ELSEVIER

Gene 289 (2002) 191–199

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Cloning, physical mapping and structural characterization of the human α_A-adaptin gene☆

Andreas Scorilas[a,b,c], Michael A. Levesque[a,b], Linda K. Ashworth[d], Eleftherios P. Diamandis[a,b,*]

[a]*Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ont. M5G 1X5, Canada*
[b]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ont. M5G 1X5, Canada*
[c]*National Center for Scientific Research 'Demokritos', IPC, Athens 153 10, Greece*
[d]*Human Genome Center, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA*

Received 30 November 2001; received in revised form 30 January 2002; accepted 19 February 2002
Received by R. Di Lauro

Abstract

Adaptins are major structural components of heterotetrameric protein complexes called adaptors, which are involved in intracellular receptor transport via clathrin-coated vesicles. In mice, one of these adaptins has been shown to be encoded by two genes, α_A-adaptin and α_C-adaptin, the former of which is expressed as two alternatively spliced transcripts. Using positional cloning gene approaches, we were able to identify the human α_A-adaptin gene, which consists of 24 exons spanning over 40 kb on chromosome 19q13.3 between the loci of the R-ras gene and the polynucleotide kinase phosphatase gene. The novel gene encodes a 977 amino acid, 107.6 kDa protein with 98% amino acid sequence identity to its murine ortholog. Human α_A-adaptin is expressed as a full-length transcript in forebrain, skeletal muscle, spinal cord, cerebellum, salivary gland, heart and colon. It is also ubiquitously expressed in tissues and in ZR-75-1 breast cancer cells and LNCaP prostate carcinoma cells as a smaller variant generated by splicing out of an exon encoding 22 amino acids in the hinge region of the protein.

Keywords: Adaptin; Adaptor; Human; Gene mapping; R-ras; Bcl2-like 12; Interferon regulatory factor 3; Gene sequence

1. Introduction

Intracellular membrane trafficking, involving budding, transport, and fusion of transport vesicles, provides the mechanism for the movement of proteins between different membrane compartments of a cell (Rothman and Weiland, 1996). The specificity in the system is conferred by a selective recognition of the cargo, its specific sequestration in appropriate transport vesicles at the budding site, and discriminative docking at the target membrane. Essential to the initiating events of vesicular trafficking are cytosolic proteins that assemble at the vesicle budding sites to form 'coats'. The best described coats are called adaptor protein complexes 1 and 2 (AP-1 and -2)[1], which mediate the recruitment of clathrin triskelions to the vesicle budding site, interact with sorting signals within the cytoplasmic domains of selected membrane proteins, and are associated with trafficking from the trans-Golgi network to the late endocytic pathway and from the plasma membrane to the early endosomes, respectively (Robinson, 1987). Two other adaptor complexes, AP-3 and -4, have also recently been described (Simpson et al., 1996; Dell'Angelica et al., 1997). Other non-clathrin-associated coats are formed by the 'coatamer' proteins COPI and COPII, which are responsible for vesicle formation in the endoplasmic reticulum or Golgi and the early secretory pathway, respectively (Schekman and Orci, 1996). All four adaptor complexes share a similar heterotetrameric structure composed of two heavy subunits (adaptins) and one medium (μ) and one small (σ) subunit (Kirchhausen, 1999): AP-1 consists of the subunits γ-adaptin, β1-adaptin,

Abbreviations: AP, adaptor protein; AP-1 and AP-2, AP complexes 1 and 2, respectively; NCBI, National Center for Biotechnology Information; BAC, bacterial artificial chromosome; JGI, Department of Energy's Joint Genome Institute; LLNL, Lawrence Livermore National Laboratory; IRF-3, interferon regulatory factor 3; PRMT1/HRMT1L2, protein arginine methyltransferase 1; PNKP, polynucleotide kinase phosphatase 1; EST, expressed sequence tag; RT-PCR, reverse transcription polymerase chain reaction; RACE, rapid amplification of cDNA ends; UPGMA, unweighted pair group method with arithmetic mean

☆ The nucleotide sequence for the human α_A-adaptin gene has submitted to the GenBank database under GenBank Accession Number AF289221.

* Corresponding author. Tel.: +1-416-586-8443; fax: +1-416-586-8628.
E-mail address: ediamandis@mtsinai.on.ca (E.P. Diamandis).

0378-1119/02/$ - see front matter
PII: S0378-1119(02)00504-8

μ1 and σ1; AP-2 contains α-adaptin, β2-adaptin, μ2, and σ2; AP-3 contains δ-adaptin, β3-adaptin, μ3, and σ3; and AP-4 contains ϵ-adaptin, β4-adaptin, μ4, and σ4. A considerable degree of sequence similarity exists between the various subunits and between the subunits in different species (Hirst and Robinson, 1998). All of the ~100 kDa adaptin proteins have a similar domain structure, consisting of a large, highly-conserved N-terminal trunk domain, a proline- and glycine-rich hinge region, and a more divergent C-terminal appendage (Hirst and Robinson, 1998). Structural studies have shown that the subunits assemble into a complex consisting of a bulky region that is formed by the N-terminal 'head' domains of the two adaptins and by the smaller subunits embedded between them and of 'ears' that symmetrically protrude from the body and correspond to the C termini of the adaptins (Kirchhausen, 1999). Each subunit in the adaptor complex is thought to have specific functions. Recognition of sorting signals, such as tyrosine-containing motifs, within cytoplasmic domains of transmembrane proteins has been shown to involve μ subunits (Bonifacino and Dell'Angelica, 1999). β-adaptins, through the clathrin-binding motifs within their hinge regions, interact with clathrin and promote its assembly and may also contribute to the recognition of sorting sequences (Bonifacino and Dell'Angelica, 1999). Through its ear domain, α-adaptin is able to interact with a number of accessory molecules, the functions of some of which are beginning to be elucidated. A binding partner for γ-adaptin, an eps15 homology domain-containing protein called γ-synergin, as well as a novel family of proteins called the GGAs which they interact with γ-synergin, have also been identified (Takatsu et al., 2000; Dell'Angelica et al. 2000).

Much of the current understanding of adaptor proteins has been derived from studies of α-adaptin, of which two isoforms have been cloned from a mouse brain DNA complementary to RNA (cDNA) library and had been previously designated as α_A-adaptin and α_C-adaptin based on their gel electrophoretic mobilities (Robinson, 1987). These isoforms are encoded by different genes and show 84% identity at the amino acid level (Robinson, 1987). α_A-Adaptin has been furthermore shown to be alternatively spliced, such that the expressed protein in most murine tissues, unlike the protein expressed in brain and skeletal muscle, is missing a 22 amino acid sequence in the hinge region encoded by a separate exon (Ball et al., 1995). In addition to the two murine α-adaptins, highly homologous cDNA sequences have also been cloned from other species including *Drosophila melanogaster* (Dornan et al., 1997), as well as from *Caenorhabditis elegans* and *Saccharomyces cerevisiae* (GenBank Accession Numbers U28742 and Z35798, respectively). Employing partially characterized genomic DNA sequences from bacterial artificial chromosome (BAC) and cosmid clones, a variety of bioinformatic and polymerase chain reaction (PCR) strategies for gene discovery, as well as cloning, sequencing, and tissue and cell line expression studies, we report here the precise mapping and detailed structural characterization of a gene on chromosome 19 that is the human ortholog of the murine α_A-adaptin gene. Further, based on these findings and the application of additional bioinformatic methods, we describe the sequence homology between human α_A-adaptin and other adaptins, we characterize the deducted amino acid sequence in terms of several functional parameters, and we reveal through tissue and cell line expression studies that the gene is alternatively spliced like its murine counterpart.

2. Materials and methods

2.1. Physical mapping and cloning

Toward identifying human α-adaptin genes, and because of the high levels of homology between adaptins, homology searches were performed using nucleotide sequences of known non-human α-adaptins as the queries for TBLASTN searches in the incomplete High Throughput Genomic Sequences database available through the National Center for Biotechnology Information. A potential α-adaptin amino terminal domain was identified in the BAC clone BC42053 sequenced by Joint Genome Institute (JGI). Genomic sequences from this clone were in the form of 87 contigs of different lengths. The clone, on chromosome 19, was obtained from the Lawrence Livermore National Laboratory and the genomic DNA was isolated. The chromosome 19 *Eco*RI restriction map (Ashworth et al., 1995), and long PCR strategies and sequencing using genomic DNA from this clone, were used to construct a contiguous sequence of the genomic area of interest. *Eco*RI sites of the sequences were determined using the WEBCUTTER2 software. Bioinformatic approaches were used as previously described (Smith and Xue, 1998; Yousef et al., 2000a–c, 2001; Scorilas et al., 2000a,b), by which a putative protein was identified having high homology with the mouse α_A-adaptin gene (Robinson, 1987). Using BLAST searching of the end sequences of the BC42053 BAC clone and based on *Eco*RI restriction digestion analysis, we identified a partially overlapping, more telomeric cosmid clone (R31181) sequenced by the JGI which allowed us to identify the 3′ extension of the new gene as well as the relative position of the new gene to the other previously identified genes R-Ras (Lowe et al., 1987), SR-A1 (Scorilas et al., 2001a), interferon regulatory factor 3 (IRF-3) (Lowther et al., 1999), Bcl2-like 12 (BCL2L12) (Scorilas et al., 2000b), protein arginine methyltransferase 1 (PRMT1/HRMT1L2) (Scorilas et al., 2000a), testis specific kinase substrate (TSKS) (Scorilas et al., 2001b), and polynucleotide kinase phosphatase (PNKP) (Karimi-Busheri et al., 1999) within the same chromosomal region (19q13.3). The sequence of the putative gene was verified by different approaches including expressed sequence tag (EST) database search, PCR screening of tissues, and DNA sequencing, as described below.

2.2. EST database searching

The predicted exons of the putative gene were subjected to homology searching using the BLASTN algorithm against the human EST database (dbEST). Clones with greater than 98% homology were obtained from the IMAGE consortium through Research Genetics Inc. (Huntsville, AL). The clones were propagated, purified and sequenced in both directions by the dideoxy method with an automated sequencer using insert-flanking vector primers.

2.3. Long PCR amplification

Because the cDNA for α_A-adaptin is predicted to be >3 kb, long PCR was employed. A pair of primers (forward: 5′-CGG TGT TCA TCT CCG ACA TC-3′ and reverse: 5′-CCA CGA AGT TCT CAG GGT TG-3′) bounding the first and the last exons of the α_A-adaptin gene, respectively, were used to amplify the predicted cDNA. PCR was performed using the Expand Long Template PCR System (Roche Molecular Systems, Indianapolis, IN) on template cDNA from human liver and brain – tissues from which the majority of ESTs had been isolated. We used a 50 μl reaction mixture containing 300 ng liver cDNA, 20 mM Tris–HCl (pH 7.5), 100 mM KCl, 2.25 mM $MgCl_2$, 1 mM dithiothreitol, 0.1 mM EDTA, 0.5% (v/v) Tween 20, 0.5% (v/v) Nonidet P40, 500 μM dNTPs, 300 nM primers and 2.6 units of enzyme mix containing thermostable Taq and Pwo DNA polymerases, on a Perkin-Elmer 9600 thermal cycler (Perkin-Elmer, Norwalk, CT). The cycling conditions were 94°C for 2 min, followed by ten cycles of 94°C for 10 s, 62°C for 30 s, 68°C for 8 min, then 20 cycles of 94°C for 10 s, 62°C for 30 s, 68°C for 8 min (increasing by 20 s per subsequent cycle), in addition to a final extension step at 68°C for 7 min. The PCR products were verified by electrophoresis on a 0.8% agarose gel. PCR products were then purified with exonuclease I plus shrimp alkaline phosphatase treatment as described elsewhere (Birren et al., 1998). Briefly, 1 μl of exonuclease I and 1 μl of shrimp alkaline phosphatase were added to the PCR products, then the mixture was incubated at 37°C for 15 min, then at 85°C for 15 min. Excess oligonucleotides and dNTPs were removed by a spin dialysis method using Centricon 30 microconcentrators (Amicon, Beverly, MA). Four sequential cycles of centrifugation and dilution were performed. After purification, PCR products were directly sequenced using an automated DNA sequencer as above.

2.4. Reverse transcription-PCR (RT-PCR) for tissue and cell line expression

Total RNA isolated from 28 different human tissues was purchased from Clontech, Palo Alto, CA. The breast cancer cell lines ZR-75-1 and T-47D as well as the prostate cancer cell line LNCaP were purchased from the American Type Culture Collection, Rockville, MD. Cells were cultured in RPMI media (Gibco BRL, Gaithersburg, MD) supplemented with glutamine (200 mmol/l), bovine insulin (10 mg/l), fetal bovine serum (10%), antibiotics and antimycotics, in plastic flasks, to near confluency. The cells were cultured for 24 h before messenger RNA (mRNA) extraction. Total mRNA was extracted from the tissues or cell lines using Trizol™ reagent (Gibco BRL) following the manufacturer's instructions. RNA concentrations were determined spectrophotometrically. A total of 2 μg RNA of each tissue was reverse-transcribed into first strand cDNA using the Superscript™ preamplification system (Gibco BRL). The final volume was 20 μl. Based on the information obtained from the genomic structure of the new gene, two gene-specific primers were designed (5′-GCC AAG CTG AAA CGC AAG AAG-3′ and 5′-AAC TGC ACC GAG GTC TTG TTG-3′) and PCR was carried out in a reaction mixture containing 1 μl of cDNA, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM dNTPs, 150 ng of primers and 2.5 units of HotStar™ DNA polymerase (Qiagen, Valencia, CA) on a Perkin-Elmer 9600 thermal cycler. The cycling conditions were 95°C for 15 min to activate the Taq DNA polymerase, followed by 35 cycles of 94°C for 30 s, 61°C for 30 s, 72°C for 1 min and a final extension step at 72°C for 10 min. Equal amounts of PCR products were electrophoresed on 2% agarose gels and visualized by ethidium bromide staining. The primers for RT-PCR spanned at least two exons to avoid contamination by genomic DNA.

2.5. 3′ and 5′ rapid amplification of cDNA ends (3′ and 5′ RACE)

In order to determine the sequences of the 3′ and 5′ ends of the α_A-adaptin gene, 3′ and 5′ RACE reactions were employed. According to the cDNA sequence and the structure of the human α_A-adaptin gene, reverse gene-specific primers were designed for both 3′ RACE (5′-GGC ACT CGA TAT TGA GCA CC-3′ and 5′-CCA CGA AGT TCT CAG GGT TG-3′) and 5′ RACE (5′-AGC CAT CCA AGG CTT TGT CT-3′ and 5′-AAT TTC CGC CTC TTT GCT CT-3′). Two rounds of RACE reactions for each pair of primers (nested PCR) were performed with 5 μl Marathon Ready™ cDNA from human thymus (Clontech) as a template. The reaction mixture and PCR conditions were selected according to the manufacturer's recommendations. In brief, the initial denaturation was for 5 min at 94°C, followed by 94°C for 5 s and 72°C for 2 min, for five cycles; then, 94°C for 5 s and 68°C for 2 min, for five cycles; then, 94°C for 5 s and 65°C for 2 min for 30 cycles for the first reaction and 25 cycles for the nested PCR reaction. Positive bands on electrophoresis were gel-purified using Qiagen Gel Purification kits (Qiagen, Chatsworth, CA) according to the manufacturer's instructions.

2.6. Cloning and sequencing of the PCR products

To verify the identity of the RT-PCR and long PCR

products, they were cloned into the pCR 2.1-TOPO vector (Invitrogen, Carlsbad, CA) as suggested by the manufacturer. The inserts were sequenced by automated DNA sequencing in both directions using vector-specific primers.

3. Results

3.1. Molecular cloning of the human α_A-Adaptin gene

Using the BC42053 BAC and R31181 cosmid clones that were obtained and processed as described above, we identified a putative adaptin gene sequence that was then blasted against the human EST database. Twenty one EST clones were found with >98% identity to one or more exons of this gene (Table 1); one of these clones was 99% identical to the last predicted exon and the 3′ untranslated region of the gene, and also contained a stretch of 40 adenine (A) nucleotides that was not found in the genomic sequence. 3′ RACE reactions and sequencing were also performed to verify the sequence of the 3′ end of the gene and the position of the poly(A) tail. A typical polyadenylation signal, AATAAA, was found and followed 14 nucleotides downstream by the poly(A) tail.

To identify the full mRNA structure of the gene and to determine the exon/intron boundaries, PCR reactions were performed using primers located in different software-predicted exons, using a panel of 28 human tissue cDNAs as templates. PCR products were sequenced. The two primers that were used for long PCR, as described in Section 2 were able to amplify the full coding region of the gene from different tissues. Comparison of the mRNA to the genomic structure indicated the presence of a gene consisting of 24 coding exons and 23 intervening introns. 5′ RACE reactions and sequencing were performed and 253 bp upstream were identified. An in-frame methionine start codon (GCCATCATGC) was found in the first exon that matches favorably with the consensus Kozak sequence (GCCA/GCCATGG). Moreover, there is a purine at position (−3) which occurs in 97% of vertebrate mRNAs (Kozak, 1987). It should also be noted that, like most other identified vertebrate adaptin genes, human α_A-adaptin does not have the consensus G nucleotide at position (+4). Translation of the mRNA sequence in all possible reading frames revealed the presence of only one frame that yields an uninterrupted polypeptide chain and that also has 98% homology with the mouse α_A-adaptin amino acid sequence.

Table 1
EST clones with >98% identity to exons of the human α_A-adaptin gene

GenBank Accession #	Tissue	IMAGE ID	Homologous exons
BE295338	Muscle	3531440	1–4
BE274905	Skin	3347201	1–5
BE273857	Kidney	3347224	1–6
AW249044	Lung	2820990	2–6
BE159471	Head and neck	–	3–8
BE397618	Lymph	3619918	4–9
BE397467	Lymph	3619406	4–9
AL042867	Testis	–	5–10
AW606713	Head and neck	–	6–7
BE312112	Brain	3161345	8–13
AI909619	Breast	–	5–9
BE397851	Lymph	43621036	5–12
AW675368	Placenta	2986179	11–15
AL045265	Testis	–	13–15
BE255864	Eye	3350448	17–24
AW905453	Nervous	–	22–24
N78384	Nervous	288873	22–24
AA469898	Testis	729961	22–24
AA232839	Fetal heart	666296	23–24
AA974267[a]	Testis	1586301	23–24
AA416916	Testis	729961	23–24

[a] Contains a poly(A) tail.

3.2. Structural characterization of the human α_A-adaptin gene and its protein product

The human α_A-adaptin gene consists of 24 coding exons and 23 intervening introns (Fig. 1), spanning an area of 39.9 kb of genomic sequence on chromosome 19q13.3. All of the exon/intron splice sites conform to the consensus sequence for eukaryotic splice sites (Iida, 1990), with the exception of the donor site of intron 19 which has the sequence AT. The lengths of the coding 24 exons are 278, 69, 143, 194, 130, 102, 109, 151, 169, 138, 183, 98, 232, 168, 161, 66, 57, 83, 92, 122, 117, 70, 135 and 432 bp, respectively. The predicted protein-coding region of the gene is 2934 bp in length and encodes a deduced 977 amino acid polypeptide with a calculated molecular weight of 107.6 kDa and an isoelectric point of 6.6.

Although the α_A-adaptin protein sequence is unique, comparative analysis revealed that it has a high degree of homology with other members of the adaptin multigene family. Human α_A-adaptin shows 98% amino acid sequence identity with mouse α_A-adaptin (Fig. 2) as well as lower levels of sequence identity with human and mouse α_C-adaptin (81%), *D. melanogaster* α-adaptin (66%) and δ-adaptin (23%), both human and mouse γ1-adaptin (26%), γ2-adaptin (25%), human δ-adaptin (21%), mouse δ-adaptin (19%), *Arabidopsis thaliana* γ1- and γ2-adaptin (26%) and human ϵ-adaptin (21%).

3.3. Splice variants of the human α_A-adaptin gene

PCR screening for the human α_A-adaptin transcripts using gene-specific primers revealed the presence of two bands in some of the tissue cDNAs examined. The two bands were gel purified, cloned and sequenced, by which the slower migrating band was shown to represent the full-length form (also called variant 1 in our GenBank submission; Accession Number AF289221) of the mRNA and the faster-migrating band (also called variant 2) was revealed to be a splice variant missing the entire 66 bp of exon 16 (Fig. 1). This splice variant, which has 98% identity with the

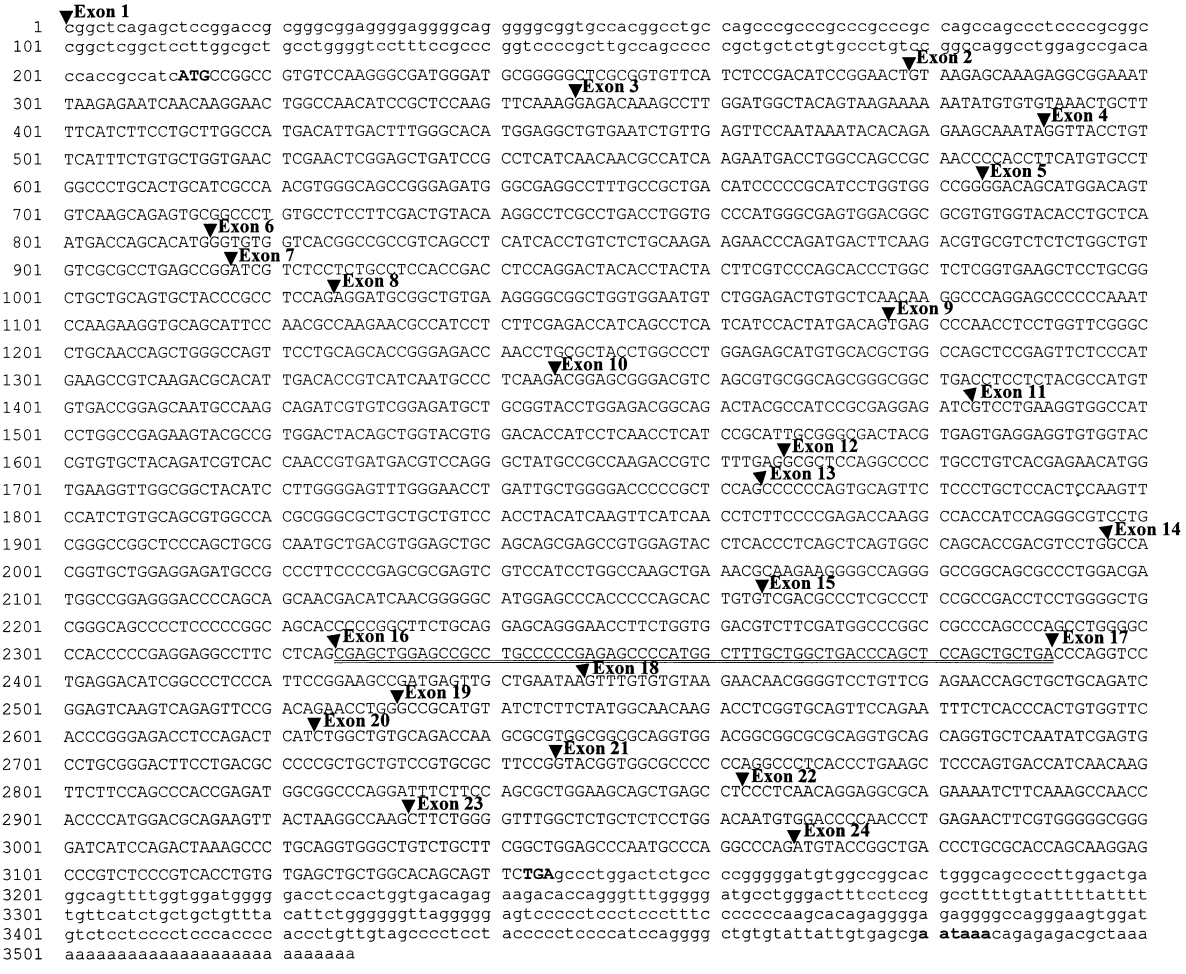

Fig. 1. Nucleotide sequence of the human α_A-adaptin cDNA based on the genomic structure. The first methionine (ATG), the stop codon (TGA), and the putative polyadenylation signal (AATAAA) are indicated by boldtype. Non-coding sequences are shown in lower case. Exon boundaries are shown with black arrowheads. Nucleotides marked by a solid double line below the sequence denote the alternatively spliced exon which is not present in the shorter variant of the α_A-adaptin mRNA. The entire gene, including intronic sequences, is reported in our GenBank submission AF289221.

smaller, alternatively spliced isoform of the mouse α_A-adaptin, is expected to encode a truncated protein product of 955 amino acids with a predicted molecular weight of 105.4 kDa and isoelectric point of 7.7.

3.4. Mapping of the of the human α_A-adaptin gene

Long PCR strategies and restriction analysis of a number of overlapping BAC clones spanning the chromosomal area of interest allowed us to locate a number of previously identified genes, together with our newly identified gene, along the *Eco*RI restriction map of the area (Ashworth et al., 1995). By identifying the position of R-Ras, SR-A1, IRF-3, BCL2L12, PRMT1/HRMT1L2, TSKS, PNKP and α_A-adaptin genes along these clones, we were able to precisely define the relative location and the direction of transcription of these eight genes (Fig. 3). R-ras is the most centromeric and its direction of transcription is from telomere to centromere. PNKP is the most telemetric and transcribe in the same direction. The human α_A-adaptin gene is localized 3.7kb downstream of TSKS and 54.1 kb upstream of PNKP gene. The direction of transcription of α_A-adaptin is from telomere to centromere, same as SR-A1, BCL2L12, PRMT1 and opposite of R-Ras, IRF-3, TSKS and PNKP genes (Fig. 3).

3.5. Expression patterns of the human α_A-adaptin gene in tissues and cell lines

RT-PCR analysis of different normal human tissues and cancer cell lines were used to establish the expression pattern of α_A-adaptin. Actin was used as a control gene. In each of the 26 adult and two fetal tissues tested, an α_A-adaptin transcript was detected (Fig. 4). The full-length form of the α_A-adaptin gene was found to be highly expressed in the fetal forebrain, as well as in the adult heart, forebrain, skeletal muscle, spinal cord and cerebellum. Lower levels of expression were also present in salivary gland and possibly colon. Expression of

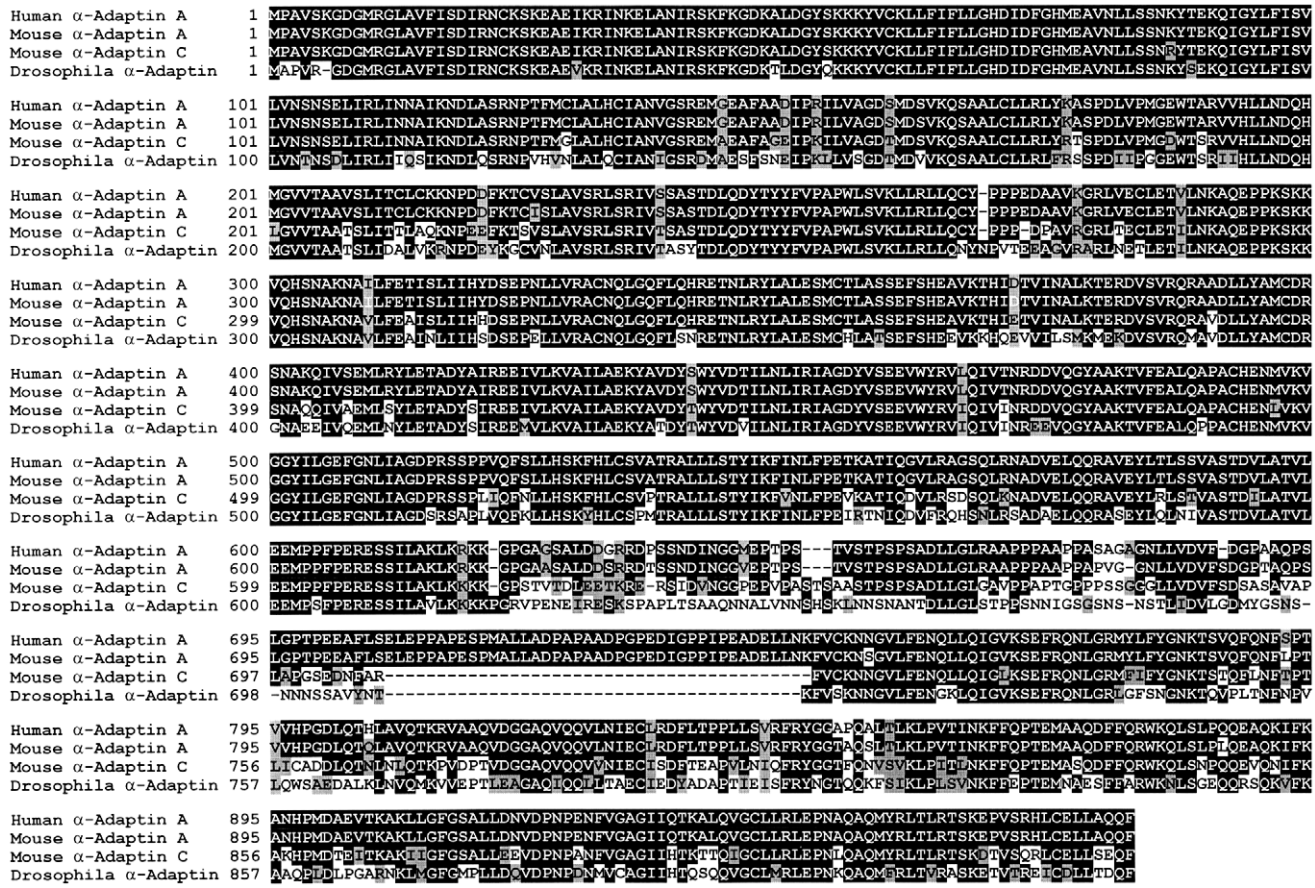

Fig. 2. Comparison of predicted amino acid sequences of human α_A-adaptin (longer variant), mouse α_A-adaptin and α_C-adaptin, and *Drosophila* α-adaptin. Identical amino acids are highlighted in black and similar residues in grey. Dashes represent gaps to bring the sequences into alignment.

this variant in other tested tissues was not found. The shorter splice variant was shown to be highly expressed in the bone marrow, skeletal muscle, spinal cord, adrenal gland, colon, lung, and stomach, and lower levels were evident in uterus, cerebellum and fetal forebrain.

Neither of the human α_A-adaptin isoforms appear to be expressed in the T-47D breast cancer cell line. Whereas a low level of expression of the shorter-length transcript was detected in the ZR-75-1 breast cancer cell line, a higher level was detectable in the LNCaP prostate carcinoma cell line (Fig. 4).

4. Discussion

In this study, we describe the cloning, precise chromosomal localization and detailed structural characterization of the human α_A-adaptin gene, identified primarily on the basis

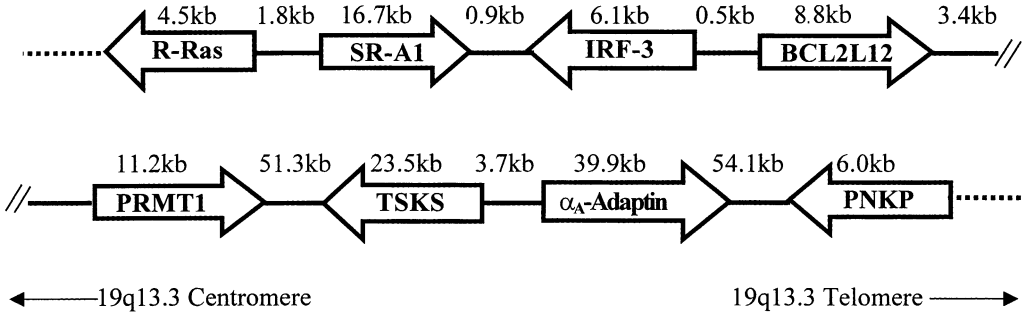

Fig. 3. Schematic representation of the position of the human α_A-adaptin gene relative to neighbouring genes on chromosome 19q13.3. Genes are represented by horizontal arrows denoting the direction of the coding sequence. Approximate lengths of genes, and distances between them, are indicated in kilobase pairs. Figure is not drawn to scale.

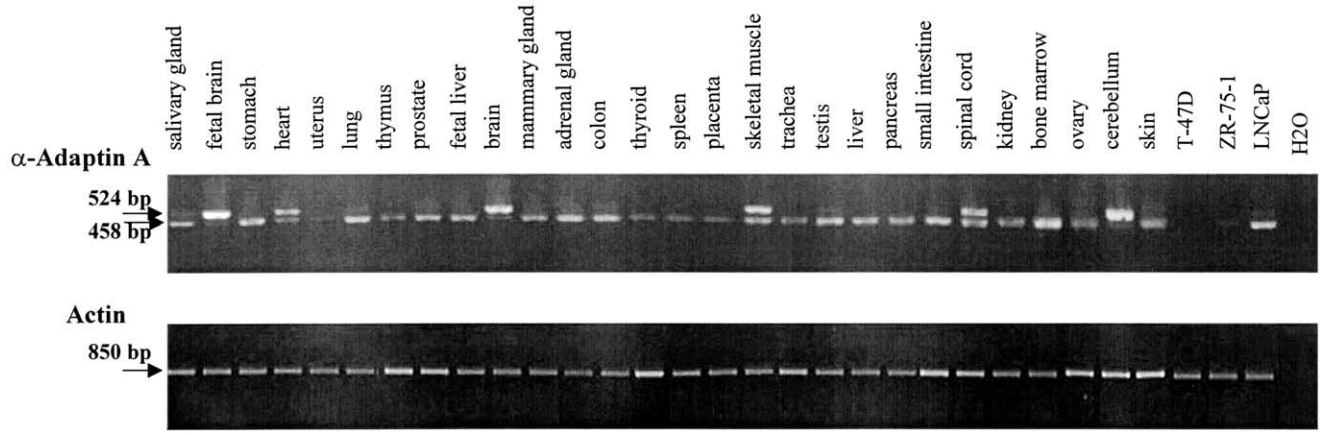

Fig. 4. Expression of the human α_A-adaptin gene in various tissues and three tumor cell lines, as determined by RT-PCR. Actin was similarly amplified as a control in each case.

of its sequence similarity to the murine α_A-adaptin gene (Robinson, 1987; Ball et al., 1995) with which it is shown to share 98% identity at the amino acid level as well as a comparable pattern by which its two splicing variants are expressed in various tissues and cell lines. A 'BLAST' search has shown 99% and 97% identity between the human α_A-adaptin variants and the hypothetical protein deposited by Wiemann et al. (2001) as well as by Wambutt et al. (Accession # CAB66859).Our findings that the 977 amino acid human α_A-adaptin protein differs from its well-studied murine ortholog by only 17 amino acids, and that these substitutions (and deletions, in two cases) occur predominantly in the flexible hinge region between amino acids ~600 and ~750 (Robinson, 1987), suggest that the two proteins may be functionally indistinct. The proteins also differ by four non-conservative amino acids in the C-terminal ear domain, the structure of which in the highly homologus murine α_C-adaptin protein has been shown to have a single binding site for a variety of ligands including amphyphysin, epsin, and eps15 that have as yet poorly understood functions (Owen et al., 1999). Studies of murine α-adaptins have also shown that, in addition to contributing along with β-adaptin to the structure of the earlike appendages of the AP-2 adaptor complex, α-adaptin contains a ~130 amino acid sequence in the N-terminus responsible for the specific interaction with its appropriate σ chains (Page and Robinson, 1995). The peptide sequence of human α_A-adaptin differs from the murine sequence of this region by only a single residue.

The human α_A-adaptin gene is expressed as either a full-length mRNA or as a variant encoding a protein in which a 22 amino acid segment in the hinge region, at the proximal end of a stretch of 41 amino acids which has no homologue in murine α_C-adaptin, has been omitted by splicing out the mRNA sequence representing the 16th exon of the 24 exon gene. This alternatively spliced exon is identical to the 66 bp sequence missing from the larger isoform of murine α_A-adaptin (Ball et al., 1995). Previous studies in mice had demonstrated, by Northern blotting, that this larger variant is expressed only in neuronal tissue and to a lesser extent in skeletal muscle, whereas both the smaller isoform and α_C-adaptin are universally expressed (Robinson, 1987). Using a large panel of human tissue mRNAs and RT-PCR, we extended these observations by showing the larger α_A-adaptin to be highly expressed not only in brain structures, spinal cord, and skeletal muscle, but also that lower abundances of the full-length α_A-adaptin message were present in cardiac muscle tissue, salivary gland, and colon tissues. Expression of α_A-adaptin in the latter two tissues had not been evaluated in earlier studies. In contrast to this restricted tissue distribution, a wide range of expression levels of the smaller isoform was found in all the other tissues tested. Interestingly, although only the smaller isoform was revealed to be expressed in two of the three cancer cell lines evaluated, expression of the smaller isoform was found to be increased in LNCaP prostate carcinoma cells after treatment with the steroids dihydrotestosterone and dexamethasone (data not shown). At least two *Drosophila* α-adaptin transcripts have also been shown to exist and exhibit patterns of tissue expression that are dependent on the developmental stage. In addition to α_A-adaptin, β2-adaptin has also been shown to be present in neural tissue as two isoforms, the smaller one of which lacks a 42 bp insert in the hinge region (Dornan et al., 1997). Based on observations that both α_A-adaptin isoforms as well as α_C-adaptin are apparently used in the same coated pits, it has been suggested that the three proteins may be functionally equivalent but expressed in response to different stimuli (Ball et al., 1995).

To our knowledge, the human α_A-adaptin gene is the first of the known adaptins in any mammalian species to be precisely mapped and characterized with respect to its genomic organization, although human β1-adaptin has been structurally characterized and localized to chromosome 22q12 and human γ1-adaptin has been assigned to chromosome 16q23. Human α_C-adaptin, β2-adaptin, γ2-adaptin, δ-

adaptin and ϵ-adaptin cDNAs have also been cloned. Comparisons of the genomic structure of the human α_A-adaptin gene to that of each of the human γ-, δ, and ϵ-adaptins, with which α-adaptin is thought to be most closely related in terms of phylogeny (Schledzewski et al., 1999), is not presently possible, but alignment of the predicted amino acid sequences encoded by their cDNAs revealed ~20–25% overall identity. In contrast, the β, μ, and σ chains exhibit greater sequence conservation (Kirchhausen, 1999). Given the degree of homology between the human and murine α_A-adaptin genes (98 and 92% at the protein and cDNA levels, respectively), it is tempting to speculate that the genomic structure of the latter would be very similar to the organization of the human α_A-adaptin gene described here. The chromosomal locus of the murine α_A-adaptin gene has not been reported, but may lie within a region on mouse chromosome 7 shown to be syntenic to human chromosome 19 (Lalley and McKusick, 1985) within which we have localized the human α_A-adaptin gene. Further support for this notion comes from our finding that the α_A-adaptin gene on human chromosome 19q13.3 is closed to R-ras and SR-A1 genes, which have been localized to murine chromosome 7 (Lowe et al., 1987; Scorilas et al., 2001a). Our analysis also showed that other proximal genes in this region include PRMT1/HRMT1L2) (Scorilas et al., 2000a), BCL2L12, TSKS (Scorilas et al., 2000b,2001b), IRF-3 (Lowther et al., 1999), centromerically and PNKP telomerically (Karimi-Busheri et al., 1999) to α_A-adaptin.

Acknowledgements

A portion of this work was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48.

References

Ashworth, L.K., Batzer, M.A., Brandriff, B., Branscomb, E., de Jong, P., Garcia, E., Garnes, J.A., Gordon, L.A., Lamerdin, J.E., Lennon, G., Mohrenweiser, H., Olsen, A.S., Slezak, T., Carrano, A.V., 1995. An integrated metric physical map of human chromosome 19. Nat. Genet. 11, 422–427.

Ball, C.L., Hunt, S.P., Robinson, M.S., 1995. Expression and localization of α-adaptin isoforms. J. Cell Sci. 108, 2865–2875.

Birren, B., Green, E.D., Klapholz, S., Myers, R.M., Roskams, J., 1998. Genome Analysis: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Plainview, NY.

Bonifacino, J.S., Dell'Angelica, E.C., 1999. Molecular bases for the recognition of tyrosine-based sorting signals. J. Cell Biol. 145, 923–926.

Dell'Angelica, E.C., Ohno, H., Ooi, C.E., Rabinovich, E., Roche, K.W., Bonifacino, J.S., 1997. An adaptor-like protein complex with ubiquitous expression. EMBO J 16, 917–928.

Dell'Angelica, E.C., Puertollano, R., Mullins, C., Aguilar, R.C., Vargas, J.D., Hartnell, L.M., Bonifacino, J.S., 2000. GGAs: a family of ADP ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex. J. Cell Biol. 149, 81–94.

Dornan, S., Jackson, A.P., Gay, N.J., 1997. α-adaptin, a marker for endocytosis, is expressed in complex patterns during *Drosophila* development. Mol. Biol. Cell 8, 1391–1403.

Hirst, J., Robinson, M.S., 1998. Clathrin and adaptors. Biochim. Biophys. Acta 1401, 173–193.

Iida, Y., 1990. Quantification analysis of 5′-splice signal sequences in mRNA precursors: mutations in 5′-splice signal sequence of human beta-globin gene and beta-thalassemia. Theor. Biol. 145, 523–533.

Karimi-Busheri, F., Daly, G., Robins, P., Canas, B., Pappin, D.J.C., Sgouros, J., Miller, G.G., Fakhrai, H., Davis, E.M., Le Beau, M.M., Weinfeld, M., 1999. Molecular characterization of a human DNA kinase. J. Biol. Chem. 274, 24187–24194.

Kirchhausen, T., 1999. Adaptors for clathrin-mediated traffic. Annu. Rev. Cell Dev. Biol. 15, 705–732.

Kozak, M., 1987. An analysis of 5′-non-coding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15, 8125–8148.

Lalley, P.A., McKusick, V.A., 1985. Report of the committee on comparative mapping. Cytogenet. Cell. Genet. 4D, 536–566.

Lowe, D.G., Capon, D.J., Delwart, E., Sakaguchi, A.Y., Naylor, S.L., Goeddel, D.V., 1987. Structure of the human and murine R-ras genes, novel genes closely related to ras proto-oncogenes. Cell 48, 137–146.

Lowther, W.J., Moore, P.A., Carter, K.C., Pitha, P., 1999. Cloning and functional analysis of the human IRF-3 promoter. DNA Cell Biol. 18, 685–692.

Owen, D.J., Vallis, Y., Noble, M.E., Hunter, J.B., Dafforn, T.R., Evans, P.R., McMahon, H.T., 1999. A structural explanation for the binding of multiple ligands by the alpha-adaptin appendage domain. Cell 97, 805–815.

Page, L.J., Robinson, M.S., 1995. Targeting signals and subunit interactions in coated vesicle adaptor complexes. J. Cell Biol. 131, 619–630.

Robinson, M.S., 1987. 100-kD coated vesicle proteins: molecular heterogeneity and intracellular distribution studies with monoclonal antibodies. J. Cell Biol. 104, 887–895.

Rothman, J.E., Weiland, F.T., 1996. Protein sorting by transport vesicles. Science 272, 227–234.

Schekman, R., Orci, L., 1996. Coat proteins and vesicle budding. Science 271, 1526–1533.

Schledzewski, K., Brinkmann, H., Mendel, R.R., 1999. Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2 and 3 and the F subcomplex of the coatamer COP1. J. Mol. Evol. 48, 770–778.

Scorilas, A., Black, M.H., Diamandis, E.P., 2000a. Genomic organization of the human arginine methyltransferase 1 (PRMT1) gene: differential expression of splice variants in breast cancer. Biochem. Biophys. Res. Commun. 278, 349–359.

Scorilas, A., Kyriakopoulou, L., Yousef, G., Ashworth, L., Kwamie, A., Diamandis, E.P., 2000b. Molecular cloning, physical mapping and expression analysis of a novel gene, BCL2L12, encoding for a proline-rich protein with a highly conserved BH2 domain of the Bcl-2 family. Genomics 72, 217–221.

Scorilas, A., Kyriakopoulou, L., Katsaros, D., Diamandis, E.P., 2001a. Cloning of a gene (SR-A1), encoding for a new member of the human Ser/Arg-rich family of Pre-mRNA splicing factors: overexpression in aggressive ovarian cancer. Br. J. Cancer 85, 190–198.

Scorilas, A., Yousef, G., Diamandis, E.P., 2001b. Identification and characterization of a novel human testis specific kinase substrate gene which is downregulated in testicular carcinomas. Biochem. Biophys. Res. Commun. 285, 400–408.

Simpson, F., Bright, N.K., West, M.A., Newman, L.S., Darnell, R.B., Robinson, M.S., 1996. A novel adaptor-related protein complex. J. Cell. Biol. 133, 749–760.

Smith, D.K., Xue, H., 1998. A major component approach to presenting consensus sequences. Bioinformatics 14, 151–156.

Takatsu, H., Yoshino, K., Nakayama, K., 2000. Adaptor γ ear homology domain conserved in γ-adaptin and GGA proteins that interact with γ-synergin. Biochem. Biophys. Res. Commun. 271, 719–725.

Wiemann, S., et al., 2001. Toward a catalog of human genes and proteins:

sequencing and analysis of 500 novel complete protein coding human cDNAs. Genome Res. 11, 422–435.

Yousef, G., Chang, A., Scorilas, A., Diamandis, E.P., 2000a. Genomic organization of the human kallikrein locus on chromosome 19q13.3–q13.4. Biochem. Biophys. Res. Commun. 276, 125–133.

Yousef, G.M., Scorilas, A., Diamandis, E.P., 2000b. Genomic organization, mapping, tissue expression and hormonal regulation of trypsin-like serine protease (TLSP PRSS20), a new member of the human kallikrein gene family. Genomics 63, 88–96.

Yousef, G., Scorilas, A., Magklara, A., Soosaipillai, A., Diamandis, E.P., 2000c. The PRSS6 gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family-Genomic characterization, mapping, tissue expression and hormonal regulation. Gene 254, 119–128.

Yousef, G., Scorilas, A., Jung, K., Aschworth, L., Diamandis, E.P., 2001. The new human kallikrein gene KLK15 is up-regulated in prostate cancer. J. Biol. Chem. 276, 53–61.