The human *TruB* family of pseudouridine synthase genes, including the *Dyskeratosis Congenita 1* gene and the novel member *TRUB1*

CINZIA ZUCCHINI*, PIERLUIGI STRIPPOLI*, ALESSIA BIOLCHI, ROSSELLA SOLMI, LUCA LENZI, PIETRO D'ADDABBO, PAOLO CARINCI and LUISA VALVASSORI

Centro di Ricerca in Genetica Molecolare 'Fondazione CARISBO' at the Istituto di Istologia ed Embriologia Generale, Via Belmeloro 8, University of Bologna, 40126 Bologna, Italy

Received August 1, 2002; Accepted September 30, 2002

Abstract. A novel human gene denominated *TruB pseudouridine (ψ) synthase homolog 1* (*E. coli*) (approved symbol, *TRUB1*) has been identified and characterized. Spanning ~40 kb on chromosome 10 and including 8 exons, *TRUB1* is the first described human ortholog of bacterial *TruB*/*ψ55*, a gene involved in tRNA pseudouridinilation. *TRUB1* gene encodes a 349-amino acid product, with a VFAVHKPKGPTSA box in positions 71-83 corresponding to motif I of the TruB family (probably involved in conserving protein structure). The TruB domain of TRUB1 lies between W104 and I255, and contains another short motif, GGTLDS AARGVLVV, including the highly conserved D residue that characterizes motif II (involved in uridine recognition and in catalytic function of ψ synthases). Northern blot analysis revealed that *TRUB1* mRNA is widely expressed in various human tissues (especially heart, skeletal muscle and liver). Phylogenetic analysis of the *TruB* domain revealed another human gene (approved symbol *TRUB2*) encoding a conserved *TruB* domain, located on human chromosome 9. Thus, the human *TruB* family includes at least three members: i.e. *DKC1* (previously identified), *TRUB1* and *TRUB2*. The TRUB1 and TRUB2 products could be the hitherto unidentified human tRNA ψ synthases. Although TRUB1 is not highly similar to DKC1/dyskerin (whose mutations cause X-linked *dyskeratosis congenita*) and putatively affects tRNA rather than rRNA modification, it is the most similar human protein to dyskerin. Study of *TRUB1* (and *TRUB2*) should facilitate understanding of the molecular mechanisms of RNA modification and the involvement of ψ synthases in human pathology, including dyskeratosis-like diseases.

Correspondence to: Dr Paolo Carinci, Istituto di Istologia ed Embriologia Generale, Via Belmeloro 8, 40126 Bologna, Italy
E-mail: carinci@alma.unibo.it

*Contributed equally

Key words: RNA modification, sequence analysis, gene family, human genome

Introduction

Pseudouridine (ψ) is the isomer of uridine, having a carbon-carbon bond between uracil C5 and ribose C1 (instead of the N1-C1 glycosyl bond in uridine). It is an abundant constituent of all kinds of RNA, apart from mRNAs (1). The isomerization of uridine in to ψ is a post-trancriptional modification carried out by a family of enzymes called ψ synthases.

Four genes encoding different ψ synthases were originally cloned from *E. coli* (2-5). The many ψ synthases that have since been cloned from various other organisms are grouped into four families on the basis of their amino acidic sequence alignments. Each family is named after one of the original *E. coli* ψ synthases: i.e. TruA, TruB, RluA and RsuA (6). The members of the four ψ synthase families do not globally present significant sequence similarity. Nevertheless, they do share short sequence motifs (6,7). In particular, motif I, which occurs in all the families except TruA, shows two highly conserved residues (one proline and one lisine) and probably has a structural role (8). Motif II, which is common to all four families, contains a highly conserved aspartic acid with an essential catalytic function (9-11). Motif III occurs only in the RluA and RsuA families (6).

Identification of ψ synthase orthologs in different species has been hampered by the overall lack of significance in sequence similarity. In the TruB family, the orthologs of *E. coli* TruB that have so far been cloned in various eukaryotes include *Pus4* and *Cbf5* in yeast (12-15), *Nop60B/minifly* in *D. melanogaster* (16,17), *NAP57* in rat (18) and, in humans, *DKC1* (*Dyskeratosis congenita 1*) (19). *DKC1* has recently been identified as the causative gene of the X-linked form of *dyskeratosis congenita* (DKC) (19-21), a rare severe hereditary disorder affecting the skin, mucous membranes and bone marrow (22,23). DKC symptoms include skin pigmentation, nail dystrophy, mucosal leukoplakia; various non-cutaneous abnormalities have also been described (23). The DKC1 protein, commonly referred to as dyskerin, is responsible for the site-specific pseudouridination of rRNA, and is also a component of the telomerase complex (24).

Herein, we report the cDNA sequence, genomic organization and mRNA expression of a novel human gene encoding a putative ψ synthase. While conducting a bio-

informatic search for novel keratinocyte-specific genes, we identified an mRNA fragment listed only in keratinocyte-derived libraries. Subsequent sequence analysis demonstrated that the novel human mRNA encodes for a protein similar to a ψ synthase. This protein is a novel human TruB member that is directly related to the bacterial TruB family. In agreement with Human Genome Organization (HUGO) Nomenclature Committee, its gene has been denominated '*TruB pseudouridine (ψ) synthase homolog 1* (*E. coli*)' (approved symbol *TRUB1*). Even though *TRUB1* bears more similarity to human *DKC1* than any other human gene so far described, the lack of overall statistical significance between the complete sequences explains why it previously escaped recognition at straightforward homology searches. Sequence analysis allowed identification of a third member of the *TruB* human gene family, officially denominated '*TruB pseudouridine (ψ) synthase homolog 2* (*E. coli*)' (approved symbol *TRUB2*). We discuss the phylogenesis and putative functional/pathological implications of the *TruB* family members from bacteria to humans.

Materials and methods

Expressed sequence tags (ESTs) and genomic database searches. The nr (non redundant) nucleotidic sequence database including Genbank, dbEST and High Throughput Genome Sequences (HTGS) databases at the National Center for Biotechnology Information World Wide Web server were searched by TBLASTN 2 (25) using as query the nucleotidic sequence of EST D29276 (default parameters). This sequence was recorded as: 'Found only in library 101: cell line: keratinocyte; cDNA sources: skin-epidermis' in Unigene [Library differential display (http://www.ncbi.nlm.nih.gov/UniGene/info/ddd.html)]. Clusters of ESTs were assembled using ESTBlast software (a tool for contig building with Expressed Sequence Tags created at Glaxo Wellcome by the Bioinformatics Department) at the Human Genome Mapping Project (HGMP) server (http://www.hgmp.mrc.ac.uk).

Reverse transcription-polymerase chain reaction (RT-PCR) amplification. To obtain the amino acidic sequence data necessary for sequence comparison, we sequenced RT-PCR products covering the mRNA open reading frame of human *TRUB1*. Total human RNA was obtained by the method of Chomczynski and Sacchi (26) from cultured keratinocytes derived from neonatal human foreskin (NHEK cells, Clonetics, Walkersville, MD, USA). For RT-PCR, 1 μg of total RNA was reverse transcribed at 42˚C for 60 min in a 25 μl final volume by cloned Moloney murine leukemia virus reverse-transcriptase 200 U (Promega, Madison, WI; used with companion buffer), 5 μM oligo dT-15 and 500 μM for each dNTP. The primers for amplification were designed using the software Amplify (http://www.wisc.edu/genetics/CATG/amplify/). The Genbank Accession Number of the sources for primer design was AL355340 (*Homo sapiens* chromosome 10 clone RP11-383C6). For human *TRUB1*, primers were as follows (progressive numbering and in 5'→3' direction): #1-GTGCACCTCCACGATGAAACAG (exon 1, 5' untraslated region - 5'-UTR, forward) used with #2-AGTCCCTCCATG CCCAATTTTC (exon 2, reverse) giving a product of 396 bp; #3-GAATGCCTTCTCCAGAATGGAC (exon 2, forward) used with #4-TCCAATGTCACTGACCAAGCTTC (exon 7, reverse) (497 bp product); #5-GATGAAGAGAGGTGA AGTCGTAG (exon 6, forward) used with #6-GTCATTCTG CATCTGCACACAG (exon 8, 3'-UTR region, reverse) (478 bp product). PCR experiments were performed in 50 μl final volume, containing 1-5 μl RT mix, 1.25 U Taq Polymerase (TaKaRa, Shiga, Japan) with companion reagents (0.2 mM each dNTP, 1.5 mM $MgCl_2$, 1X PCR buffer), and 0.3 μM each primer. An initial denaturation step of 2 min at 94˚C was followed by amplification for 40 cycles, (30 sec at 94˚C, 30 sec at 61˚C, 30 sec at 72˚C) and final extension for 7 min at 72˚C.

cDNA sequencing. All RT-PCR products obtained as above were gel analysed following standard methods, purified using a Nucleospin extraction kit (Clontech, Palo Alto, CA, USA) and then subjected to automated sequence analysis of both DNA strands for each fragment, with the same primers used in the respective PCR reactions. The BigDye chain-terminator method was used with an automated ABI 377 DNA sequencer (Perkin-Elmer, Foster City, CA). Each region was sequenced twice using two independent amplification reaction products as template.

Northern blot analysis. In order to study *TRUB1* mRNA expression, a nylon membrane containing 2 μg poly(A)$^+$ RNA samples from human tissues (multiple tissue Northern blot - MTN human 12-lane, Clontech) was hybridized with a *TRUB1* probe. The *TRUB1* probe was a 396 bp RT-PCR product containing 36% of the coding region, obtained as described above, between primers #1 and #2, purified by Nucleospin Extraction kit (Clontech) and ^{32}P-labeled by random priming using MegaPrime kit (Amersham Pharmacia Biotech, Uppsala, Sweden). The filter was hybridized and washed at high stringency, essentially as described by Church and Gilbert (27), except that albumin was omitted, and exposed to X-ray film for 2 to 6 days.

Sequence analysis. The amino acidic sequences were aligned by ClustalW software (version 1.7) (28). Profile, motif, and pattern searches were conducted by updated tools and databases. A SMART analysis was run at the P. Bork group server (http://coot.embl-heidelberg.de/SMART). Pfam (Version 4.1, July 1999, 1488 families) (29) collection of protein families and domains was searched at the server of The Sanger Centre (http://www.sanger.ac.uk/Pfam/search.shtml) (Hinxton, Cambridge, UK). PSI-BLAST (Position-specific iterated BLAST search) bioinformatics analysis of pattern (25) was run at the NCBI World Wide Web server.

Molecular evolution analysis. The coding sequences were subjected to neighbour-joining analysis at the server of EBI (European Bioinformatics Institute, Hinxton, Cambridge, UK). The default parameters offered by the server were used (http://www.ebi.ac.uk/clustalw/; blosum matrix; Kimura correction; penalty for gap: open 10, extension 0.05, distance 8; bootstrap value 1,000). TreeView PC was used as tree visualization software (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

Results

ESTs and genomic database searches. The ESTBlast process, starting from EST #D29276, allowed us to identify a partial coding sequence with a putative 3'-UTR. The encoded sequence was highly similar to the hypothetical murine protein BAB27569. Using the murine sequence data as query, a significant match was found in the unfinished human genomic clone AL355340. Several ESTs matching this genomic clone were then retrieved and assembled, providing the putative mRNA and protein sequence of the *TRUB1* gene. The actual cDNA sequence of the new *TRUB1* gene (see below) was confirmed by many EST clones related to our construct. We assigned 37 human ESTs to *TRUB1*. We found that corresponding clusters are listed in the Unigene database as Hs.321052 and Hs.88678.

The EST distribution showed the expression of *TRUB1* in several neoplastic human tissues: brain (AI422325, AI418848, BE858241), ovary (BF059067, BF116232, BF058155), pancreas (BE736287), small intestine (BF982548), prostata (BG180560), uterus (BE885960) cancers. *TRUB1*-related EST can also be found in very early development stages in human foetal heart (AI274244), as well as in total foetus (AA393936). They are then found in the testis (AI208778, BF979739, BG719608), colon (AW014385), germinal centre B cells (AA282195, AA769039, AA737549, AA279917, AA768828, AA279406), melanocytes (N35167, AI274244), prostate (BE645624, BE856143, AI401692, AI745545), kidney (AI685024, AI640751), pregnant uterus (AI274244) and breast (BF746364).

RT-PCR amplification and cDNA sequencing. RT-PCR products of the expected size, as determined by bioinformatics analysis, were obtained from cultured keratinocytes derived from neonatal human foreskin (NHEK cells, Clonetics). The complete coding sequence of *TRUB1* was determined by assembling the respective sequences from RT-PCR products. The sequence, which has been deposited in the GenBank database with the accession number AF448144 (*TRUB1*), is exactly included between the 3' ends of primers #1 and #6.

TRUB1 cDNA sequence has a continuous open reading frame (ORF) of 1,047 nucleotides (349 amino acids). The first AUG shows a good Kozak consensus sequence for the translation initiation (4 positions, with both A/G in -3 and G in +4 with respect to the consensus GCCGCCRCCAUGG, where R = purine and AUG = translation initiation codon) (30). Molecular weight estimated from the predicted 349 amino acid protein is 37.3 kDa (theoretical pI: 8.44). Alignment of TRUB1 with orthologs from the different biological groups is shown in Fig. 1.

TRUB1 intron/exon boundaries were determined by BLAST comparison of our cDNA sequence with the sequence of an unfinished clone (AL355340 version 15, GI:13897058) containing a sequence that is complementary to our cDNA. All sequence data regarding AL355340 were produced by the Human Chromosome 10 Group at Sanger Centre, Hinxton, UK. No differences are observable between our cDNA sequence (AF448144) and the matching genomic sequence. All introns conform to the GT/AG rules; their limit (with respect to the above-cited genomic sequence) being: exon 1 from 137124 to 136800 (325 bp) and intron 1 from 136799 to 132556 (4,244 bp); exon 2 from 132555 to 132457 (99 bp) and intron 2 from 132456 to 124107 (8,350 bp); exon 3 from 124106 to 124051 (56 bp) and intron 3 from 124050 to 115438 (8,613 bp); exon 4 from 115437 to 115356 (82 bp) and intron 4 from 115355 to 104796 (10,560 bp); exon 5 from 104795 to 104723 (73 bp) and intron 5 from 104722 to 103029 (1,694 bp); exon 6 from 103028 to 102889 (140 bp) and intron 6 from 102888 to 100835 (2,054 bp); exon 7 from 100834 to 100778 (57 bp) and intron 7 from 100777 to 100041 (737 bp); exon 8 from 100040 to 97487 (2,554 bp, as determined by matching poly(A) sequence-containing EST D29276).

The 3'-untraslated region (3'-UTR) is 2,607 nucleotides long from the stop codon to the first A of the polyadenilate tail, as determined by comparison of two EST sequences containing a poly(A) stretch (AW504496, AI299932). A polyadenilation recognition signal fulfilling the AATAAA consensus sequence lies 20 nucleotides upstream at the start of the poly(A) tail (from 97507 to 97512 in the genomic sequence).

Northern blot analysis. Hybridization bands for *TRUB1* mRNA were visible in all RNA lanes corresponding to twelve human tissues (Fig. 2). Three bands were consistently observed even after high-stringency washes; comparison with provided marker points using a semi-logarithmic graph showed that their sizes were 5.2, 3.8 kb and 1.75 kb. The highest expression was observed in heart, skeletal muscle and liver tissues. Lower levels of transcripts were detected in the lung, small intestine, kidney and spleen; in the brain, colon, thymus, placenta and peripheral blood leukocytes, the expression was barely detectable. A uniform expression pattern for the three transcripts was seen in many of the tissues (brain, skeletal muscle, colon, thymus, spleen, kidney, liver, placenta and peripheral blood). In the remaining tissues, particular bands gave a significantly stronger signal (the 5.2 and 1.75 kb mRNA in the heart and small intestine, the 1.75 kb mRNA in the lung).

Motif searches. PSI-BLAST analysis of the TRUB1 amino acid sequence was conducted for nine iterations, identifying many proteins with known tRNA ψ synthase function. The protein most similar to human TRUB1 were mainly bacterial tRNA pseudouridine 55 synthases (Expect ≤4e-66). In the same analysis, a novel hypothetical protein (NP_056494) was identified. This protein is predicted to start from the human cDNA sequence AK001956 recorded in the context of a human cDNA sequencing project at Helix Research Institute, Chiba, Japan. BLAST analysis shows that the corresponding gene (approved symbol *TRUB2*) is located on human chromosome 9. Intron/exon boundaries were determined by BLAST comparison of its cDNA sequence with the sequence of the finished clone AL359091. All sequences and numbering data referred to AL359091 were produced by the chromosome 9 Mapping Group at Sanger Centre, Hinxton, UK. All introns conform to the GT/AG rule, and their limits being (with respect to the above-cited genomic sequence): exon 1 from 21614 to 21496 (119 bp) and intron 1 from 21495 to 20927 (569 bp); exon 2 from 20926 to 20795 (132 bp) and intron 2 from 20794 to 16423 (4,372 bp); exon 3 from 16422 to 16348

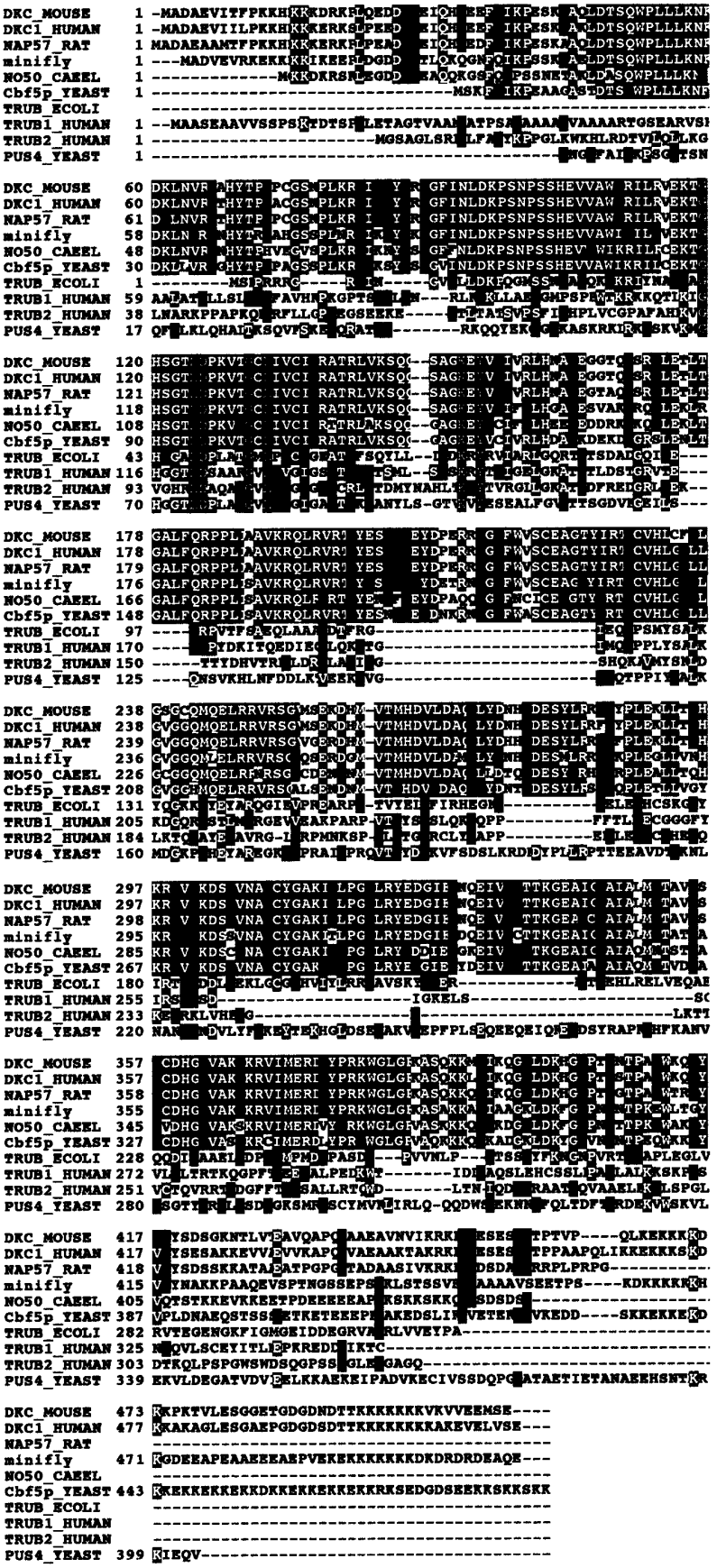

Figure 1. The TruB protein family sequence ClustalW alignment. Yellow letters, identical and light blue background, conserved or similar (pink background) amino acids in at least 50% of the sequence (MacBoxShade, default parameters). There is high conservation from bacteria to humans of certain specific amino acids: namely (with respect to TRUB1 positions), G115, L120-D121, G126, L128, K147 and Y149. Comparison with published data shows that the LD conserved residues belong to the ψ synthase motif II. A broader conserved region, which includes motif II near its amino terminus, matches the described TruB domain from W104 to I255. CAEEL, *C. elegans*; ECOLI, *E. coli*; HUMAN, *H. sapiens*; YEAST, *S. cerevisiae*; minifly, *D. melanogaster* Nop60B; MOUSE, *M. musculus*; RAT, *R. norvegicus*.

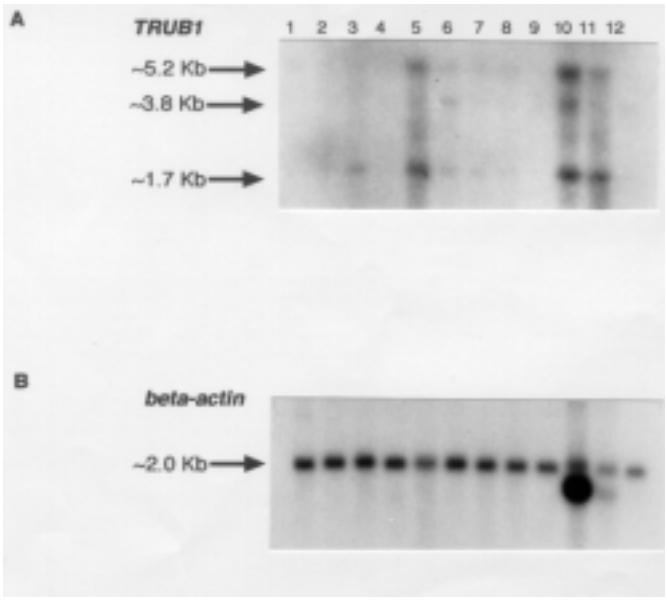

Figure 2. Northern blot analysis of poly(A)$^+$ RNA from 12 human tissues (lane number in brackets): peripheral blood leukocytes (1), lung (2), placenta (3), small intestine (4), liver (5), kidney (6), spleen (7), thymus (8), colon (9), skeletal muscle (10), heart (11), and brain (12). A, hybridization with a cDNA probe for human *TRUB1* mRNA. B, murine ß-actin probe hybridization, subsequentially performed as a control for the amounts of RNA loaded in each lane; a lower 1.8 kb α- or γ-actin mRNA is known to be expressed in some tissues (e.g. skeletal muscle).

(75 bp) and intron 3 from 16347 to 14825 (1,523 bp); exon 4 from 14824 to 14763 (62 bp) and intron 4 from 14762 to 13086 (1,677 bp); exon 5 from 13085 to 13004 (82 bp) and intron 5 from 13003 to 10798 (2,206 bp); exon 6 from 10797 to 10725 (73 bp) and intron 6 from 10724 to 10220 (505 bp); exon 7 from 10219 to 10083 (137 bp) and intron 7 from 10082 to 9072 (1,011 bp); exon 8 from 9071 to 8313 (759 bp, as determined by matching poly(A) sequence-containing EST #AI589701).

Table I. Domain hits of TruB family protein sequences.

Protein	Domain	Stretch	Score
TruB_ECOLI	-	30-180	1.40e-94
DKC1	TruB	107-247	6.6e-69
	PUA	295-370	1.8e-25
TRUB1	TruB	133-255	5.8e-45
TRUB2	TruB	86-233	5.5e-05

Stretch is the matched interval (amino acid positions); scores are given as E-value for Pfam. TruB, the family pseudouridine synthase N terminal domain; PUA, an RNA binding domain.

The 3'-untraslated region (3'-UTR) is 436 nucleotides long from the stop codon to the first A of the polyadenilate tail, as determined by comparison with two other EST sequences containing a poly(A) stretch (H11393, AA428128). A poly-adenilation recognition site fulfilling the AATAAA consensus sequence lies 26 nucleotides 5' to the beginning of poly(A) tail (from 22751 to 22756 in genomic sequence). BLASTP versus nr analysis of TRUB2 gave as best match a *Mus musculus* protein (AAH15285), which probably represents the murine TRUB2 ortholog. The other matches were with ψ synthase-related proteins of various species. PROSITE search failed to identify any pattern corresponding to *TRUB1* or *TRUB2* gene products when the option 'exclude patterns with a high probability of occurrence' was selected. SMART/Pfam databases search results are listed in Table I. A consistent similarity with TruB domain was detected in both the novel human proteins. An RNA binding domain was detected only in the *DKC1* human gene and not in *TRUB1* or *TRUB2* (Fig. 3).

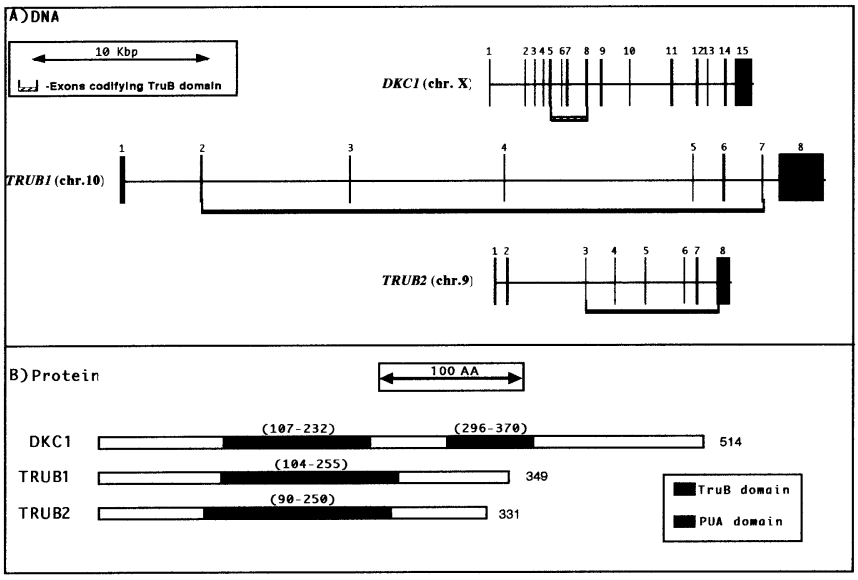

Figure 3. A, schematic presentation of the genomic organization of the three known human *TruB* family genes. The originally identified *TruB* family member, *DKC1*, is represented alongside the newly identified genes, *TRUB1* and *TRUB2*. Exons are symbolized as vertical bars. B, schematic comparison of the three human proteins containing TruB domains (i.e. DKC1, TRUB1 and TRUB2).

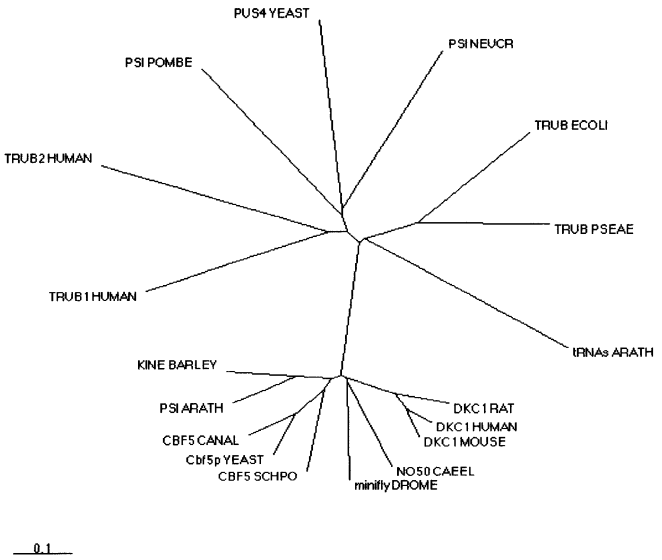

Figure 4. Phylogenetic tree obtained from the described amino acid sequences of TruB showing the positions of the human TRUB1 and TRUB2 proteins. The tree presents two distinct groups of gene products: the human *TRUB1* and *TRUB2* genes are both situated on the (upper) tRNA ψ synthase ramification of the tree, which appears to have separated from the (lower) rRNA ψ synthase ramification. ARATH, *A. thaliana*; BARLEY, *H. vulgare*; CAEEL, *C. elegans*; CANAL: *C. albicans*; ECOLI, *E. coli*; HUMAN, *H. sapiens*; YEAST, *S. cerevisiae*; DROME, *D. melanogaster*; MOUSE, *M. musculus*; NEUCR, *N. crassa*; POMBE, *S. pombe*; PSEASE, *P. aeuriginosa*; RAT, *R. norvegicus*; SCHPO, *S. pombe.*

Sequence alignment. The gene family alignment of amino acid sequences obtained by ClustalW analysis is displayed in Fig. 1. The most prominent feature is the high conservation from bacteria to humans of certain specific amino acids: namely, G115, L120-D121, G126, L128, K147 and Y149 (numbered with respect to TRUB1 positions). Comparison with published data shows that the LD conserved residues belong to the ψ synthase motif II. A broader conserved region, which includes motif II near its amino terminus, matches the described TruB domain from to W104 to I255. Although this domain is characterized by a low degree of similarity among the various family members, K147 and Y149 are conserved from bacteria to humans. Moreover, a motif I containing conserved KP residues is present in the TRUB1 amino acid sequence from V71 to A83. BLAST 2 sequence comparison indicates that the predicted TRUB1 protein shares 24% identity (39% similarity) with DKC1 and 30% identity (45% similarity) with TRUB2.

Molecular evolution analysis. The tree resulting from phylogenetic analysis of amino acid sequences is shown in Fig. 4. Two groups of gene products appear to be related to the ancestral bacterial *TruB*. One group is related to yeast PUS4, showing more similarity to original eubacterial tRNA ψ synthases. The other group is related to yeast Cbfp5, and includes proteins with described functions in rRNA binding and the centromere. Human TRUB1 and TRUB2 are in connection with the tRNA ψ synthase group, and are distinct from the group that includes DKC1.

Discussion

The isomerization of uridine contained in rRNA and tRNA to its C-glycoside isomer ψ is catalized by ψ synthases. Although ψ synthase genes are well characterized in prokaryotes (2-5, 9,31,32) and in yeasts (13,15,33,34), little is known about ψ synthase-like genes in higher eukaryotes (16-19,35). Conservation of primary amino acid sequences during the course of evolution is limited to restricted motifs, the different combinations of which define the four families of ψ synthase proteins described in prokaryotes (6). The weak evolutionary conservation of the overall primary amino acid sequence in ψ synthase proteins has hampered the identification of orthologs in different species by means of global sequence analysis.

In prokaryotes, the TruB ψ synthases are involved in isomerization of U55 in tRNAs (3), while in the lower eukaryotes the TruB family includes two different classes of proteins. One class is related to rRNA modification, and includes members such as yeast Cbf5 (12,13,36) which have gained an RNA binding domain. The other class, which includes yeast Pus4 (15), is more directly related to the original, single TruB found in bacteria. No ψ synthase gene specifically involved in tRNA formation has yet been identified in humans.

The present work reports the identification and characterization of a novel gene, named *TRUB1*, which is the first described human ortholog of bacterial TruB/ψ55, and encodes a product probably related to tRNA pseudouridinilation. This new human TruB-related gene was recognized during analysis of mRNA fragments (i.e. ESTs) found only in keratinocyte libraries aimed at identifying genes of relevance to epithelial molecular biology. Starting from these mRNA sequence fragments, we first reconstructed a new putative mRNA by systematic use of a collection of bioinformatic tools. The characterization was then completed by the actual cloning of cDNA, the determination of the genomic structure of *TRUB1*, and by analysis of its expression pattern and phylogenetic relationships.

The *TRUB1* gene spans ~40 kb on chromosome 10, and includes 8 exons. Its cDNA sequence codes for a product of 349 amino acids. A VFAVHKPKGPTSA box is present in positions 71-83. It includes the two invariant residues KP and corresponds to motif I of the TruB family, which is probably involved in the maintenance of protein structure (8). The TruB domain of TRUB1 lies between W104 and I255, and includes another short motif, GGTLDSAARGVLVV. This includes the highly conserved D residue that characterizes motif II. The D residue is involved in uridine recognition and is important for the catalytic function of ψ synthases (9-11,37).

We investigated the expression pattern of *TRUB1* mRNA by Northern blot analysis. *TRUB1* mRNA turned out to be widely expressed in various types of human tissue. Although the *TRUB1* gene probably codes for a housekeeping function such as RNA modification, we found that the *TRUB1* mRNA expression pattern shows quantitative differences in different tissue types. Among the twelve tissues studied by us, the highest levels of expression were in heart, skeletal muscle and liver. Three hybridization bands were seen, probably representing different maturation stages of mRNA. One of these (the 3.8 kb band) corresponds to the mature mRNA size

predicted by summing the coding sequence to the 3'-UTR calculated on the basis of EST analysis. The smallest of the three transcripts appears to have a specific quantitative regulation, and may represent an isoform generated via alternative splicing or alternative polyadenilation.

Phylogenetic analysis of the TruB domain was performed to explain the origin of the *TRUB1* gene. In order to identify representative members of the TruB family from bacteria to humans, we performed a systematic sequence comparison of the TRUB1 amino acid sequence. An unexpected consequence of this analysis was the identification of a third human gene encoding a conserved TruB domain. This gene, denominated *TRUB2*, is located on human chromosome 9. Thus, the TruB family can now be said to include at least three members in humans: i.e. *DKC1* (previously identified, mutated in the X-linked *dyskeratosis congenita*) plus *TRUB1* and *TRUB2*.

In the original classification of the four ψ synthase families, Koonin (6) noted that even though the entire amino acid sequences of the synthases within each family are not significantly similar, analysis of regions with assigned biological functions suggests that each family of proteins has evolved from a common ancestor. Similarly, whereas *TRUB1* and *TRUB2* show only a weak global similarity to *DKC1*, analysis of single domains and motifs clearly reveals that the three proteins are actually closely related members of the TruB family. Remarkably, motifs I and II have been conserved in all products of the Rlua, RsuA, TruB families, including TRUB1 and TRUB2.

Comparison of TruB members in different species provides some clues regarding the function of TRUB1 and TRUB2. In eukaryotes, the single prokaryotic member of the TruB family appears to have diverged in the two different genes encoding proteins with different functions. Thus, in yeasts Cbf5 presents an rRNA binding domain and a nucleolar localization, suggesting that rRNA is its specific target (13). By contrast, PUS4 acts on tRNA pseudouridination and appears to represent the real yeast ortholog of prokaryotic TruB (15). In humans, DKC1 is the ortholog of yeast Cbf5, as confirmed by its nucleolar localization (38). On the other hand, TRUB1 and TRUB2 are more directly related to bacterial ψ 55 synthase, and could be the hitherto unidentified tRNA ψ synthases in humans, where a further duplication of tRNA modification related genes seems to have occurred. Functional studies should clarify the different activities of TRUB1 and TRUB2 on tRNA.

Since ψ synthase-like proteins participate in the modifications required for tRNA structure completion, mutations in their genes might have been expected to cause severe alterations in protein synthesis. Surprisingly, however, disruption of the ψ 55 synthase gene in yeast and *E. coli* did not give rise to a significant phenotype in the resulting ψ 55-deficient tRNA (15,39,40). It is likely that in higher eukaryotes TruB family proteins gain new and/or modified functions, such as in telomere maintenance (24), and that they are responsible for disease when mutated (19-21). Human X-linked form of *dyskeratosis congenita* is caused by mutations in *DKC1* (38). Even though TRUB1 lacks high similarity to DKC1 and putatively affects tRNA rather than rRNA modification, it is still the most similar product to dyskerin found in humans. Since the ψ synthases of higher eukaryotes have acquired different functions, involvement of TRUB1 in dyskeratosis-like diseases or even in dyskeratosis itself cannot be excluded. Study of human TRUB1 (and also TRUB2) should allow us to gain an understanding of the molecular mechanisms of RNA modification in higher eukaryotes, and to investigate the involvement of ψ synthases in human pathology.

Acknowledgments

We are grateful to Gabriella Mattei for her excellent technical assistance with automatic sequencing, and to Dr Elspeth Bruford (HUGO Nomenclature Committee) for her kind and expert assistance in gene nomenclature. The Authors thank Mr Robin Cooke for helping revise the manuscript. This work was supported by grants from MURST 'ex 60%' (to C.Z.) and from the 'Fondazione CARISBO Center for Research into Molecular Genetics'.

References

1. Grosjean H and Benne R: In: Modification and Editing of RNA. ASM Press, Washington DC, 1998.
2. Marvel CC, Arps PJ, Rubin BC, Kammen HO, Penhoet EE and Winkler ME: hisT is part of a multigene operon in *Escherichia coli* K-12. J Bacteriol 161: 60-71, 1998.
3. Nurse K, Wrzesinski J, Bakin A, Lane BG and Ofengand J: Purification, cloning, and properties of the tRNA psi 55 synthase from *Escherichia coli*. RNA 1: 102-112, 1995.
4. Wrzesinski J, Nurse K, Bakin A, Lane BG and Ofengand J: A dual-specificity pseudouridine synthase: an *Escherichia coli* synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). RNA 1: 437-448, 1995.
5. Wrzesinski J, Bakin A, Nurse K, Lane BG and Ofengand J: Purification, cloning, and properties of the 16S RNA pseudouridine 516 synthase from *Escherichia coli*. Biochemistry 34: 8904-8913, 1995.
6. Koonin EV: Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. Nucleic Acids Res 24: 2411-2415, 1996.
7. Gustafsson C, Reid R, Greene PJ and Santi DV: Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. Nucleic Acids Res 24: 3756-3762, 1996.
8. Spedaliere CJ, Hamilton CS and Mueller EG: Functional importance of motif I of pseudouridine synthases: mutagenesis of aligned lysine and proline residues. Biochemistry 39: 9459-9465, 2000.
9. Huang L, Pookanjanatavip M, Gu X and Santi DV: A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. Biochemistry 37: 344-351, 1998.
10. Ramamurthy V, Swann SL, Paulson JL, Spedaliere CJ and Mueller EG: Critical aspartic acid residues in pseudouridine synthases. J Biol Chem 274: 22225-22230, 1999.
11. Conrad J, Niu L, Rudd K, Lane BG and Ofengand J: 16S ribosomal RNA pseudouridine synthase RsuA of *Escherichia coli:* deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases. RNA 5: 751-763, 1999.
12. Jiang W, Middleton K, Yoon HJ, Fouquet C and Carbon J: An essential yeast protein, CBF5p, binds *in vitro* to centromeres and microtubules. Mol Cell Biol 13: 4884-4893, 1993.
13. Lafontaine DL, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M and Tollervey D: The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev 12: 527-537, 1998.
14. Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M and Luhrmann R: Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. RNA 4: 1549-1568, 1998.

15. Becker HF, Motorin Y, Planta RJ and Grosjean H: The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res 25: 4493-4499, 1997.
16. Phillips B, Billin AN, Cadwell C, *et al:* The Nop60B gene of Drosophila encodes an essential nucleolar protein that functions in yeast. Mol Gen Genet 260: 20-29, 1998.
17. Giordano E, Peluso I, Senger S and Furia M: Minifly, a Drosophila gene required for ribosome biogenesis. J Cell Biol 144: 1123-1133, 1999.
18. Meier UT and Blobel G: NAP57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. J Cell Biol 127: 1505-1514, 1994.
19. Heiss NS, Knight SW, Vulliamy TJ, *et al:* X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet 19: 32-38, 1998.
20. Knight SW, Heiss NS, Vulliamy TJ, *et al:* X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene. Am J Hum Genet 65: 50-58, 1999.
21. Vulliamy TJ, Knight SW, Heiss NS, Smith OP, Poustka A, Dokal I and Mason PJ: Dyskeratosis congenita caused by a 3' deletion: germline and somatic mosaicism in a female carrier. Blood 94: 1254-1260, 1999.
22. Drachtman RA and Alter BP: Dyskeratosis congenita: clinical and genetic heterogeneity. Report of a new case and review of the literature. Am J Pediatr Hematol Oncol 14: 297-304, 1992.
23. McKusick VA: Mendelian Inheritance in Man. Catalogs of Human Genes and Genetic Disorders. Baltimore, Johns Hopkins University Press, 1998.
24. Mitchell JR, Wood E and Collins K: A telomerase component is defective in the human disease dyskeratosis congenita. Nature 402: 551-555, 1999.
25. Altschul SF, Madden TL, Schaffer AA, Zhan J, Zhang Z, Miller W and Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402, 1997.
26. Chomczynski P and Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156-159, 1987.
27. Church GM and Gilbert W: Genomic sequencing. Proc Natl Acad Sci USA 81: 1991-1995, 1984.
28. Thompson JD, Higgins DG and Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680, 1994.
29. Bateman A, Birney E, Durbin R, Eddy SR, Finn RD and Sonnhammer EL: Pfam 3.1: 1313 multiple alignments match the majority of proteins. Nucleic Acids Res 27: 260-262, 1999.
30. Kozak M: An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15: 8125-8148, 1987.
31. Conrad J, Sun D, Englund N and Ofengand J: The rluC gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23 S ribosomal RNA. J Biol Chem 273: 18562-18566, 1998.
32. Kammen HO, Marvel CC, Hardy L and Penhoet EE: Purification, structure, and properties of *Escherichia coli* tRNA pseudouridine synthase I. J Biol Chem 263: 2255-2263, 1988.
33. Simos G, Tekotte H, Grosjean H, Segref A, Sharma K, Tollervey D and Hurt EC: Nuclear pore proteins are involved in the biogenesis of functional tRNA. EMBO J 15: 2270-2284, 1996.
34. Lecointe F, Simos G, Sauer A, Hurt EC, Motorin Y and Grosjean H: Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of psi 38 and psi 39 in tRNA anticodon loop. J Biol Chem 273: 1316-1323, 1998.
35. Chen J and Patton JR: Cloning and characterization of a mammalian pseudouridine synthase. RNA 5: 409-419, 1999.
36. Cadwell C, Yoon HJ, Zebarjadian Y and Carbon J: The yeast nucleolar protein Cbf5p is involved in rRNA biosynthesis and interacts genetically with the RNA polymerase I transcription factor RRN3. Mol Cell Biol 17: 6175-6183, 1997.
37. Raychaudhuri S, Niu L, Conrad J, Lane BG and Ofengand J: Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. J Biol Chem 274: 18880-18886, 1999.
38. Heiss NS, Girod A, Salowsky R, Wiemann S, Pepperkok R and Poustka A: Dyskerin localizes to the nucleolus and its mislocalization is unlikely to play a role in the pathogenesis of dyskeratosis congenita. Hum Mol Genet 8: 2515-2524, 1999.
39. Cole JR, Olsson CL, Hershey JW, Grunberg-Manago M and Nomura M: Feedback regulation of rRNA synthesis in *Escherichia coli*. Requirement for initiation factor IF2. J Mol Biol 198: 383-392, 1987.
40. Sands JF, Regnier P, Cummings HS, Grunberg-Manago M and Hershey JW: The existence of two genes between infB and rpsO in the *Escherichia coli* genome: DNA sequencing and S1 nuclease mapping. Nucleic Acids Res 16: 10803-10816, 1988.