tryptophan (SD-Trp). The expression of Trap B fusion protein was confirmed by western blotting using anti-c-myc monoclonal antibody. The possibility that Trap B activated transcription and limited yeast strain growth by toxicity was tested and eliminated. Following the CLONTECH manual, 50 ml of saturated culture of yeast strain AH109 bearing Bait B was collected and combined with 1 ml of pretransformed human liver cDNA library in yeast strain Y187, yeast from the mating culture was harvested and spread on plates with QDO media (SD/-Ade/-His/-Leu/-Trp). The transformed yeast cells were grown at 30°C for 8 to 16 days. Transformants grown on the QDO plates were streaked and re-streaked to new QDO plates with X-β-gal to reveal positive colonies. The pBGKT7 plasmid with the interaction trap sequence (kanamycin-resistant) and library plasmid DNA (ampicillin-resistant) mixture was isolated from positive yeast colonies. The library plasmid was recovered by introducing into *Escherichia coli* DH5α cells by electroporation under the selection pressure of ampicillin. Plasmid DNAs were sequenced and identified by Blast search. The positive clones were confirmed by retesting the Trap B-protein interaction through cotransformation of the interaction trap and library DNA plasmids into AH109 yeast. The strength of the interaction was quantified by β-galactosidase activity as directed by The Yeast Handbook from CLONTECH. Trap B-interacting substrates were tested for interaction with full-length LRPPRC by co-transformation with cDNAs for Trap B-interactors with a cDNA in which the coding sequence for full-length LRPPRC was fused with the DNA binding domain and ability of co-transformants to grow on QDO plates.

Database Searches and Informatic Analysis

Repeat motif searches were performed using motif discovery tool MEME (Multiple Expectation-Maximization for Motif Elicitation, http://meme.sdsc.edu/meme/website/). MEME represents motifs, a sequence pattern that occurs repeatedly in a group of related protein sequences, as position-dependent letter-probability matrices describing the probability of each possible letter at each position in the pattern. MEME uses statistical modeling techniques to automatically choose the best width and description for each motif [55]. Secondary structure was predicted by PHD, which includes protein family information in the prediction [56, http://www.embl-heidelberg.de/predictprotein/]. Assignment of homology domains was based on Pfam HMM search results. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. Alignments are based on some evolutionary conserved structure thought to have implications for protein function. Profile-hidden Markov models (profile HMMs) built from the Pfam alignments were employed to recognize that a new protein belongs to an existing family, even if the homology is weak. Unlike standard pairwise alignment methods as BLAST and FASTA, Pfam HMMs deal rationally with multidomain proteins [57, http://pfam.wustl.edu/].

New amino acid and DNA sequences generated from the yeast two-hybrid system were initially subjected to BLAST search [58], then to the MAXHOM multiple sequence alignment and 3D-PSSM (three-dimensional position-specific scoring matrix) search methods. MAXHOM scans protein sequence databases for similarities that are significant in terms of protein structure or function [http://www.embl-heidelberg.de/predictprotein]. Homology using MAXHOM was expressed as percentage of pair-wise sequence identity and percentage of weighted similarity of residues that is a function of alignment length [59]. The 3D-PSSM combines the power of multiple sequence profiles with secondary structure matching and solvation potentials, therefore, can recognize structural and functional relationships beyond amino acid sequence [60, http://www.bmm.icnet.uk/~3dpssm].

REFERENCES

1. DiSorbo D, Shi EG, McKeehan WL. Purification form human hepatoma cells of a 130-kDa membrane glycoprotein with properties of the heparin-binding growth factor receptor. Biochem. Biophys. Res. Commun. 1988; 157:1007–1014. [PubMed: 2462864]
2. Hou J, Wang F, McKeehan WL. Molecular cloning and expression of the gene for a major leucine-rich protein from human hepatoblastoma cells (HepG2). In Vitro Cell. Dev. Biol. Anim. 1994; 30A: 111–114. [PubMed: 8012652]
3. Ghiso NS, Lennon GG. LRP130 gene assigned to chromosome 2. In Vitro Cell. Dev. Biol. Anim. 1994; 30A:744. [PubMed: 7881626]
4. Suzuki Y, Wanaka A, Tohyama M, Takagi T. Identification of differentially expressed mRNAs during neuronal differentiation of P19 embryonal carcinoma cells. Neurosci. Res. 1995; 23:65–71. [PubMed: 7501302]
5. Kan M, et al. High and low affinity binding of heparin-binding growth factor to a 130-kDa receptor correlates with stimulation and inhibition of growth of a differentiated human hepatoma cell. J. Biol. Chem. 1988; 263:11306–11313. [PubMed: 2457020]
6. Small ID, Peeters N. The PPR motif -- a TPR-related motif prevalent in plant organellar proteins. TIBS. 2000; 25:45–47.
7. Groves MR, Barford D. Topological characteristics of helical repeat proteins. Curr. Opin. Struct. Biol. 1999; 9:383–389. [PubMed: 10361086]
8. Mancebo R, Zhou X, Shillinglaw W, Henzel W, Macdonald PM. BSF binds specifically to the *bicoid* mRNA 3' untranslated region and contributes to stabilization of *bicoid* mRNA. Mol. Cell. Biol. 2001; 21:3462–3471. [PubMed: 11313472]
9. Neuwald AF, Hirano T. HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. Genome Res. 2000; 10:1445–1452. [PubMed: 11042144]
10. Stommel JM, et al. A leucine-rich nuclear export signal in the p53 tetramerization domain: regulation of subcellular localization and p53 activity by NES masking. EMBO J. 1999; 18:1660–1672. [PubMed: 10075936]
11. Leo C, Chen JD. The SRC family of nuclear receptor coactivators. Gene. 2000; 245:1–11. [PubMed: 10713439]
12. Kay BK, Yamabhai M, Wendland B, Emr SD. Identification of a novel domain shared by putative components of the endocytic and cytoskeletal machinery. Prot. Sci. 1999; 8:435–438.
13. Itoh T, et al. Role of the ENTH domain in phosphatidylinositol-4,5-bisphosphate binding and endocytosis. Science. 2001; 291:1047–1051. [PubMed: 11161217]
14. Hyman J, Chen H, Di Fiore PP, De Camilli P, Brunger AT. Epsin 1 undergoes nucleocytosolic shuttling and its eps15 interactor NH(2)-terminal homology (ENTH) domain, structurally similar to Armadillo and HEAT repeats, interacts with the transcription factor promyelocytic leukemia Zn(2)+ finger protein (PLZF). J. Cell Biol. 2000; 149:537–546. [PubMed: 10791968]
15. Linder B, Gerlach N, Jackle H. The *Drosophila* homolog of the human AF10 is an HP1-interacting suppressor of position effect variegation. EMBO Rep. 2001; 2:211–216. [PubMed: 11266362]
16. Halachmi N, Lev Z. The Sec1 family: a novel family of proteins involved in synaptic transmission and general secretion. J. Neurochem. 1996; 66:889–897. [PubMed: 8769846]
17. Schroer A, Schneider S, Ropers H, Nothwang H. Cloning and characterization of UXT, a novel gene in human Xp11, which is widely and abundantly expressed in tumor tissue. Genomics. 1999; 56:340–343. doi:10.1006/geno.1998.5712. [PubMed: 10087202]
18. Ngsee JK, Scheller RH. Isolation and characterization of two homologous cDNA clones from Torpedo electromotor neurons. DNA. 1989; 8:555–561. [PubMed: 2480872]
19. Tucker RP. The roles of microtubule-associated proteins in brain morphogenesis: a review. Brain Res., Brain Res. Reviews. 1990; 15:101–120.
20. Lien LL, et al. Mapping of human microtubule-associated protein 1B in proximity to the spinal muscular atrophy locus at 5q13. Proc. Natl. Acad. Sci. USA. 1991; 88:7873–7876. [PubMed: 1881920]

21. Rienitz A, et al. Neuraxin, a novel putative structural protein of the rat central nervous system that is immunologically related to microtubule-associated protein 5. EMBO J. 1989; 8:2879–2888. [PubMed: 2555150]
22. Noble M, Lewis SA, Cowan NJ. The microtubule binding domain of microtubule-associated protein MAP1B contains a repeated sequence motif unrelated to that of MAP2 and tau. J. Cell Biol. 1989; 109:3367–3376. [PubMed: 2480963]
23. Jansen RP. RNA-cytoskeletal associations. FASEB J. 1999; 13:455–466. [PubMed: 10064612]
24. Zhou B, Boudreau N, Coulber C, Hammarback J, Rabinovitch M. Microtubule-associated protein 1 light chain 3 is a fibronectin mRNA-binding protein linked to mRNA translation in lamb vascular smooth muscle cells. J. Clin. Invest. 1997; 100:3070–3082. [PubMed: 9399954]
25. Burd CG, Dreyfuss G. Conserved structures and diversity of functions of RNA-binding proteins. Science. 1994; 265:615–621. [PubMed: 8036511]
26. Thomas GH, et al. Intragenic duplication and divergence in the spectrin superfamily of proteins. Mol. Biol. Evol. 1997; 14:1285–1295. [PubMed: 9402739]
27. Roberts RG, et al. Characterization of DRP2, a novel human dystrophin homologue. Nat. Genet. 1996; 13:223–226. [PubMed: 8640231]
28. Fernandez I, et al. Three-dimensional structure of an evolutionarily conserved N-terminal domain of syntaxin 1A. Cell. 1998; 94:841–849. [PubMed: 9753330]
29. Faix J, et al. Cortexillins, major determinants of cell shape and size, are actin-bundling proteins with a parallel coiled-coil tail. Cell. 1996; 86:631–642. [PubMed: 8752217]
30. Sato R, et al. Assignment of the membrane attachment, DNA binding, and transcriptional activation domains of sterol regulatory element-binding protein-1 (SREBP-1). J. Biol. Chem. 1994; 269:17267–17273. [PubMed: 8006035]
31. Parraga A, Bellsolell L, Ferre-D'Amare AR, Burley SK. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution. Structure. 1998; 6:661–672. [PubMed: 9634703]
32. Becker S, Groner B, Muller CW. Three-dimensional structure of the Stat3beta homodimer bound to DNA. Nature. 1998; 394:145–151. [PubMed: 9671298]
33. Martinez-Moczygemba M, Gutch MJ, French DL, Reich NC. Distinct STAT structure promotes interaction of STAT2 with the p48 subunit of the interferon-alpha-stimulated transcription factor ISGF3. J. Biol. Chem. 1997; 272:20070–20076. [PubMed: 9242679]
34. Zylka MJ, Reppert SM. Discovery of a putative heme-binding protein family (SOUL/HBP) by two-tissue suppression subtractive hybridization and database searches. Mol. Brain Res. 1999; 74:175–181. [PubMed: 10640688]
35. Kwon HJ, Bennik MH, Demple B, Ellenberger T. Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. Nat. Struct. Biol. 2000; 7:424–430. [PubMed: 10802742]
36. Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL. Arac/XylS family of transcriptional regulators. Microbiol. Mol. Biol. Rev. 1997; 61:393–410. [PubMed: 9409145]
37. Taketani S, et al. Molecular characterization of a newly identified heme-binding protein induced during differentiation of murine erythroleukemia cells. J. Biol. Chem. 1998; 273:31388–31394. [PubMed: 9813049]
38. Footz TK, et al. Analysis of the cat eye syndrome critical region in humans and the region of conserved synteny in mice: A search for candidate genes at or near the human chromosome 22 pericentromere. Genome Res. 2001; 11:1053–1070. [PubMed: 11381032]
39. Urrutia R, Henley JR, Cook T, McNiven MA. The dynamins: redundant or distinct functions for an expanding family of related GTPases? Proc. Natl. Acad. Sci. USA. 1997; 94:377–384. [PubMed: 9012790]
40. Prakash B, Renault L, Praefcke GJ, Herrmann C, Wittinghofer A. Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. EMBO J. 2000; 19:4555–4564. [PubMed: 10970849]
41. van der Bliek AM. Functional diversity in the dynamin family. Trends Cell Biol. 1999; 9:96–102. [PubMed: 10201074]
42. Hinshaw JE. Dynamin and its role in membrane fission. Annu. Rev. Cell Dev. Biol. 2000; 16:483–519. [PubMed: 11031245]

43. Aravind L, Landsman D. AT-hook motifs identified in a wide variety of DNA-binding proteins. Nucleic Acids Res. 1998; 26:4413–4421. [PubMed: 9742243]
44. Jones MH, Hamana N, Nezu J-i. Shimane M. A novel family of bromodomain genes. Genomics. 2000; 63:40–45. doi:10.1006/geno.1999.6071. [PubMed: 10662543]
45. Dhalluin C, et al. Structure and ligand of a histone acetyltransferase bromodomain. Nature. 1999; 399:491–496. [PubMed: 10365964]
46. Jacobson RH, Ladurner AG, King DS, Tjian R. Structure and function of a human TAFII250 double bromodomain module. Science. 2000; 288:1422–1425. [PubMed: 10827952]
47. York JD, Ponder JW, Majerus PW. Definition of a metal-dependent/Li(+)-inhibited phosphomonoesterase protein family based upon a conserved three-dimensional core structure. Proc. Natl. Acad. Sci. USA. 1995; 92:5149–5153. [PubMed: 7761465]
48. Lopez-Coronado JM, Belles JM, Lesage F, Serrano R, Rodriguez PL. A novel mammalian lithium-sensitive enzyme with a dual enzymatic activity, 3'-phosphoadenosine 5'-phosphate phosphatase and inositolpolyphosphate 1-phosphatase. J. Biol. Chem. 1999; 274:16034–16039. [PubMed: 10347153]
49. Kozlov G, et al. Structure and function of the C-terminal PABC domain of human poly(A)-binding protein. Proc. Natl. Acad. Sci. USA. 2001; 10 1073/pnas.071024998v1.
50. Kuhn U, Pieler T. Xenopus poly(A) binding protein: functional domains in RNA binding and protein-protein interaction. J. Mol. Biol. 1996; 256:20–30. [PubMed: 8609610]
51. Wiley HS, Burke PM. Regulation of receptor tyrosine kinase signaling by endocytic trafficking. Traffic. 2001; 2:12–18. [PubMed: 11208164]
52. Jans DA, Hassan G. Nuclear targeting by growth factors, cytokines, and their receptors: a role in signaling? Bioessays. 1998; 20:400–411. [PubMed: 9670813]
53. Prudovsky IA, Savion N, LaVallee TM, Maciag T. The nuclear trafficking of extracellular fibroblast growth factor (FGF)-1 correlates with the perinuclear association of the FGF receptor-1alpha isoforms but not the FGF receptor-1beta isoforms. J. Biol. Chem. 1996; 271:14198–14205. [PubMed: 8662999]
54. Feng S, Xu J, Wang F, Kan M, McKeehan WL. Nuclear localization of a complex of fibroblast growth factor(FGF)-1 and an NH2-terminal fragment of FGF receptor isoforms R4 and R1alpha in human liver cells. Biochim. Biophys. Acta. 1996; 1310:67–73. [PubMed: 9244177]
55. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. Bioinformatics. 1998; 14:48–54. [PubMed: 9520501]
56. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. J. Mol. Biol. 1993; 232:584–599. [PubMed: 8345525]
57. Bateman A, et al. The Pfam protein families database. Nucleic Acids Res. 2000; 28:263–266. [PubMed: 10592242]
58. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25:3389–3402. [PubMed: 9254694]
59. Sander C, Schneider R. Database of homology derived protein structures and the structural meaning of sequence alignment. Proteins. 1991; 9:56–68. [PubMed: 2017436]
60. Kelley LA, MacCallum RM, Sternberg MJE. Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol. 2000; 299:501–522.

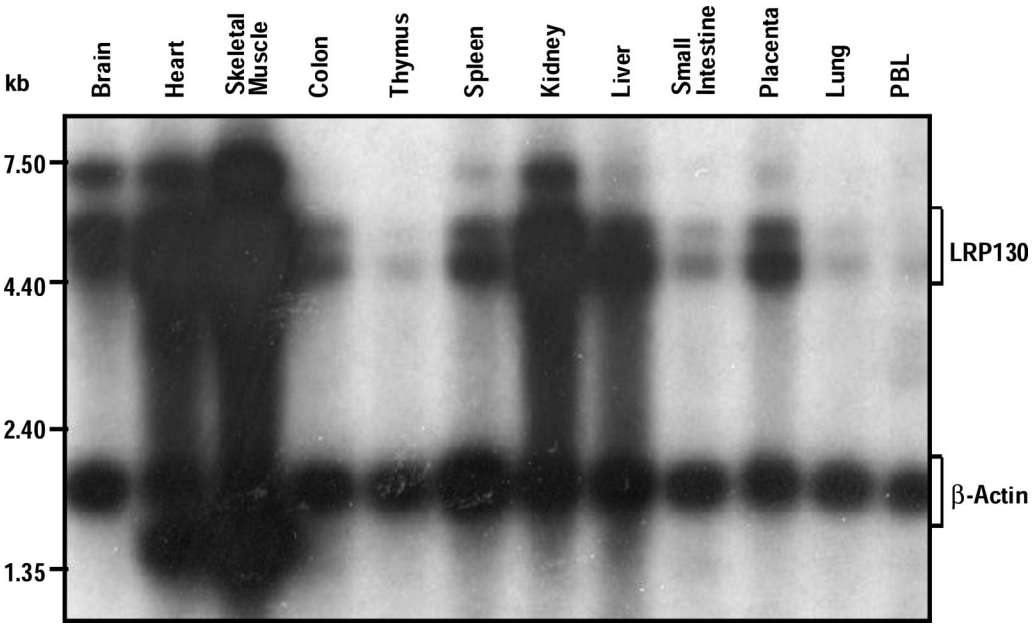

FIG. 1.

Expression of LRPPRC mRNA. Poly(A) mRNA from the indicated tissues was subjected to Northern hybridization analysis with an LRPPRC cDNA spanning the coding region employed in the yeast two-hybrid screen (Trap B, Fig. 2B) and β-actin cDNA as described in Materials and Methods. PBL, peripheral blood leukocytes.

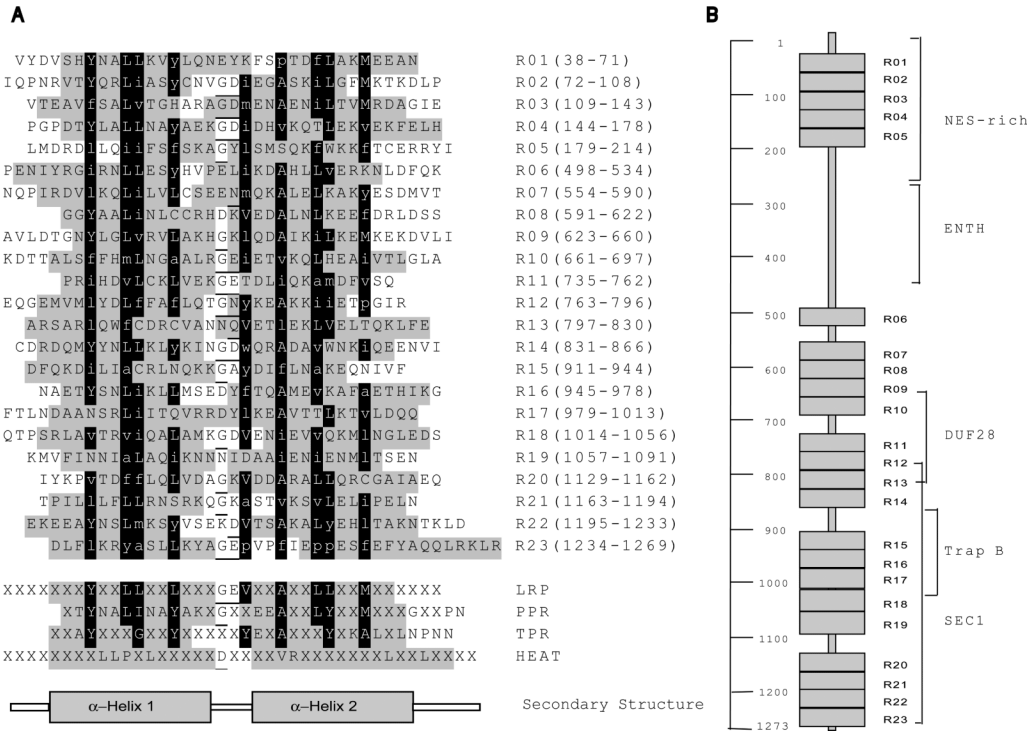

FIG. 2.

Domain structure of LRPPRC predicted from sequence. (**A**) Homology and secondary structure of the LRP tandem repeats. Repeat sequences and inclusive residues are numbered at right. Residues within an α-helix predicted by PHD are shaded and conserved residues according to type of amino acid are in black. A consensus LRP repeat is indicated showing the most common residue at a conserved site throughout the 23 repeats. Similar, but not identical, residues are indicated in lower case. The consensus, most commonly inter-helical, gly and basic residues (E/D) are underlined. The consensus TPR [6], PPR [6] and HEAT [7] repeat is also indicated. (**B**) Schematic of the LRP repeat and sequence homology domain structure of LRPPRC.

FIG. 3.

Predicted homologies to ENTH, DUF28 and SEC1 structural domains. Examples and alignments are from the Pfam search. The GenBank entry name of examples is indicated followed by the inclusive range of sequence compared. Residues in black indicate identity and shaded residues indicate amino acids of similar property in respect to hydrophobicity, charge and/or size. (**A**) ENTH family proteins. (**B**) DUF28 family proteins. (**C**) SEC1 family proteins.

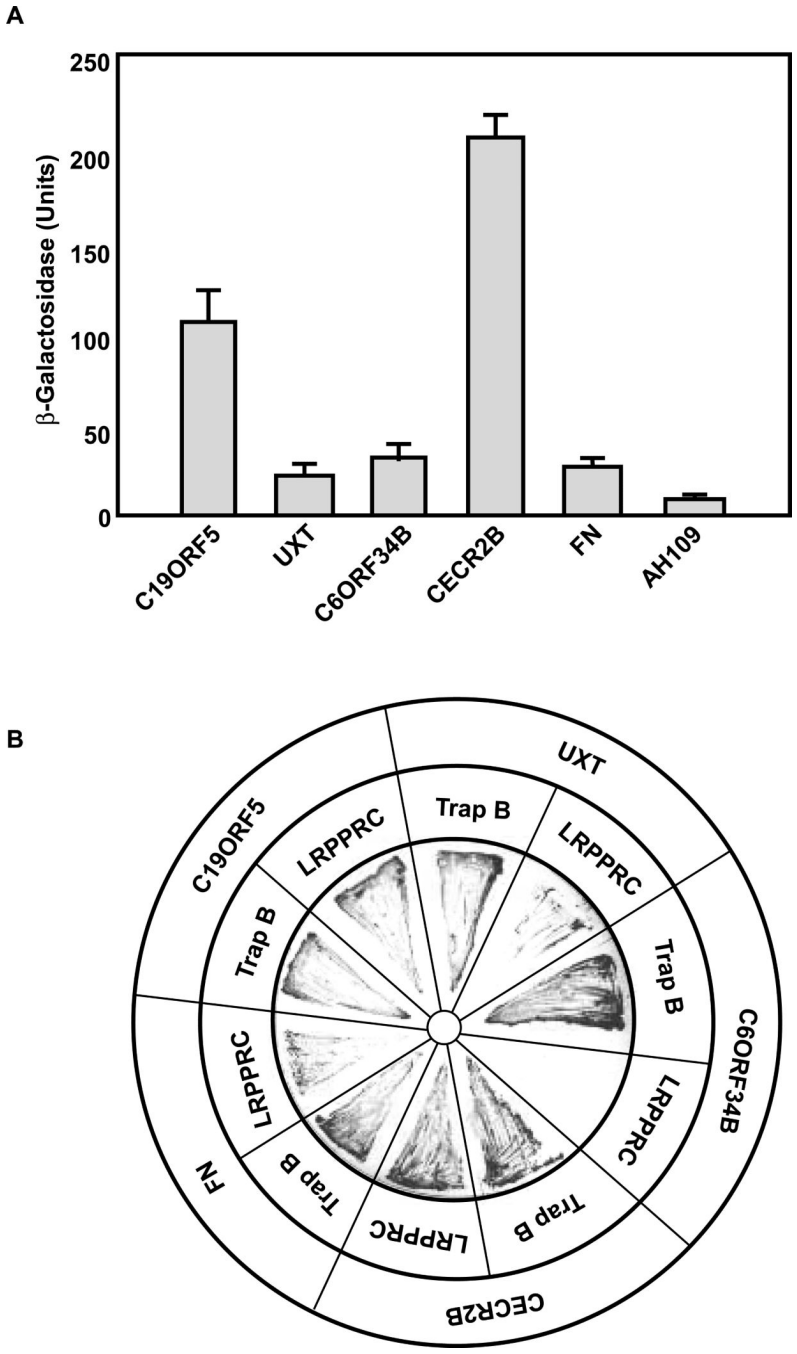

FIG. 4.

Interaction of expression products from cloned human liver cDNAs LRPPRC in the yeast two-hybrid system. (A) Interaction with Trap B containing the SEC1 subdomain of LRPPRC. The five plasmids carrying genes coding for C19ORF5, UXT, C6ORF34B, CECR2B and fibronectin (FN) fused with the activation domain that survived the positive selection series in the yeast two-hybrid complementation system described in the text were co-transformed with the pGBKT7-BD-trap B and full-length LRPPRC fusion construct into yeast AH109 cells. The co-transformants were initially grown on SD/-Leu-Trp plates and re-streaked on QDO plates (Materials and Methods). The β-galactosidase activity in cell lysates was measured spectrophotometrically using the substrate *o*-nitrophenyl-β-D-

galactopyranoside. The data indicated is the mean (± SE) of three independent co-transformations. (B) Interaction of SEC1 interactive substrates with full length LRPPRC. The interaction of the five gene products with the Trap B and entire LRPPRC were compared by ability of co-transformed yeast to grow on QDO plates.

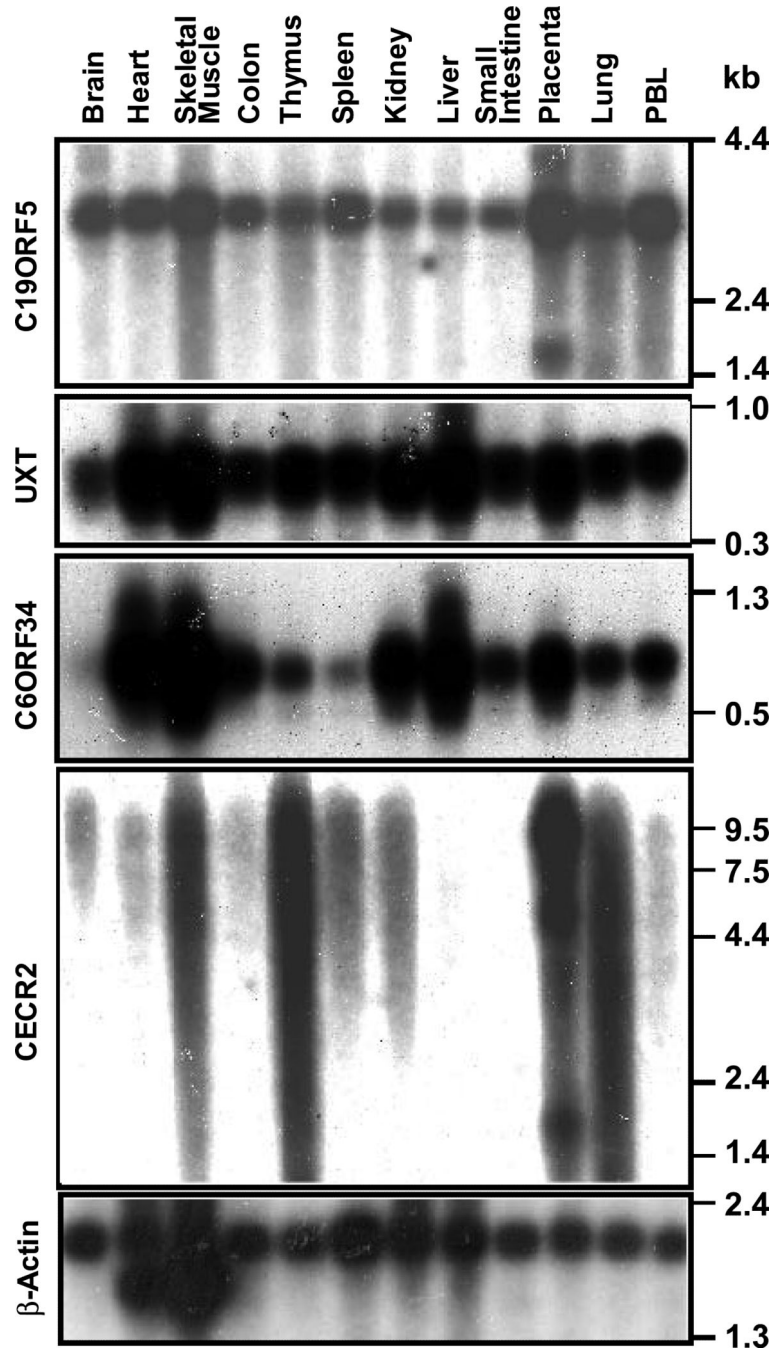

FIG. 5.

Expression of the LRPPRC-interacting partners *C19ORF5*, *UXT*, *C6ORF34*, *CECR2* and control β-actin mRNAs in multiple human tissues. Specific probes were described in Materials and Methods. Three *CECR2* probes were employed with similar results.

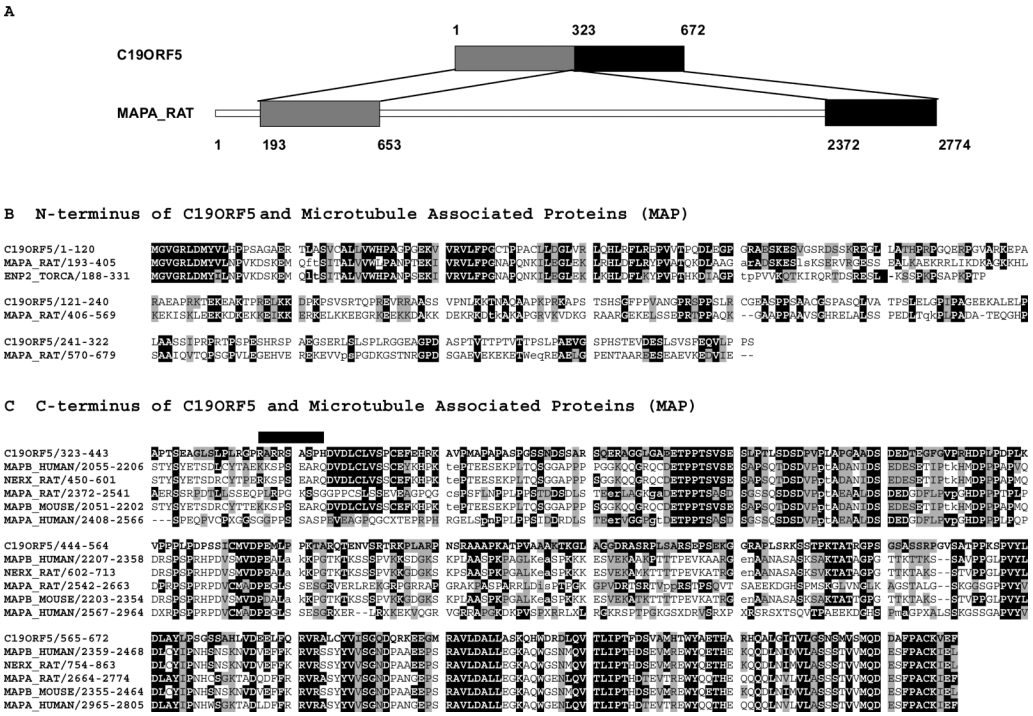

FIG. 6.

Homology of C19ORF5 to microtubule-associated proteins (MAP). (A) Schematic of domain homology to MAP1A (MAPA_RAT). (B) Homology of the N-terminus of C19ORF5. (C) The C-terminus of C19ORF5. The solid bar indicates a potential RNA binding motif. The indicated alignment was based on the MAXHOM multiple sequence alignment program. The GenBank entry name of examples is indicated followed by the inclusive range of sequence compared. Residues in black indicate identity and shaded residues indicate amino acids of similar property in respect to hydrophobicity, charge and/or size. Identical and similar residues are bolded or shaded, respectively. Two lowercase residues indicate an omission of a stretch of sequence with insignificant homology between the residues.

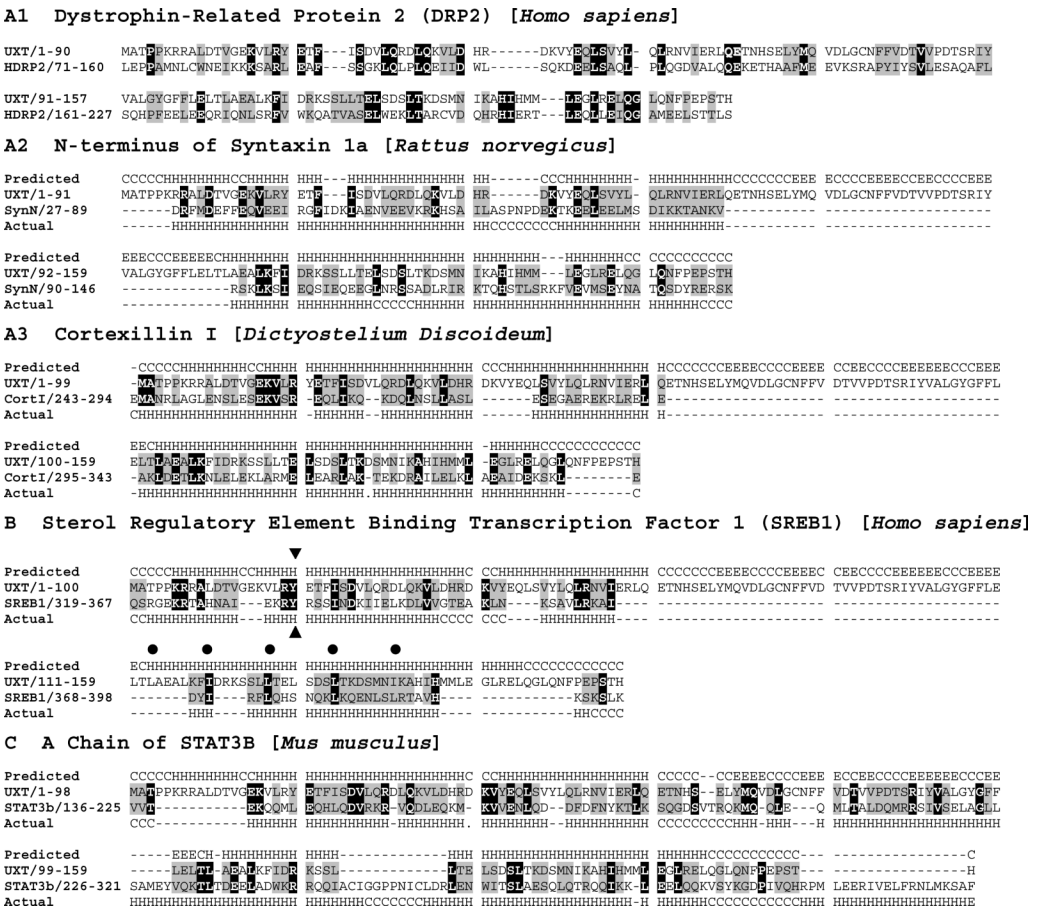

FIG. 7.

Examples of UXT homology to spectrin repeat-containing proteins and two membrane-bound transcription factors. (A) Spectrin repeat proteins. Alignment to 954-residue human dystrophin-related protein 2 (DRP2) was based on the MAXHOM multiple sequence alignment program, while that to 288-residue syntaxin 1a and 444-residue cortexillin I was predictions of structural homology based on sequence using the 3D-PSSM program. (B, C) Mobile transcription factors SREB1 and STAT3b. Full-length SREB1 and STAT3b consist of 1147 and 770 residues, respectively. Alignment was also based on the 3D-PSSM program. The signature tyr-335 in the basic region of SREB1 is noted by triangle and the L/I residues of the four heptad repeats in UXT are noted by solid circle. Residue identity and similarities are indicated in black and gray, respectively. The secondary structure (C, loop; E, β-strand; and H, α-helix) to which each residue is predicted to contribute is indicated above each residue. The actual secondary structure from x-ray or NMR for each homologue is indicated below each residue

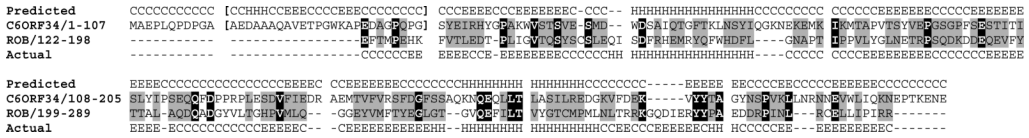

FIG. 8.

Homology of C6ORF34 to the regulatory domain of the *E. coli* transcriptional factor Rob. The alignment is based on the 3D-PSSM analysis with the predicted and actual structures indicated as described in Fig. 7.

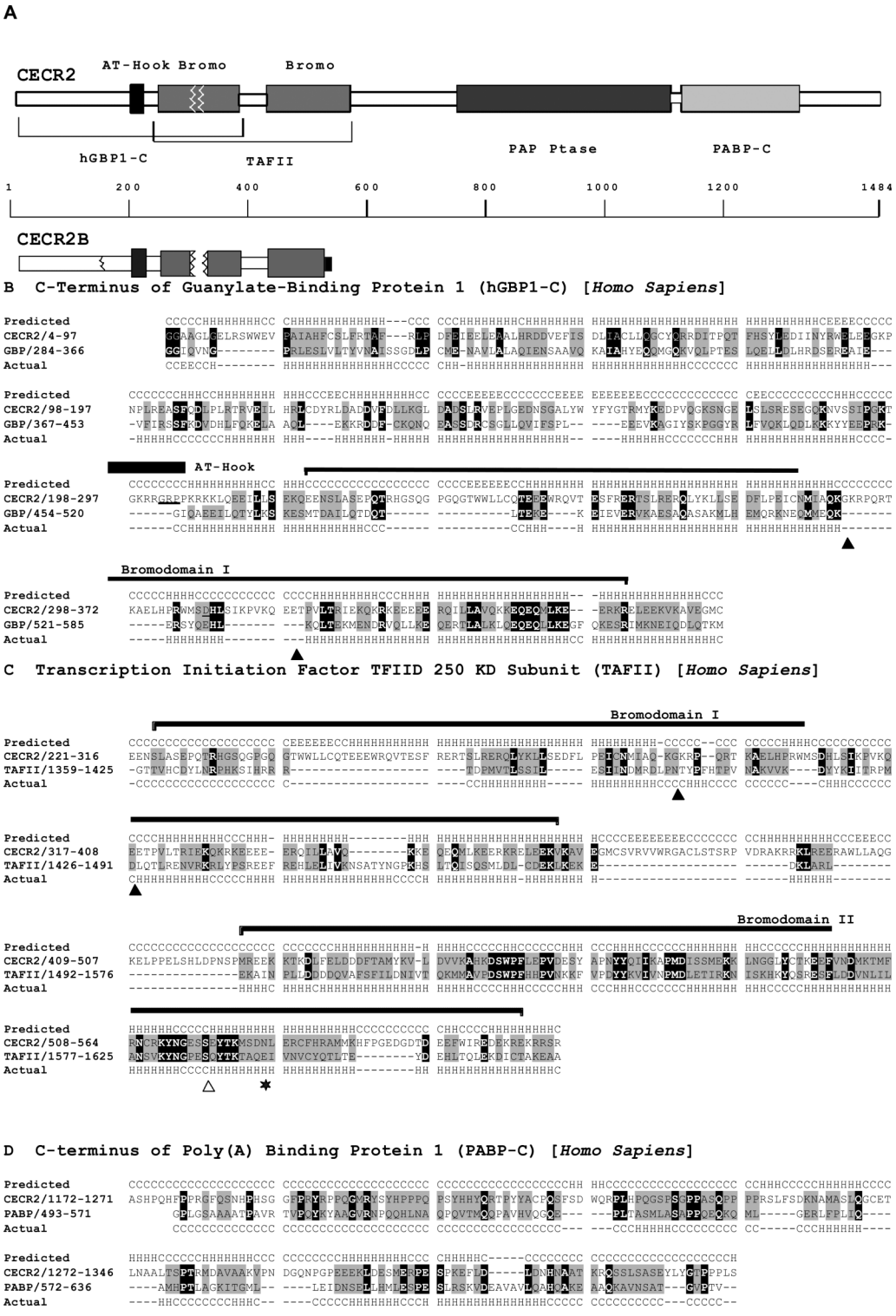

FIG. 9.

Homology domains of the CECR2 protein. (A) Schematic of the homology domain structure. The indicated homology domains are shown in proportion to the full-length 1484-residue CECR2 protein at top. The deletion in the bromodomain I of CECR2 due to exon skipping is indicated (white wavy lines). The black wavy line at left in CECR2B indicates

the point of N-terminal fusion with the library expression vector. (B) GBP1 and the AT-Hook Homology. Alignment to the C-terminus of 592-residue human GBP1 is based on the 3D-PSSM analysis with the predicted and actual structures indicated as described in Fig. 7. The AT-Hook sequence is overlined, its highly conserved signature GRP residues are underlined and bromodomain I is indicated. The deletion which results from exon skipping is between the solid triangles. (C) Example of homology to a two-bromodomain transcription factor. Both Pfam and 3D-PSSM searches detected the TFIID homologue because of its tandem bromodomains (overlined with bar). Alignment was performed with 3D-PSSM analysis. The point at which the alternate exon in CECR2 begins is indicated by the open triangle. and the C-terminus with a star. (D) Structural domain homology with PABP-C. Alignment to the C-terminus of 636-residue human poly A binding protein 1 was also based on the 3D-PSSM search. Residue identity and similarities predicted and actual secondary structure are indicated as described in Fig. 7.