Acta Biochimica Polonica

Vol. 55 No. 3/2008, 473–478

on-line at: www.actabp.pl

Regular paper

Characterization of a novel transcript variant of human *STAU1* gene★

Nie Fa-hui[1, 2], Yao Hai-feng[1], Qi Rui-feng[3], Li Xin[4]✉ and Wu Cai-bin[2]✉

[1]*State Key Laboratory of Pollution Control and Resource Reuse, Tongji University, Shanghai, PR China;* [2]*Institute of Environment Engineering Research, East China Jiao Tong University, Nanchang, PR China;* [3]*Institute of Protein Research, Tongji University, Shanghai, PR China;* [4]*Institute of Biomedical Science, Fudan University, Shanghai, PR China*

Received: 09 March, 2008; revised: 11 August, 2008; accepted: 15 August, 2008
available on-line: 20 September, 2008

Human STAU1 is one member of the family of double-stranded RNA (dsRNA)-binding proteins. It is thought to function in transporting mRNA, controlling translation and eliciting mRNA decay in neurons, and to function in infection of influenza virus and human immunodeficiency virus type 1 (HIV-1). Four transcripts coding two isoforms have been identified before. In this study, we have isolated a novel transcript of STAU1, coding a novel isoform that has six amino acids more (SFPLKQ) than isoform a. In order to examine the tissue distribution of this novel isoform, we have performed RT-PCR experiments and the analysis showed that it was highly expressed in heart, liver, kidney and pancreas.

Keywords: STAU1, expression pattern

INTRODUCTION

Drosophila Staufen, homologous human STAU1, is the first RNA-binding protein proven to play a role in RNA localization. Staufen and its homologous genes play a key role in many biological phenomena. One of the best known examples is determination of the anteroposterior axis during the early development of *Drosophila melanogaster*. Staufen plays a role in the transportation and location of *bicoid* and *oskar* mRNAs to the anterior and posterior poles, respectively (Schuldt *et al.*, 1998**)**. The specfic localization of mRNAs is also essential in the definition of cell asymmetries in cell division and differentiation. Staufen transports *prospero* mRNA to basal area of fly mitotic neuroblasts (Li *et al.*, 1997). Cell comunication events are also affected by the localization of specific mRNAs. Human Stau1 can interact with many proteins and locates in the dendrites of differentiated neuroblasts (Kiebler & DesGroseillers, 2000). Human *STAU1* functions in infection by influenza virus and human immunodeficiency virus type 1 (HIV-1). It can interact with NS1 protein of influenza virus in infected cells (Falcón *et al.*, 1999), and it can also interact with the NC domain of HIV-1 $pr55^{Gap}$ and may be make the virus generate infectious viral particles (Chatel-Chaix *et al.*, 2004). In addition to these biological processes, *Drosophila* Staufen function in the translation derepression of *oskar* mRNA once the mRNA has been located (Micklem *et al.*, 2000). Human STAU1 causes the decay of many mRNAs through binding Upf1 and interacting with the 3′ untranslated regions of mRNAs (Kim *et al.*, 2005).

Drosophila Staufen has five double-stranded RNA-binding domains (dsRBD) (Gibson *et al.*, 1994). Each domain contains a 65- to 68-amino-acid consensus sequence and folds into a compact $\alpha\beta\beta\beta\alpha$ structure. The analysis of crystal structure shows that dsRBD recognizes the shape of A-form dsRNA through two regions between β2-β1 and β3-α2 (Ramos *et al.*, 2000). The composition of Stau-containing RNA has been identified in neuroblasts, including ribosomes and ribosome components (ribosomal protein S6, Ll2, FMRP, PABP,

✉Corresponding authors: Li Xin, lxstaroad@yahoo.com or Wu Cai-bin, wucaibin@ecjtu.jx.cn

★The nucleotide sequence reported in this paper has been submitted to GenBank under accession number AY536098.

Abbreviation: dsRBD, double-stranded RNA-binding domain; G3PD, glycaraldehyde-3-phosphate dehydrogenase gene; NS1, nonstructural protein; RT-PCR, reverse transcriptional PCR; Stau, Staufen protein.

S40, S60 etc); proteins from the cytoskeleton, motor proteins and regulatory proteins (tubulins, actin, internexin, myosin, IQGAP-1, cdc42, rac1, RasGAP, dynein, kinesin etc.); and other RNA-binding proteins normally localized to the nucleus (nucleolin, RNA helicase A, hnRNP U etc.) (Villace *et al.*, 2004). In addition, STAU1 can bind protein phosphatase-1 in primary hippocampal neurons (Monshausen *et al.*, 2002), Upf1 and Arf1 proteins in HeLa cells, NS1 protein of influenza virus and $pr55^{Gap}$ protein of HIV-1 (Falcon *et al.*, 1999; Chatel-Chaix *et al.*, 2004; Kim *et al.*, 2005). Therefore, the component is very complex and human Stau1 may be plays vital roles.

Four transcript variants resulting from human *STAU1* gene and encoding two isoforms (a, b) with different N-terminal ends have been described (Marión *et al.*, 1999). Three of these variants encode the same isoform (a), however, differ in their 5' UTR. In contrast to *Drosophila* Staufen, human STAU1 has only four dsRBDs. However, the significance of the various transcripts is not known today. In a recent study, we have isolated a novel transcript variant of *STAU1* through a large-scale sequencing analysis of the 18-week human fetal brain cDNA library. We termed it *STAU1* transcript variant T5 (isoform c). Here, we report the cloning and initial characterization of this new transcript variant, together with data in its genomic structure and mRNA tissue distribution.

MATERIALS AND METHODS

The cloning and bioinformatics analysis. cDNA library construction and DNA sequencing were performed as previously described (Li *et al.*, 2005). DNA sequence homology searches and comparisons were performed using BLAST-N and BLAST-X at the National Center for Biotechnology Information (NCBI) network service (http://www.ncbi.nlm.nih.gov/blast). BLAST-N searching in the human genome was performed to identify the chromosomal localization and the gene structure of this novel gene. The predicted amino-acid sequence of the *STAU1* T5 was compared against the profile entries to find the occurrence of known profiles (http://www.expasy.ch/pfscan). To identify the chromosomal localization and the gene structure, multiple alignments were performed by the GeneDoc program (http://www.psc.edu/biomed/genedoc/). Other databases from Genbank, Swiss Pro, PDB, and EXPASY were also used. The software includes Gene Runner, Primer Premier 5.0 *et al.*

Assessment of human *STAU1* T5 mRNA tissue distribution. Human Multiple Tissue cDNA (MTC) panels (CLONTECH) were used as PCR templates according to the manufacturer's protocol. Thirty six PCR cycles for *STAU1* T5 and 30 PCR cycles for *G3PDH* (as control) were performed using *Taq* Plus polymerase (Sangon) in the following program: 0.5 min at 94ºC, 0.5 min at 65ºC, 1.0 min at 72ºC. The PCR products of *STAU1* T5 and *G3PDH* were then electrophoresed on a 2% agarose gel. PCR primers are indicated from 5′ to 3′ as the following: *STAU1* T5 sense CTCTCTCGGCTCCCGCTTCCTTT; *STAU1* T5 antisense ACCTGTTTCAGAGGGAAAGACTCG; *G3PDH* sense: TGAAGGTCGGAGTCAACGGATTTGGT; *G3PDH* antisense CATGTGGGCCATGAGGTCCACCAC. The sense and antisense primers of human *STAU1* T5 span 688 bp in the cDNA from 61 to 748 bp.

RESULTS AND DISCUSSION

Through the large-scale cDNA sequencing, we cloned a novel transcript variant from the constructed human fetal brain cDNA library. This cDNA is composed of an open reading frame from nucleotide 363 to 1868, encoding a 502 amino-acid protein with a molecular mass of 55 kDa (Genbank: AY546099) (Fig. 1). The deduced protein shows high identity with the mouse and *Drosophila* Staufen2 proteins. Alignment was performed between STAU1 T5, other two isoforms in human, and its orthologs from mouse and *Drosophila*. Like other Staufens, the deduced protein of T5 also contains four dsRNA-binding domains (dsRBDs) (Fig. 2). Universally, dsRBD2 contains an insertion sequence at an identical position in all species, and the domain does not bind RNA when this insertion is deleted. Surprisingly, we found another insertion of 18 bp in the sequence, which is deduced into six amino acids between dsRBD1 and dsRBD2. This new insertion may suggest a new kind of mechanism or method of RNA localization during the cell cycle or embryo development. Moreover, through ortholog analysis (not shown), the insertion of the 18 bp sequence is also present in certain Staufen protein in *Drosophila*, which suggests that this protein may be conserved in evolution and that this insertion may have something to do with the specificity of its binding in different tissues of human and *Drosophila*. However, on its own Staufen binds mRNA without apparent specificity, indicating that the specificity might be provided by auxiliary factors.

Four transcript variants resulting from alternative splicing of the *STAU1* gene and encoding two isoforms with different N-terminal ends have been reported. Three of these variants encode the same isoform, however, differ in their 5' UTR.

```
1                                                                   ac

3     ttcctgccgggctgcgggcgcctgagcgctcttcagcgtttgcgcggcggctgcgcgtctctctcggctcccgcttcctttgaccgcctc

93    ccccccccggcccggcggcgcccgcctcctccacggccactccgcctcttccctccttcgtccttcttcctctccctttttttccttct

183   tccttcccctcctcgccgccaccgcccaggaccgccggccgggggacgagctcggagcagcagccagagtttattaaccacttaacctct

273   cagaactgaacaaagacaacattgttcctggaacgccctctttttaaaaaagaaagcataacccctactgtagaactaaatgcactgtgc

363   atgaaacttggaaaaaaaccaatgtataagcctgttgacccttactctcggatgcagtccacctataactacaacatgagaggaggtgct

        M  K  L  G  K  K  P  M  Y  K  P  V  D  P  Y  S  R  M  Q  S  T  Y  N  Y  N  M  R  G  G  A

453   tatcccccgaggtacttttacccatttccagttccacctttactttatcaagtggaactttctgtgggaggacagcaatttaatggcaaa

        Y  P  P  R  Y  F  Y  P  F  P  V  P  P  L  L  Y  Q  V  E  L  S  V  G  G  Q  Q  F  N  G  K

543   ggaaagacaagacaggctgcgaaacacgatgctgctgccaaagcgttgaggatcctgcagaatgagcccctgccagagaggctggaggtg

        G  K  T  R  Q  A  A  K  H  D  A  A  A  K  A  L  R  I  L  Q  N  E  P  L  P  E  R  L  E  V

633   aatggaagagaatccgaagaagaaaatctcaataaatctgaaataagtcaagtgtttgagattgcacttaaacggaacttgcctgtgaat

        N  G  R  E  S  E  E  E  N  L  N  K  S  E  I  S  Q  V  F  E  I  A  L  K  R  N  L  P  V  N

723   ttcgagtctttccctctgaaacaggtggcccgggagagtggcccaccccacatgaagaactttgtgaccaaggtttcggttggggagttt

        F  E  S  F  P  L  K  Q  V  A  R  E  S  G  P  P  H  M  K  N  F  V  T  K  V  S  V  G  E  F

813   gtgggggaaggtgaagggaaaagcaagaagatttcaaagaaaaatgccgccatagctgttcttgaggagctgaagaagttaccgcccctg

        V  G  E  G  E  G  K  S  K  K  I  S  K  K  N  A  A  I  A  V  L  E  E  L  K  K  L  P  P  L

903   cctgcagttgaacgagtaaagcctagaatcaaaaagaaaacaaaacccatagtcaagccacagacaagcccagaatatggccaggggatc

        P  A  V  E  R  V  K  P  R  I  K  K  K  T  K  P  I  V  K  P  Q  T  S  P  E  Y  G  Q  G  I

993   aatccgattagccgactggcccagatccagcaggcaaaaaaggagaaggagccagagtacacgctcctcacagagcgaggcctcccgcgc

        N  P  I  S  R  L  A  Q  I  Q  Q  A  K  K  E  K  E  P  E  Y  T  L  L  T  E  R  G  L  P  R

1083  cgcagggagtttgtgatgcaggtgaaggttggaaaccacactgcagaaggaacgggcaccaacaagaaggtggccaagcgcaatgcagcc

        R  R  E  F  V  M  Q  V  K  V  G  N  H  T  A  E  G  T  G  T  N  K  K  V  A  K  R  N  A  A

1173  gagaacatgctggagatccttggtttcaaagtcccgcaggcgcagcccaccaaacccgcactcaagtcagaggagaagacacccataaag

        E  N  M  L  E  I  L  G  F  K  V  P  Q  A  Q  P  T  K  P  A  L  K  S  E  E  K  T  P  I  K

1263  aaaccaggggatggaagaaaagtaaccttttttgaacctggctctggggatgaaaatgggactagtaataaagaggatgagttcaggatg

        K  P  G  D  G  R  K  V  T  F  F  E  P  G  S  G  D  E  N  G  T  S  N  K  E  D  E  F  R  M

1353  ccttatctaagtcatcagcagctgcctgctggaattcttcccatggtgcccgaggtcgcccaggctgtaggagttagtcaaggacatcac

        P  Y  L  S  H  Q  Q  L  P  A  G  I  L  P  M  V  P  E  V  A  Q  A  V  G  V  S  Q  G  H  H

1443  accaaagattttaccagggcagctccgaatcctgccaaggccacggtaactgccatgatagcccgagagttgttgtatgggggcacctcg
```

```
      T  K  D  F  T  R  A  A  P  N  P  A  K  A  T  V  T  A  M  I  A  R  E  L  L  Y  G  G  T  S

1533  cccacagccgagaccattttaaagaataacatctcttcaggccacgtaccccatggacctctcacgagaccctctgagcaactggactat

      P  T  A  E  T  I  L  K  N  N  I  S  S  G  H  V  P  H  G  P  L  T  R  P  S  E  Q  L  D  Y

1623  ctttccagagtccagggattccaggttgaatacaaagacttccccaaaaacaacaagaacgaatttgtatctcttatcaattgctcctct

      L  S  R  V  Q  G  F  Q  V  E  Y  K  D  F  P  K  N  N  K  N  E  F  V  S  L  I  N  C  S  S

1713  cagccacctctgatcagccatggtatcggcaaggatgtggagtcctgccatgatatggctgcgctgaacatcttaaagttgctgtctgag

      Q  P  P  L  I  S  H  G  I  G  K  D  V  E  S  C  H  D  M  A  A  L  N  I  L  K  L  L  S  E

1803  ttggaccaacaaagtacagagatgccaagaacaggaaacggaccaatgtctgtgtgtgggaggtgctgaacctttctggccatgaacca

      L  D  Q  Q  S  T  E  M  P  R  T  G  N  G  P  M  S  V  C  G  R  C

1893  ttataaaatcccaacatatatactgaaaatactgaaactgctttgaaaatttggaatttctgatacctccagtgggccgagagacacggt

1983  gggtaaaggatgtgggcagcagcagggaagacaacagaaacacaaggaggcggctgtggccgggctggactgtgcgggggtttgttgtga

2073  tggccactcggtgacctggcggtccctacgcaatagcagctgcctgtggggaagaggggctgcccagccagctggttctcccgggacacc

2163  agcagatccacaccctgggcacctccgtgtttggtctttttttttcccctgtgtgaaagaagaaacggcacgacccccttctcaagctggct

2253  cactcagacacattgggacaaaccctggacagccatgccagagagaggcctttgaccggccccagagctaaaagcaccagagaaaatcaa

2343  atgcttcctactcagcgtgacccaacttttctagtgtgccacggccccaccacctcctgcagtacccacaccatcaccactgctttctct

2433  tccaacagtgatctgtattcttagtttcattattttctttgattgatatgacactatataaaattttcatttgagaatttctcaattgt

2523  atctagttaaatagcacagtttggaaacttgtctgagactgactttatcaataatctaaccgacaaagatcatatccatgtgtatgtggt

2613  tagacatttttatttcattgactaacccaggacagtttcagtgatgcaaattgtgtgccctctggttcagctgaaacagtcctggacttt

2703  caaaaaccttgaataagtctcccacagttgtataaattggacaatttaggaattttaaactttagatgatcatttggttccatttttatt

2793  tcatttttatttttgttaatgcaaacaggacttaaatgaactttgatctctgttttaaagattattaaaaaacattgtgtatctatacat

2883  atggctcttgaggacttagctttcactacactacaggatatgatctccatgtagtccatataaacctgcagagtgattttccagagtgct

2973  cgatactgttaattacatctccattagggctgaaaagaatgacctacgtttctgtatacagctgtgttgcttttgatgttgtgttactgt

3063  acacagaagtgtgtgcactgaggctctgcgtgtggtccgtatggaaagcctggtagccctgcgagttaagtactgcttccattcattgtt

3153  tacgctggaatttttctccccatggaatgtaagtaaaacttaagtgtttgtcatcaataaatggtaatactaaatttttttgttaattta

3243  ttctcaaatgccactactgctaggttggtcccctcccaacttgc
```

Figure 1. The transcript variant cDNA and deduced amino-acid sequences of *STAU1* T5.
The nucleotides are numbered at the left. The extra amino-acid residues of Stau1 T5 compared to other isoforms are boxed.

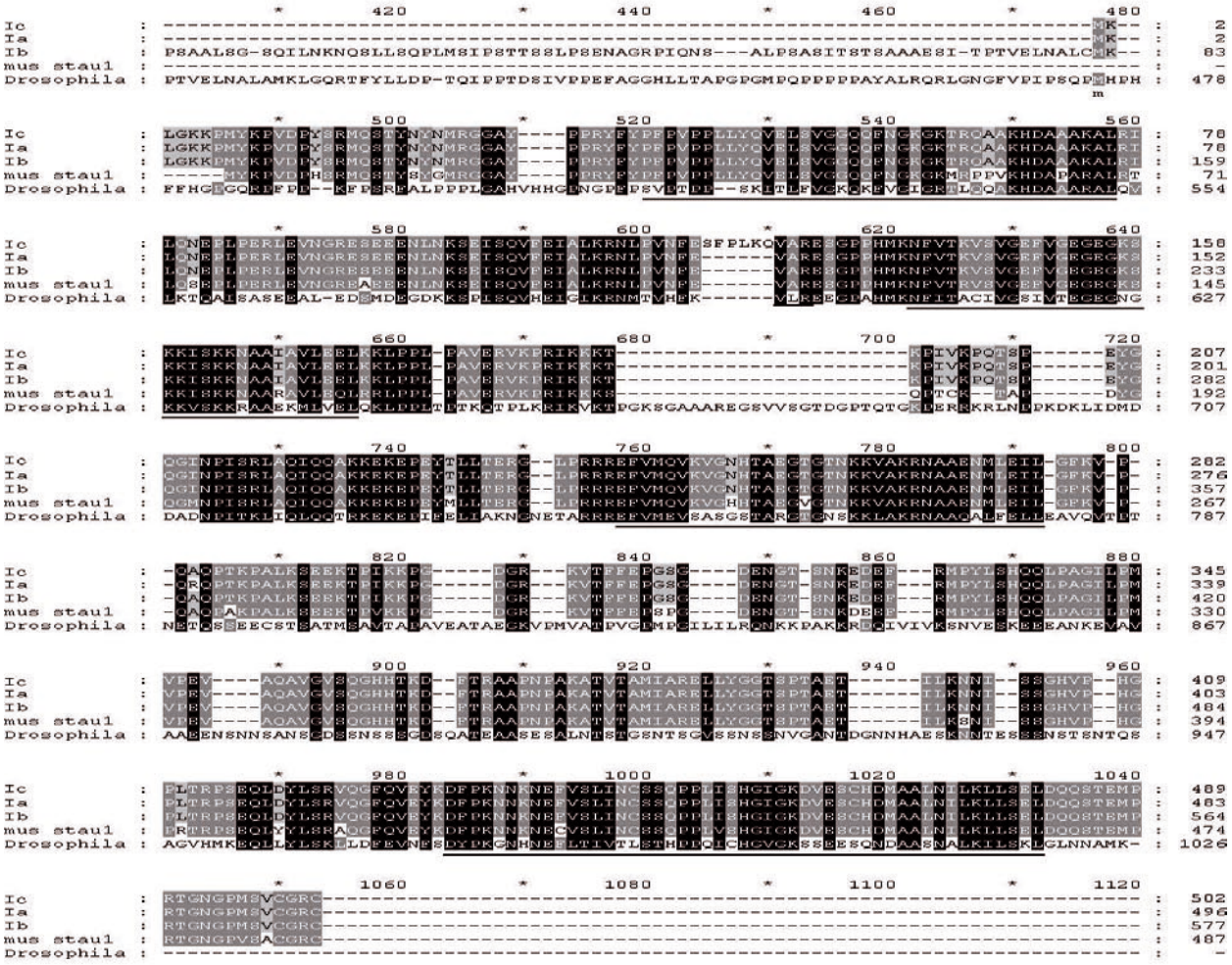

Figure 2. Alignment of STAU1 c and a (NP_004593), b (NP_059347) and its orthologs in mouse Stau1 (NP_035620) and *Drosophila* (NP_476751).
The alignment was performed by GeneDoc program (http://www.psc.edu/biomed/genedoc/): Black (100% similarity), grey (80–90% similarity); light grey (60–70% similarity). Domains of the 5 dsRBDs are underlined.

Table 1. The exon-intron analysis of the human *STAU1* T5 gene.

Exon		Splicing acceptor	Splicing donor	Intron	
Number	Length (bp)			Number	Length (bp)
1	249		CAGCAGCCAG***gt***gaggcggc	1	13846
2	75	tattttcc***ag***AGTTTATTAA	TTTAAAAAAG***gt***acatataa	2	20123
3	139	tttcttgc***ag***AAAGCATAAC	ATCCCCCGAG***gt***atgtgttt	3	2185
4	166	ttatttct***ag***GTACTTTTAC	GAGGCTGGAG***gt***gaggagtt	4	15650
5	99	attcttac***ag***GTGAATGGAA	GAATTTCGAG***gt***aagctaat	5	11227
6	231	tttcactc***ag***TCTTTCCCTC	CATAGTCAAG***gt***gagaactt	6	1139
7	144	tgttcctc***ag***CCACAGACAA	TGTGATGCAG***gt***gggtccgc	7	2963
8	147	ctttgtgc***ag***GTGAAGGTTG	AGAGGAGAAG***gt***gagtgctg	8	1573
9	76	ctttttta***ag***ACACCCATAA	AATGGGACTA***gt***aagtgtga	9	232
10	320	tccctccc***ag***GTAATAAAGA	GGGATTCCAG***gt***aactgtct	10	528
11	123	tccccctt***ag***GTTGAATACA	CCATGATATG***gt***acgtcaca	11	1258
12	86	tccccact***ag***GCTGCGCTGA	ACCAATGTCT***gt***gtgagtgc	12	888
13	1431	ttccttaa***ag***GTGTGGGAGG			

Intron and exon nucleotide sequences are shown in lower-case and upper-case letters, respectively. Bold italics lettering indicates donor and acceptor splice site.

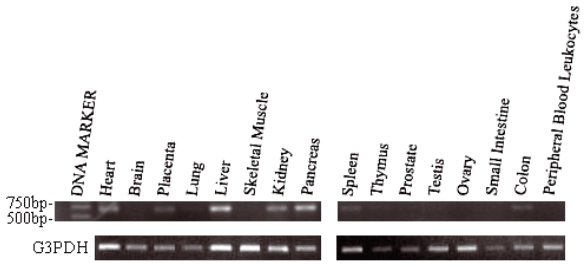

Figure 3. Tissue distribution of human *STAU1* T5 mRNA.
Reverse transcription-PCR analysis of human cDNA for *STAU1* T5 and *G3PDH* (as a control). Prenormalized cDNAs from sixteen human adult tissues were purchased from CLONTECH and employed as templates in PCR reactions containing *STAU1* T5 and *G3PDH*-specific primers described in Materials and Methods.

Isoform (a) associates with 40S and 60S ribosomal subunits and colocalizes with rough endoplasmic reticulum in neuroblasts (Wickham *et al*., 1999). Transcript variant T5 has an insertion of 18 bases at position 729 compared to other four transcript variants of *STAU1*. This introduces an insertion of six amino acids in the deduced protein, causing T5 to encode a protein isoform (c) of 502 amino-acids protein. So we termed it human Stau1 variant 5 and the deduced protein isoform (c). All the sequences of the exon-intron junctions are consistent with the AG-GT rule (Table 1).

Expression pattern of the *STAU1* T5 was analyzed by RT-PCR. The data demonstrate that it is especially expressed in certain tissues we used. The expression level in heart, liver, kidney and pancreas are relatively high, while there is slightly lower expression in placenta, colon (Fig. 3). It is different from the other four transcripts that are expressed in heart, brain, pancreas, skeletal muscles, liver, placenta, lung, kidney (Wickham *et al*., 1999). Therefore, further studies will be necessary to define the precise roles of three isoforms.

REFERENCE

Chatel-Chaix L, Clément JF, Martel C, Bériault V, Gatignol A, DesGroseillers L, Mouland AJ (2004) Identification of Staufen in the human immunodeficiency virus type 1 Gag ribonucleoprotein complex and a role in generating infectious viral particles. *Mol Cell Biol* **24**: 2637–2648.

Falcón AM, Fortes P, Marión RM, Beloso A, Ortín J (1999) Interaction of influenza virus NS1 protein and the human homologue of Staufen *in vivo* and *in vitro*. *Nucleic Acids Res* **27**: 2241–2247.

Gibson TJ, Thompson JD (1994) Detection of dsRNA-binding domains in RNA helicase A and Drosophila maleless: implications for monomeric RNA helicases. *Nucleic Acids Res* **22**: 2552–2556.

Kiebler MA, DesGroseillers L (2000) Molecular insights into mRNA transport and local translation in the mammalian nervous system. *Neuron* **25**: 19–28.

Kim YK, Furic L, Desgroseillers L, Maquat LE (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3′UTRS so as to elicit mRNA decay. *Cell* **120**: 195–208.

Li P, Yang X, Wasser M, Cai Y, Chia W (1997) Inscuteable and Staufen mediate asymmetric localization and segregation for *prospero* RNA during Drosophila neuroblast cell divisions. *Cell* **90**: 437–447.

Li X, Ji C, Gu J, Xu J, Jin Z, Sun L, Zou X, Lin Y, Sun R, Wang P, Gu S, Mao Y (2005) Molecular cloning and characterization of AAAS-V2, a novel splice variant of human AAAS. *Mol Biol Rep* **32**: 127–131.

Marión RM, Fortes P, Beloso A, Dotti C, Ortín J (1999) A human sequence homologue of Staufen is an RNA-binding protein that is associated with polysomes and localizes to the rough endoplasmic reticulum. *Mol Cell Biol* **19**: 2212–2219.

Micklem DR, Adams J, Grünert S, St Johnston D (2000) Distinct roles of two conserved Staufen domains in *oskar* mRNA localization and translation. *EMBO J* **19**: 1366–1377.

Monshausen M, Rehbein M, Richter D, Kindler S (2002) The RNA-binding protein Staufen from rat brain interacts with protein phosphatase-1. *J Neurochem* **81**: 557–564.

Ramos A, Grünert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G (2000) RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* **19**: 997–1009.

Schuldt AJ, Adams JH, Davidson CM, Micklem DR, Haseloff J, St Johnston D, Brand AH (1998) Miranda mediates asymmetric protein and RNA localization in the developing nervous system. *Genes Dev* **12**: 1847–1857.

Villacé P, Marión RM, Ortín J (2004) The composition of Staufen-containing RNA granules from human cells indicates their role in the regulated transport and translation of messenger RNAs. *Nucleic Acids Res* **32**: 2411–2420.

Wickham L, Duchaîne T, Luo M, Nabi IR, DesGroseillers L (1999) Mammalian Staufen is a double-stranded-RNA- and tubulin-binding protein which localizes to the rough endoplasmic reticulum *Mol Cell Biol* **19**: 2220–2230.