814–821 Nucleic Acids Research, 1997, Vol. 25, No. 4

Cloning and characterization of HUPF1, a human homolog of the *Saccharomyces cerevisiae* nonsense mRNA-reducing UPF1 protein

Steven E. Applequist[+], Manuel Selg, Chander Raman[1] and Hans-Martin Jäck*

Department of Microbiology and Immunology and Program in Molecular Biology, Stritch School of Medicine, Loyola University of Chicago, Maywood, IL 60153, USA and [1]Hospital for Special Surgery, Cornell University, New York, NY 10021, USA

Received October 18, 1996; Revised and Accepted December 18, 1996

DDBJ/EMBL/GenBank accession no. U59323

ABSTRACT

Levels of most nonsense mRNAs are normally reduced in prokaryotes and eukaryotes when compared with that of corresponding functional mRNAs. Genes encoding polypeptides that selectively reduce levels of nonsense mRNA have so far only been identified in simple eukaryotes. We have now cloned a human cDNA whose deduced amino acid sequence shows the highest degree of homology to that of UPF1, a *bona fide Saccharomyces cerevisiae* group I RNA helicase required for accelerated degradation of nonsense mRNA. Based on the total sequence of the shorter yeast UPF1 protein, the overall identity between the human protein and UPF1 is 51%. Besides NTPase and other RNA helicase consensus motifs, UPF1 and its human homolog also share similar putative zinc finger motifs that are absent in other group I RNA helicases. Northern blot analysis with the human cDNA probe revealed two transcripts in several human cell lines. Further, antibodies raised against a synthetic peptide of the human polypeptide detected a single 130 kDa polypeptide on Western blots from human and mouse cells. Finally, immunofluorescence and Western blot analyses revealed that the human and mouse polypeptides, like yeast UPF1, are expressed in the cytoplasm, but not in the nucleus. We have thus identified the first mammalian homolog of yeast UPF1, a protein that regulates levels of nonsense mRNA, and we tentatively name this protein human HUPF1 (for human *h*omolog of *UPF1*).

INTRODUCTION

An mRNA with a premature translational stop codon (nonsense mRNA) can originate from mistakes during post-transcriptional events such as RNA editing and RNA splicing (reviewed in 1). Alternatively, nonsense mRNA can be transcribed from a germline or somatically mutated gene, from a pseudo-gene (2) or from non- productively rearranged immunoglobulin (Ig) and T cell receptor (TCR) genes (reviewed in 3). A nonsense mRNA encodes a truncated polypeptide that, if it accumulates, can affect normal cellular processes in a dominant-negative or gain-of-function fashion. Thus, high levels of nonsense mRNA might influence the growth, differentiation or other physiological functions of a cell (4–6).

Cytoplasmic levels of most nonsense mRNAs are normally reduced in higher and simple eukaryotes, as well as prokaryotes, when compared with that of their corresponding functional mRNAs (reviewed in 1,7–9). Genes that encode polypeptides required to reduce levels of nonsense mRNA have so far only been identified in the yeast *Saccharomyces cerevisiae* (10–13) and *Caenorhabditis elegans* (6). One example of such a gene is the *S.cerevisiae UPF1* gene (for *up-f*rameshift mutation *1*; 10), also known as the *NAM7* gene (for *n*uclear *a*ccommodation of *m*itochondria; 14). However, a mammalian gene encoding a protein that controls levels of nonsense mRNA has not previously been identified or cloned.

Using combined comparative genomics and cDNA library screenings, we isolated and characterized a human cDNA clone that encodes a structural homolog of yeast UPF1. Hence, we suggest naming this protein human HUPF1 (for human *h*omolog of *UPF1*).

MATERIALS AND METHODS

Cell culture and cell lines

All cell lines were grown in complete RPMI medium as described (15). VXH is a murine B cell hybridoma line (16). Human cell lines used in this study are the plasmacytoma line MC/CAR (American Type Culture Collection no. CRL8083), the heart muscle line HA-VSMC (ATCC no. CRL-1999), the glioma line Cla (established from a grade 4 neuroblastoma by Dr Len Erickson, Indiana University), the T lymphoma line Jurkat (ATCC no. TIB152), the B lymphoma line Raji (ATCC no. CCL86) and the monocyte line U-937 (ATCC no. CRL 1593).

Isolation of λ phage cDNA clones

About 2.5×10^5 recombinant λ phages from each cDNA library were plated and screened with ^{32}P-nick-translated DNA probes as described (17). A 1.5 kb *Hin*dIII–*Not*I fragment from EST clone R13609, which was obtained from the IMAGE Consortium through Genome Systems Inc (St Louis, MO), was used to isolate clone 3.6 from an amplified oligo(dT)-primed human HeLa cDNA library in the excisable phage vector λYES (18). A 1.4 kb *Xho*I–*Sal*I fragment

*To whom correspondence should be addressed. Tel.: +1 708 216 5816; Fax: +1 708 216 9574; Email: hjaeck@wpo.it.luc.edu

[+]Present address: Department of Pathology, University of Uppsala, Uppsala, Sweden

from clone 3.6 was used to isolate clone 5.5 from an amplified random hexamer/oligo(dT)-primed human Jurkat cDNA library in the excisable phage vector λZapII (Stratagene, La Jolla, CA).

DNA sequencing

Double-stranded nucleotide sequencing was performed on two identical inserts by the Sanger dideoxy chain termination method (19) and primer walking (20). Percent nucleic acid and amino acid identities and similarities were determined using the BLAST program (21).

RNA analyses

Total RNA was prepared with the RNeasy Total RNA Isolation kit from Qiagen Inc. (Chatsworth, CA) and 10 µg RNA was analyzed by Northern blotting as described (22). Bands were detected by autoradiography.

Antibodies, antiserum production and purification

An anti-human HUPF1 peptide serum was generated by immunizing a rabbit with a synthetic human HUPF1 peptide (codons 106–123 in Fig. 1A) coupled to keyhole limpet hemocyanin. Anti-human HUPF1 peptide antibodies were purified on a peptide affinity column. The affinity-purified antibodies were used at a 1:200 dilution on Western blots and at a 1:25 dilution in immunofluorescence analysis. The generation of rabbit anti-mouse BiP antibodies was as previously described (23). Anti-proliferating cell nuclear antigen (PCNA) antibody was purchased from Novocastra Laboratories (Burlingame, CA), horseradish peroxidase (HRP)-conjugated goat anti-rabbit IgG from BioRad (Richmond, CA), HRP-conjugated goat anti-mouse IgG from Southern Biotechnolgy (Birmingham, AL) and fluorescein (FITC)-conjugated goat anti-rabbit (IgG) antibodies from Gibco-BRL (Gaithersburg, MD).

Isolation of nuclei from cultured mammalian cells

Nuclei were prepared according to the citric acid/Triton X-100 method (24). Cells were collected at 150 *g* for 5 min and washed twice in ice-cold PBS. Cells were resuspended in 5 ml ice-cold 25 mM citric acid, 1% Triton X-100, allowed to swell on ice for 5 min and subsequently homogenized (10 strokes in a type-B dounce homogenizer) on ice. The homogenate was layered on top of a 0.88 M sucrose, 25 mM citric acid cushion. Nuclei were pelleted at 800 *g* for 5 min at 3°C, resuspended in 5 ml ice-cold 25 mM citric acid, 0.25 M sucrose and pelleted again through a 0.88 M sucrose, 25 mM citric acid cushion. Nuclei were then resuspended in 25% glycerol, 5 mM magnesium acetate, 0.1 mM EDTA and 50 mM Tris, pH 8.0. The integrity of nuclei and the absence of cytoplasmic tags were confirmed by phase contrast microscopy. Prior to subjecting the SDS–PAGE, nuclei were lysed in 3× SDS sample buffer (0.2 M Tris, pH 6.8, 30% glycerol, 15% β-mercaptoethanol, 0.006% bromophenol blue and 7.5% SDS) and boiled for 3 min. To prevent protein degradation, all buffers contained 1 µM leupeptin, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 µM pepstatin, 100 µM EDTA and 0.072 µg/µl aprotinin.

Immunofluorescence analysis and Western blotting

Cytoplasmic immunofluorescence was performed as described (25). Proteins were separated on a 10% SDS–polyacrylamide gel and analyzed by Western blotting as described (15), with the exception that cells were lysed in the presence of the protease inhibitors leupeptin (1 µM), PMSF (1 mM), pepstatin (1 µM), EDTA (100 µM) and aprotinin (0.072 µg/µl) and protein–antibody complexes were visualized with HRP-conjugated goat anti-rabbit IgG antibodies (1:10,000; BioRad, Hercules, CA) and an Enhanced Chemiluminescence kit from Boehringer Mannheim (Indianapolis, IN).

In vitro translation

A 3.5 kb *Xho*I fragment containing the complete coding region of human *HUPF1* was isolated from clone 3.6 and cloned into the *Xho*I site of plasmid pGEM-7Zf(–) (Promega Corp., Madison, WI). The linearized plasmid was subjected to an *in vitro* coupled transcription–translation reaction in the presence of Tran^{35}S label (ICN Biochemicals, Costa Mesa, CA) using the TnT® Coupled Reticulocyte Lysate System from Promega (Madison, WI). Translation products were separated on a reducing 10% SDS–polyacrylamide gel and detected by fluorography.

RESULTS

Detection and isolation of a human *UPF1*-like cDNA clone by XREFdb searches and cDNA library screening

To clone a mammalian *UPF1* homolog, we used a comparative genomics approach, provided by XREFdb (26) through the World Wide Web site http://www.ncbi.nlm.nih.gov/xrefdb. When we searched the human, mouse and rat database subsets of expressed sequence tags (ETS) of random cDNA clones (dbEST; 27) with the complete UPF1 amino acid sequence, XREFdb identified a 1.5 kb EST cDNA fragment with sequence homology to *UPF1*. The 5′ 'single pass' 396 nt sequence of the 1.5 kb cDNA fragment (GenBank accession no. R13609) showed 63% identity and 71% similarity at the amino acid level to that of the RNA helicase region of yeast UPF1 (data not shown). Using the 1.5 kb R13609 fragment, we isolated a λ phage containing a 3.6 kb *Eco*RI insert (clone 3.6) from a human HeLa cDNA library and determined the complete nucleic acid sequence of its insert (Fig. 1A). Clone 3.6 contains from nt 176–3529 a 3354 bp open reading frame (ORF). The first N-terminal ATG codon of clone 3.6 from nt 176–178 is very likely the true translation initiation codon, because it is flanked by sequences that predict strong translational initiation (28) and preceded by an in-frame TAG stop codon at nt 107–109. The ORF encodes a polypeptide of 1118 amino acids with a calculated molecular mass of 123 kDa and a calculated pI of 6.06. When we analyzed *in vitro* translated products from a template of clone 3.6 by SDS–PAGE, we detected a single band with an apparent molecular mass of ~130 kDa (Fig. 2). This suggests that clone 3.6 encodes a 130 kDa polypeptide. The difference between predicted (123 kDa) and apparent (130 kDa) molecular weight might be due to the fact that acidic polypeptides run aberrantly in SDS–polyacrylamide gels (23). Finally, we found at the 3′-end of clone 3.6 one classical polyadenylation signal (AATAAA) from nt 3592–3597. However, residual adenine residues could not be detected at the 3′-terminal end, suggesting that clone 3.6 lacks part of the 3′-untranslated region.

The human cDNA encodes a structural homolog of yeast UPF1, a group I RNA helicase that is required to degrade nonsense mRNA

When we searched the translated non-redundant GenBank, PDB, SwissProt and PIR sequence databases (release date July 22,

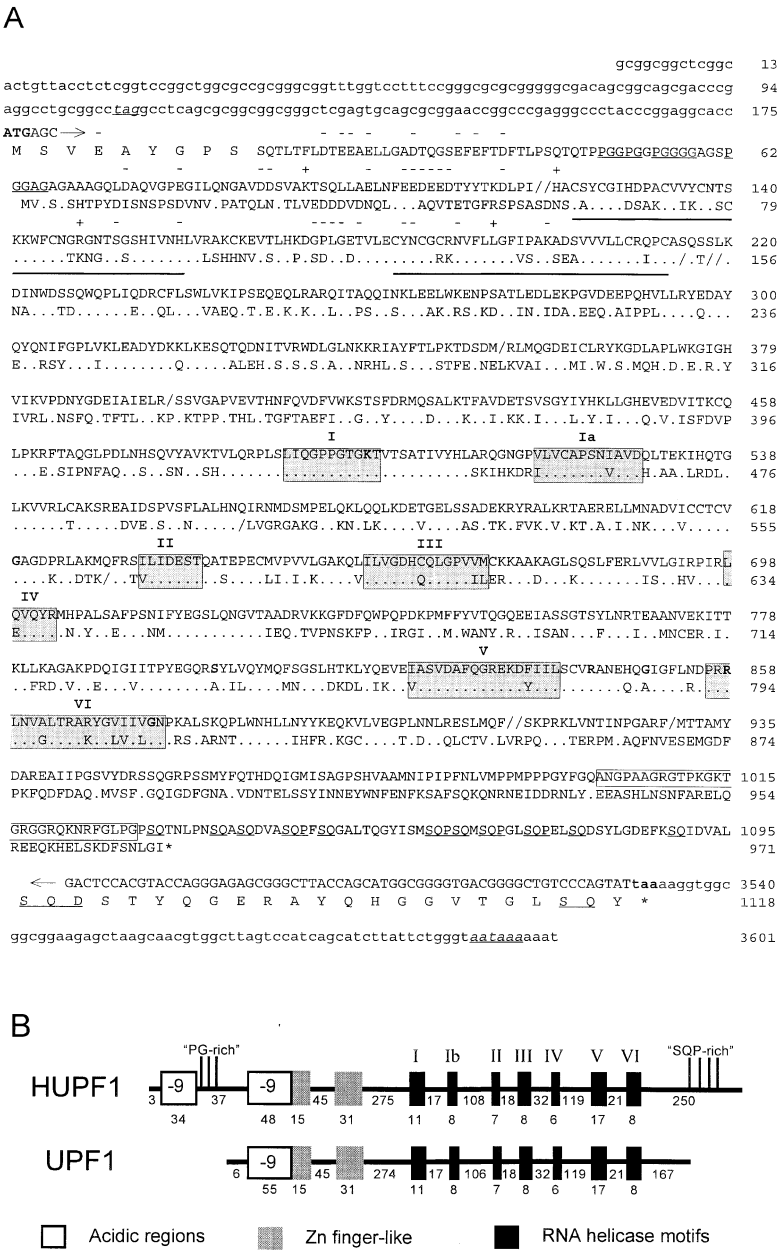

Figure 1. Sequence comparison of human HUPF1 and yeast UPF1. (**A**) Partial complementary nucleotide and entire deduced amino acid sequences of human HUPF1 and comparison with the *S.cerevisiae* UPF1 amino acid sequence (SwissProt accession no. P30771). The numbers on the right indicate nucleotide and amino acid positions. Untranslated regions are shown in lower case and translated regions in upper case letters. Arrows indicate the region of the *HUPF1* nucleotide sequence that was omitted from the figure and can be retrieved from GenBank sequence file U59323. The start and stop codons of *HUPF1*, as well as the amino acid residues critical for the function of yeast UPF1 (10,32), are in bold and the in-frame stop codon upstream of the ATG as well as the poly(A) addition signal site are underlined and italicized. The glycine/proline-rich tandemly repeated motifs and the serine/glutamine (SQ) and serine/glutamine/proline (SQP) repeats are underlined. The deduced amino acid sequences of HUPF1 and UPF1 (in single letter IUPC code) are shown below the partial nucleotide sequence. In the alignments (top, HUPF1; bottom, UPF1), a period indicates an identical amino acid residue and a forward slash a gap of one amino acid introduced for optimal alignment. Charged amino acids at the N-terminal ends are indicated above the HUPF1 and below the UPF1 sequence by + and – and the positively charged short amino acid stretch in HUPF1 is boxed. Two putative zinc finger motifs are underlined in bold and the seven (I, Ib and II–VI) group I RNA helicase motifs (29,30) are indicated by gray shaded boxes. (**B**) Domain maps of human HUPF1 and yeast UPF1. The domains are indicated by boxes. Numbers between and below the domains and at the ends indicate amino acid distances and numbers in the boxes the net charge of the domain. PG, proline/glycine-rich; SQP, serine/glutamine and serine/glutamine/proline repeats.

1996) with the deduced amino acid sequence of clone 3.6 using the BLAST program (21), we found that the human polypeptide showed the highest degree of homology to the UPF1 (NAM7) protein of *S.cerevisiae* (10,14) and a UPF1-like hypothetical 105.6 kDa protein whose ORF was identified in a cosmid clone of *Schizosaccharomyces pombe* (SwissProt accession no. W02865). Based on the complete amino acid sequence of the shorter yeast UPF1 protein, the overall identity between the human and the yeast UPF1 proteins is 51%. The human polypeptide and UPF1, as well as the UPF1-like hypothetical 105.6 kDa protein from

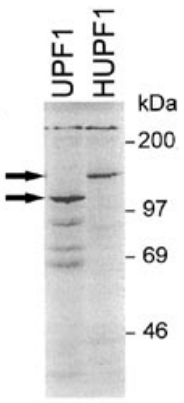

Figure 2. Analysis of radiolabeled *in vitro* translation products. 10 µl of ^{35}S-labeled *in vitro* translation products generated from a human *HUPF1* and a yeast *UPF1* template were separated on a reducing 10% SDS–polyacrylamide gel and detected by fluorography. Positions of protein molecular weight standards in kilodaltons (kDa) are indicated on the right and positions of the 130 kDa HUPF1 and the 109 kDa UPF1 band are indicated by arrows.

S.pombe (which we abbreviate to UPF1-like *S.pombe* protein), each contain putative zinc finger motifs at their N-terminal ends and central RNA helicase regions (Figs 1 and 3). The human coding sequence from codon 121 to codon 918, which contains the zinc finger and RNA helicase regions and represents 71.3% of its ORF, is 60% identical and 74% similar to the corresponding *UPF1* (*NAM7*) coding region from codon 60 to codon 856.

The RNA helicase region of the human polypeptide from codon 489 to codon 875 is 63% identical and 85% similar to that of UPF1 (codon 427 to codon 811) and contains in the same order all seven RNA helicase motifs that are found in UPF1 (10,14,29) and other group I RNA helicases (29–31). Remarkably, the distances between six of the seven RNA helicase consensus motifs are preserved between the human polypeptide and yeast UPF1, and the distances between domains Ib and II differ only by two amino acid residues (Fig. 1B). Additionally, all UPF1 amino acid residues that are known to be critical for UPF1 function (10,32) are preserved in both the human polypeptide and UPF1 (Fig. 1A, amino acids of interest are indicated in bold). Finally, the BLAST search revealed that all seven RNA helicase consensus motifs in the human polypeptide are very similar to that found in other putative eukaryotic group I RNA helicases (29), such as the UPF1-like hypothetical *S.pombe* protein, SEN1 of *S.cerevisiae*, mouse MOV-10, the human and mouse µ switch region binding proteins SMBP-2, a putative human DNA helicase and a hamster insulin enhancer binding protein (Fig. 3A). Thus,

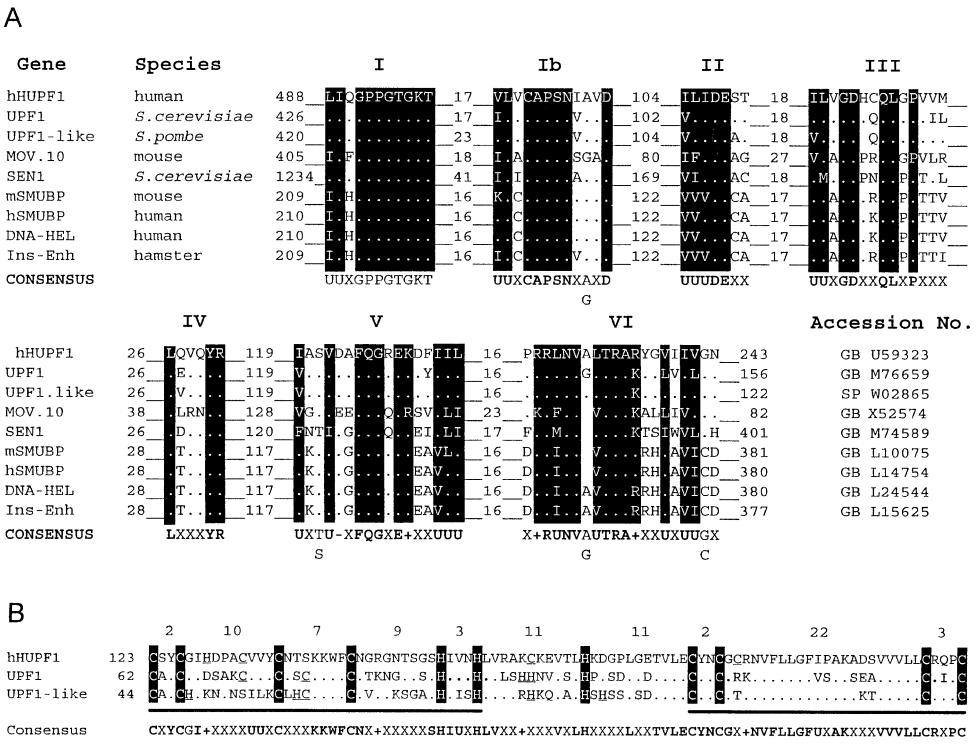

Figure 3. Alignments of consensus motifs shared between human HUPF1 and other putative group I RNA helicases. (**A**) Alignment of the seven helicase motifs I, Ia and II–VI shared between HUPF1 and all eukaryotic members of the RNA helicase superfamily I (29–31). Amino acid sequences of the seven motifs were retrieved from databases and aligned manually with the corresponding amino acid sequence of human HUPF1 (hHUPF1). In the alignments a period indicates amino acid identity and numbers show the distances between the domains and the lengths of the N- and C-terminal ends. Identical or similar amino acid residues are indicated in white on black. The consensus sequences were deduced from the nine alignments and are similar to that described by E.Koonin (29). In the consensus sequence U indicates a bulky hydrophobic (I, L, V, M, F, Y or W; 29), X any, + a basic (H, K or R) and – an acidic (D or E) amino acid residue; conserved residues are indicated in bold. The GenBank (GB) and SwissProt (SP) accession numbers are listed after each sequence. It should be noted that UPF1 is the only member of superfamily I that has been shown to be a *bona fide* RNA helicase (32). (**B**) Alignment of amino acid sequences of the cysteine/histidine-rich regions of HUPF1, UPF1 and the UPF-like *S.pombe* protein. In the alignments a period indicates amino acid identity and the numbers show the position of the first amino acid in each sequence and the distances between conserved C and H residues. Cysteine and histidine residues conserved in all three sequences are shown in white on black. Other C and H residues are underlined and two potential zinc finger motifs are underlined in bold.

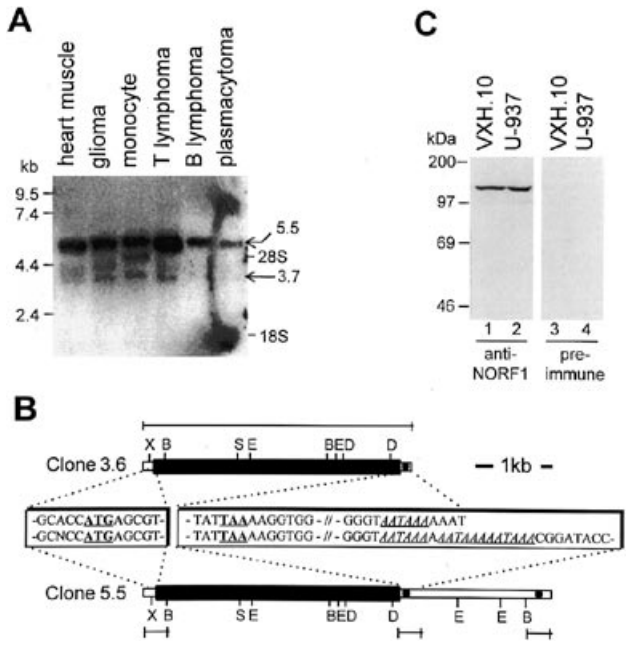

Figure 4. Analysis of *HUPF1* expression in cell lines. (**A**) Northern blot analysis of *HUPF1* mRNA in human cell lines. The blot was hybridized with the complete 3.6 kb human *HUPF1* probe (B). RNA size markers are indicated in kilobases (kb) on the left and the positions of the 18S and 28S RNA and the two transcripts that hybridize with the human *HUPF1* probe on the right. The band between the 5.5 and 3.6 kb transcripts is due to cross-hybridization of the *HUPF1* probe with 28S RNA. (**B**) Restriction maps of the human *HUPF1* clones 3.6 and 5.5. ORFs are indicated by black boxes, untranslated regions by white boxes and polyadenylation signal sites by black circles. Sequenced regions are indicated by brackets and partial sequences of clone 3.6 (upper sequence) and clone 5.5 (lower sequence) are shown in boxes. Restriction cutting sites are: X, *Xho*I; B, *Bst*XI; S, *Sal*I; E, *Eco*47III; D, *Dra*I. (**C**) Western blot analysis of cellular extracts from mouse VXH (lanes 1 and 3) and human U-937 cells (lanes 2 and 4) with rabbit anti-HUPF1 peptide antibodies (lanes 1 and 2) and preimmune serum (lanes 3 and 4). Molecular masses of protein standards are indicated in kDa on the left.

we conclude that the human 3.6 kb cDNA clone encodes a putative group I RNA helicase.

In addition to the seven RNA helicase motifs, which are present in all group I RNA helicases, the human polypeptide, *S.cerevisiae* UPF1 and the UPF1-like *S.pombe* protein contain a cysteine/histidine-rich region with putative zinc finger motifs (see Discussion) that is absent in other group I RNA helicase family members. The cysteine/histidine-rich region in the human polypeptide from codons 123 to 213 (Fig. 1A) is 63% identical and 90% similar to that of UPF1 (residues 62–152), and the distances between the cysteine/histidine-rich and RNA helicase regions of UPF1 and the human protein differ by only one amino acid residue (Fig. 1B). The cysteine/histidine-rich region is preceded in both polypeptides by an acidic stretch of 47 and 55 amino acids, respectively, with limited homology between each other and a net charge of –10. A short N-terminal acidic amino acid stretch with a net charge of –5 is also found in the UPF1-like *S.pombe* protein (not shown).

Further, compared with UPF1, HUPF1 has 63 and 83 additional amino acids at its N- and C-terminal ends, respectively. The N-terminal end contains an additional acidic stretch of 34 amino acids with a net charge of –9, followed by three tandem repeats of a proline/glycine-rich motif (PGGXG). The C-terminal tail contains several SQ and SQP repeats, for which a structural or functional role has not yet been determined.

In summary, we have cloned a human gene that encodes a putative group I RNA helicase and the first structural UPF1 homolog identified in mammalian cells. Hence, we suggest naming the protein human HUPF1 (for human *h*omolog of *UPF1*).

Chromosomal location of the human and mouse *HUPF1* genes

When we performed an XREFdb search with the complete amino acid sequence of yeast UPF1 and HUPF1, another human EST cDNA probe (GenBank accession no. F06433), whose deduced amino acid sequence showed 98 and 54% identity to the RNA helicase regions of human HUPF1 and UPF1, respectively, was identified (data not shown). This XREFdb search also revealed that the F06433 probe was selected by the XREF project to localize the mouse homolog of *UPF1* (mouse *HUPF1*) to chromosome 8 (XREFdb mouse map report for D8Xrf83, unpublished results). This inferred that the human homolog is located on chromosome 19. Indeed, the same probe (F06433) hybridized to a somatic hybrid cell line containing only human chromosome 19 (XREFdb human map report for D8Xrf83, unpublished results). Finally, the XREFdb search using the human HUPF1 amino acid sequence revealed two overlapping human EST clones (GenBank accession nos H13969, H13971), both of which have been isolated by exon trapping from a cosmid clone containing part of human chromosome 19 (19p12–p13.1; 33). The deduced amino acid sequences of the two clones (195 and 187 amino acid residues, respectively), differed from that of the corresponding region in human HUPF1 by only one and two amino acids, respectively (data not shown). These findings and the results from unpublished XREFdb mapping reports strongly suggest that the human *HUPF1* gene is located on chromosome 19.

Expression of *HUPF1*

If HUPF1 is required to reduce levels of nonsense mRNAs, we would expect that the *HUPF1* gene is expressed in every tissue, because any cell can potentially generate nonsense mRNA, for example via imprecise pre-mRNA splicing. When we used the human 3.6 kb *HUPF1* probe (clone 3.6 in Fig. 4B) to perform a Northern blot analysis of total RNA from several human cell lines representing various human tissues, we detected a predominant ~5.5 kb and a minor ~3.7 kb transcript in all analyzed samples (Fig. 4A), suggesting that *HUPF1* is indeed expressed in many tissues.

To isolate a *HUPF1* cDNA that represents the 5.5 kb transcript, we screened a human Jurkat λ phage cDNA library with a 5′ probe of clone 3.6 (*Xho*I–*Sal*I fragment in Fig. 4B). We identified one clone with a ~5.5 kb *Eco*RI insert (clone 5.5), whose size corresponds very well with that of the 5.5 kb transcript in Figure 4A. Restriction mapping and partial DNA sequencing of clone 5.5 revealed that clone 3.6 overlaps entirely with clone 5.5. For example, we detected in clone 5.5 all restriction sites that are present in clone 3.6 (Fig. 4B and data not shown). Moreover, we verified by partial DNA sequencing in clone 5.5 sequences of clone 3.6 extending from nucleotide positions 1 to 190 (Figs 1A and 4B, contains the translational start codon) and from nucleotide positions 3514 to 3601 (Figs 1A and 4B, contains the translational stop codon and the polyadenylation site). We also detected at the 3′-end of clone 5.5 an additional classical

polyadenylation signal site (AATAAA), however, a poly(A) tail could not be detected (data not shown). Therefore, clones 3.6 and 5.5 contain the same ORF, but clone 5.5 has a longer 3′-untranslated region with an extra polyadenylation site at its 3′-end. Thus, we conclude that clone 5.5 represents a partial clone of the 5.5 kb transcript, whereas clone 3.6 represents either a partial clone of the 5.5 kb transcript or a full-length clone of the 3.7 kb transcript. However, we cannot exclude the possibility that the 3.7 kb transcript is encoded by another RNA helicase gene with considerable sequence homology to the human *HUPF1* sequence.

Western blot analysis of human cell extracts with anti-HUPF1 peptide antibodies revealed a single band with an apparent molecular mass of 130 kDa (Fig. 4C, lane 2). The size of this band corresponds with the 130 kDa band that was detected in products translated *in vitro* from a template of clone 3.6 (Fig. 2). Because we also detected a 130 kDa band in mouse cells (Fig. 4C, lane 1), the 130 kDa polypeptide is, at least in part, conserved between mouse and man. Thus, we conclude that both the human and mouse *HUPF1* genes encode a single 130 kDa polypeptide.

Intracellular location of HUPF1 in human and mouse cells

UPF1, the yeast homolog of human HUPF1, is located in the cytoplasm but not in the nucleus (34). Immunofluorescence analysis with anti-HUPF1 peptide antibodies revealed a similar expression pattern of the HUPF1 protein in ethanol-fixed mouse VXH cells (Fig. 5A) as well as human Raji and U937 cells (data not shown), i.e. we detected a typical cytoplasmic staining, but little or no nuclear staining. The dim nuclear staining in some cells is very likely due to non-specific staining or staining that resulted from cytoplasm squashed on top of the nuclei during the mounting procedure, because a similar nuclear staining pattern was detected with antibodies against immunoglobulins (data not shown). The absence of nuclear HUPF1 in mouse VXH cells was confirmed by Western blot analysis of purified nuclei using the anti-HUPF1 peptide antibodies (Fig. 5B, lane 2). However, HUPF1 could, as expected, be easily detected in post-nuclear extracts of these cells (lane 1). The same results were obtained when the human monocytic cell line U-937 was examined by Western blotting (data not shown). To isolate nuclei from cultured cell lines, we used the citric acid/Triton X-100 method that yields intact nuclei lacking the outer nuclear membrane with attached endoplasmic reticulum and polysomes. That the nuclei prepared by this method were intact and free of contaminating endoplasmic reticulum was confirmed by the presence of PCNA protein (Fig. 5B, lane 2) and the absence of the 72 kDa BiP protein (35), an endoplasmic reticulum-resident chaperone (Fig. 5B, lane 2), respectively.

DISCUSSION

In summary, we have cloned a putative human RNA helicase that represents the first mammalian structural homolog of yeast UPF1, a protein that is required for accelerated degradation of nonsense mRNA. These conclusions are based on three observations. First, HUPF1 shows the highest degree of sequence homology at the amino acid level to the *bona fide* yeast RNA helicase UPF1. Second, both proteins exhibit a very similar overall polypeptide domain structure, i.e. each contains several putative zinc finger motifs and seven motifs characteristic of group I RNA helicases. Putative zinc finger and RNA helicase consensus motifs have so far only been found together on the same polypeptide chain in the group I RNA helicase UPF1 (NAM7) of *S.cerevisiae* and the UPF1-like protein of *S.pombe*, as well as the group II DEAD box RNA helicase GLH-1 of *C.elegans* (36). Finally, HUPF1, like UPF1, is expressed in the cytoplasm, but not within the nucleus.

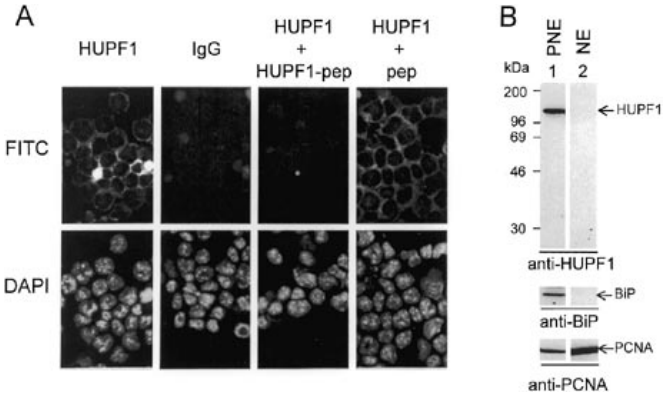

Figure 5. Intracellular localization of HUPF1 in mouse cells by immunofluorescence staining and Western blot analysis. (**A**) Ethanol-fixed mouse VXH cells were stained with either isotype-matched control antibodies (IgG) or anti-HUPF1 peptide antibodies (HUPF1), followed by incubation with FITC-conjugated goat anti-rabbit IgG antibodies and DAPI. Specificity of the anti-HUPF1 staining was confirmed by incubating cells with anti-HUPF1 peptide antibodies and 10 µg uncoupled HUPF1 peptide (HUPF-pep). No fluorescence was detected. In contrast, an irrelevant peptide (pep) did not prevent binding of anti-HUPF1 antibodies. The same sections were photographed to detect either fluorescein (FITC) or DAPI staining. The affinity-purified isotype-matched control antibody (IgG) was raised against a synthetic UPF1 peptide (Fig. 1A, codons 920–935). (**B**) Western blot analysis of post-nuclear (PE) and nuclear (N) extracts from mouse VXH cells. The same blot was stepwise incubated with anti-hHUPF1 peptide antibodies (HUPF1, 1:200), with anti-PCNA antibodies (anti-PCNA) and finally with anti-mouse BiP (anti-BiP, 1:200) antibodies.

Evidence that HUPF1 is not only a structural but very likely also a functional homolog of the nonsense mRNA-reducing yeast UPF1 protein comes from a study published during the review of this manuscript (37). In this study, Perlick and colleagues reported the cloning and sequencing of the *Rent1* (for *re*gulator of *n*onsense *t*ranscripts) gene, which is identical to our *HUPF1* gene. Using a modified allosupression growth assay these authors also showed that a chimeric yeast UPF1/human RENT1 protein restored the $UPF1^+$ phenotype in UPF1-deficient yeast.

Structural and functional domains shared between yeast UPF1 and human HUPF1

RNA helicase motifs. UPF1 and HUPF1 contain seven conserved motifs that are characteristic of RNA helicases (Figs 1 and 3A). RNA helicases are enzymes that recognize and unwind double-stranded RNA regions. These enzymes are involved in gene transcription and recombination, DNA replication and repair, RNA processing, transport and stability, as well as protein translation (reviewed in 38). To facilitate the displacement of paired RNA strands, RNA helicases must contain an ATPase activity and single-stranded (ss) RNA binding sites (discussed in 39). Additionally, extra RNA binding sites might be required to confer substrate specificity to the RNA helicase. Some of these activities, such as ATPase, RNA binding and helicase activities, were previously identified in affinity-purified UPF1 (32). The RNA helicase consensus motifs I and II, which are also commonly

found as A box and B box elements in NTPases (40), are critical for ATP hydrolysis and helicase activity of UPF1 (32), as well as of other RNA helicases (41). The helicase motif VI, which is critical for *in vivo* function of UPF1 (10), might be a candidate for a ss RNA binding site because mutations in this motif abolished ATPase, helicase and RNA binding activities of yeast UPF1 (42).

Potential zinc finger motifs. Other nucleic acid binding sites might be formed by putative zinc finger motifs that can be located in the cysteine/histidine-rich region of HUPF1, UPF1 and the UPF1-like *S.pombe* protein (Fig. 3B). Putative zinc finger motifs can be partially aligned with the zinc finger/knuckle motif, $CX_{2–5}CX_{4–12}C/HX_{2–4}C/H$, found in some transcription factors as well as RNA binding proteins (43). For example, two potential zinc finger motifs within the cysteine/histidine-rich region of HUPF1 (starting at codon 123 and codon 183, respectively), as well as of UPF1 (starting at codon 62 and codon 122, respectively) and the UPF1-like *S.pombe* protein (starting at codon 44 and codon 104, respectively), can be written as $CX_2CX_{28}HX_3H$ (C indicates a cysteine, H a histidine and X any amino acid residue) and $CX_2CX_{22}CX_3C$ (Fig. 3B) respectively. The distances between these two motifs are preserved in all three polypeptides (Fig. 3B).

The first zinc finger motif ($CX_2CX_{28}HX_3H$), which was not recognized in UPF1 (NAM7) by other researchers (10,14), resembles the motif of a CC/HH zinc finger protein (reviewed in 44), such as the 5S RNA binding proteins TFIIIA (45,46) and p43 (47). Two additional cysteine residues in this motif are conserved in HUPF1, UPF1 and the UPF1-like *S.pombe* protein (Fig. 3B), indicating that these residues might also be critical for protein function.

The second zinc finger motif ($CX_2CX_{25}CX_3C$), which was previously recognized in UPF1 (NAM7) (10,14), approximates a motif characteristic of CC/CC zinc finger proteins, such as glucocorticoid receptors (48) and the eukaryotic translational initiation factor eIF-2β (49). In eIF-2β, the zinc finger motif $CX_2CX_{18}CX_2C$ is required to recognize the AUG start codon during the ribosome-mediated translational scanning process. Therefore, it is tempting to speculate that the two zinc finger motifs of HUPF1 and UPF1 interact with ds or ss RNA regions and assist in the recognition of nonsense codons. For example, the zinc fingers might facilitate interaction between UPF1 or HUPF1 and ribosomes. Once a nonsense codon is recognized by the translocating ribosome, a conformational change of the zinc fingers might activate the helicase domain, which results in unwinding of ds RNA regions and, ultimately, in accelerated decapping and degradation of the nonsense transcript (50). Alternatively, the zinc fingers might facilitate recognition of an mRNA instability sequence located downstream of the nonsense codon (reviewed in 8), which also leads to unwinding of ds RNA regions, decapping and degradation.

A third potential zinc finger motif ($CX_2CX_6CX_3C$, starting at codon 62), which fits the zinc finger/knuckle consensus motif $CX_{2–5}CX_{4–12}C/HX_{2–4}C/H$ very well, was previously recognized in UPF1 (10,14). The same motif is present in HUPF1 (starting at codon 123), but absent from the UPF1-like *S.pombe* protein and, therefore, might not be critical for function (Fig. 3B)

Other motifs. Both yeast UPF1 and human HUPF1 contain at their N-terminal ends short acidic amino acid stretches without much sequence homology (Fig. 1A). Acidic amino acid stretches of limited homology are also found in many transcription factors and might be involved in homodimer formation (discussed in 14 and reviewed in 51). Similarly, the acidic stretches in UPF1 and HUPF1 could facilitate formation of a multicomponent nonsense mRNA-reducing complex. It has recently been shown that UPF1 interacts with UPF2 (11,12), another UPF factor required for nonsense codon-mediated degradation of mRNA in yeast (10).

Motifs unique to HUPF1

HUPF1 contains two additional putative structural and functional motifs that are absent in UPF1. First, three tandemly repeated glycine-rich motifs (PGGXG, Fig. 1A) are located in the N-terminal tail. Glycine-rich regions have been proposed to facilitate protein–protein (52–54) or protein–RNA interactions (reviewed in 43). Second, a short positively charged 27 amino acid residue stretch (Fig. 1A, codons 1003–1029), which contains seven basic (arginine or lysine) and 11 small amino acid residues (glycine or alanine), can be arranged into one conserved and two degenerated RGG motifs. Such motifs are found in some RNA binding domains (reviewed in 43).

Subcellular localization of nonsense mRNA reduction

There is strong experimental evidence that UPF1-dependent reduction of nonsense mRNA occurs in the cytoplasm of yeast cells (reviewed in 8). For example, nonsense mRNA is associated with polysomes (55), a dominant-negative form of UPF2 interferes with the function of UPF1 in yeast cells when it is targeted to the cytoplasm, but not when it is targeted to the nucleus (12), and yeast UPF1 is located in the cytoplasm, co-purifies with polysomes and is not detected in the nucleus (34).

The finding that HUPF1 is expressed in the cytoplasm, but not within the nucleus of mammalian cells, suggests that HUPF1, like yeast UPF1, exerts its function in the cytoplasm. Therefore, HUPF1, in concert with other HUPF factors, might only induce degradation of nonsense mRNAs, such as nonsense Ig μ (Li and Jäck, in preparation) and adenine phosphoribosyl transferase (APRT) mRNAs (56), once these transcripts have reached the cytoplasm. However, our analysis does not exclude the possibility that HUPF1 is also associated with the outer nuclear membrane or the nuclear pore complex and that small amounts of HUPF1, which escaped our detection methods, could be transported into the nucleus. Therefore, HUPF1 might also induce translation-dependent reduction of nucleus-associated nonsense mRNAs, such as nonsense triosephosphate isomerase (TPI) mRNA (57,58), and interfere with the splicing of nonsense pre-mRNA (59,60,61; reviewed in 1). The intracellular localization of HUPF1 by immunoelectron microscopy, as well as the analysis of cytoplasmic and nuclear nonsense mRNA levels and turnover rates in mammalian cells that lack the HUPF1 gene, will distinguish between these possibilities.

Why is a nonsense mRNA surveillance mechanism preserved during evolution?

The isolation of a human gene whose deduced amino acid sequence has a striking homology to the nonsense mRNA-reducing UPF1 factor implies a critical role of nonsense mRNA-reducing factors in mammalian organisms. For example, HUPF proteins reduce cytoplasmic levels of nonsense mRNA that could compete with functional mRNA for RNA processing and regulation factors as well as for translating ribosomes. Additionally, reducing nonsense mRNA levels prevents the accumulation of high levels of truncated polypeptides that can interfere with the

function of the corresponding full-length polypeptide chain. For example, a shorter form of the myosin heavy chain, which is translated from stable nonsense mRNA, causes, even in the presence of its corresponding full-length chain, abberant development of muscle cells in *C.elegans* (6). Within the immune system, HUPF proteins might be critical to reduce levels of nonsense Ig and TCR transcripts in lymphocytes, which are predominantly encoded by non-productively rearranged Ig and TCR genes (3). Because non-productively rearranged Ig and TCR genes are often generated during somatic rearrangement of Ig and TCR gene segments, lymphocytes that contain on one chromosome a productive and on the other chromosome a non-productively rearranged Ig or TCR gene can easily be detected in lymphoid organs (reviewed in 3). If nonsense Ig and TCR transcripts were not removed from a lymphocyte, the accumulating shorter Ig or TCR polypeptides could compete with their corresponding functional chains for other polypeptides that are required for the assembly of signal transducing Ig or TCR complexes. Because signaling through Ig and TCR complexes is obligatory for either a lymphocyte precursor to develop into an antigen-responsive mature lymphocyte or for a mature lymphocyte to respond to foreign antigen (reviewed in 62), an intact nonsense mRNA surveillance system should be critical for B and T cell development. Thus, it will be interesting to see whether deleting the *HUPF1* gene in a mouse affects lymphocyte development.

ACKNOWLEDGEMENTS

We thank Laura Hartwell for help in preparing rabbit antisera, Drs G.Napolitano, C.Blobel, G.Weskamp, P.Tucker, G.Wu and S.Elledge for genomic and cDNA libraries, Drs T.Ellis, R.Pieper and R.Shankar for human cell lines, Dr M.Wabl for discussions and Drs I.Haas and S.Amero for critical reading of the manuscript. This work was supported in part by the National Institutes of Health (R29CA56772-01A1), the Cancer Society of America and the Tobacco Research Council of America.

REFERENCES

1 Maquat,L.E. (1995) *RNA*, **1**, 453–465.
2 Rowen,L., Koop,B.F. and Hood,L. (1996) *Science*, **272**, 1755–1762.
3 Alt,F.W., Blackwell,T.K. and Yancopoulos,G.D. (1987) *Science*, **238**, 1079–1087.
4 Kazazian,H.H,Jr, Dowling,C.E., Hurwitz,R.L., Coleman,M., Stopeck,E. and Adams,J.G. (1992) *Blood*, **79**, 3014–3018.
5 Thein,S.L. (1992) *Br. J. Haematol.*, **80**, 273–277.
6 Pulak,R. and Anderson,P. (1993) *Genes Dev.*, **7**, 1885–1897.
7 Higgins,C.F., Peltz,S.W. and Jacobson,A. (1992) *Curr. Opin. Genet. Dev.*, **2**, 793–747.
8 Peltz,S.W., He,F., Welch,E. and Jacobson,A. (1994) *Prog. Nucleic Acid Res. Mol. Biol.*, **47**, 271–298.
9 Ross,J. (1995) *Microbiol. Rev.*, **59**, 423–450.
10 Leeds,P., Wood,J.M., Lee,B.-S. and Culbertson,M. (1992) *Mol. Cell. Biol.*, **12**, 2165–2177.
11 Cui,Y., Hagan,K.W., Zhang,S. and Peltz,S.W. (1995) *Genes Dev.*, **9**, 423–436.
12 He,F. and Jacobson,A. (1995) *Genes Dev.*, **9**, 437–454.
13 Lee,B.-S. and Culbertson,M.R. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 10354–10358.
14 Altamura,N., Groudinsky,O., Dujardin,G. and Slonimski,P.P. (1992) *J. Mol. Biol.*, **224**, 575–587.
15 Keyna,U., Beck,E.G., Jongstra,J., Applequist,S.E. and Jäck,H.-M. (1995) *J. Immunol.*, **155**, 5536–5542.
16 Jäck,H.-M., Berg,J. and Wabl,M. (1989) *Eur. J. Immunol.*, **19**, 843–847.
17 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
18 Elledge,S.J., Mulligan,J.T., Ramer,S.W., Spottswood,M. and Davis,R.W. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1731–1735.
19 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
20 Slightom,J.L., Siemieniak,D.R., Sieu,L.C., Koop,B.F. and Hood,L. (1994) *Genomics*, **20**, 149–168.
21 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
22 Jäck,H.-M. and Wabl,M. (1988) *EMBO J.*, **7**, 1041–1046.
23 Bornemann,K., Brewer,J.W., Beck-Engeser,G.B., Corley,R.B., Haas,I.G. and Jäck,H.-M. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 4912–4916.
24 Birnie,G.D. (1986) In Macgillivray,A.J. and Birnie,G.D. (eds) *Nuclear Structures: Isolation and Characterization*, Butterworths, London, Boston, pp. 13–26.
25 Jäck,H.-M., Beck-Engeser,G.B., Sloan,B., Wong,H.L. and Wabl,M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11688–11691.
26 Bassett,D.J., Boguski,M.S., Spencer,F., Reeves,R., Goebl,M. and Hieter,P. (1995) *Trends Genet.*, **11**, 372–373.
27 Boguski,M.S., Lowe,T.M. and Tolstoshev,C.M. (1993) *Nature Genet.*, **4**, 332–333.
28 Kozak,M. (1994) *Biochimie*, **76**, 815–821.
29 Koonin,E.V. (1992) *Trends Biochem. Sci.*, **17**, 495–497.
30 Hodgman,T.C. (1988) *Nature*, **333**, 22–23.
31 Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1989) *Nucleic Acids Res.*, **17**, 4713–4730.
32 Czaplinski,K., Weng,Y., Hagan,K.W. and Peltz,S.W. (1995) *RNA*, **1**, 610–623.
33 Li,Q.-L., Lennon,G.G. and Brook, J.D. (1996) *Genomics*, **32**, 218–224.
34 Atkins,A.L., Altamura,N., Leeds,P. and Culbertson,M.R. (1995) *Mol. Biol. Cell*, **6**, 611–625.
35 Haas,I.G. and Wabl,M.R. (1983) *Nature*, **306**, 387–389.
36 Roussell,D.L. and Bennett,K.L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9300–9304.
37 Perlick,H.A., Medghalchi,S.M., Spencer,F.A., Kendzior,R.J. and Dietz,H.C. (1996) *Proc. Natl. Acad.Sci. USA*, **93**, 10928–10932.
38 Schmid,S.R. and Linder,P. (1992) *Mol. Microbiol.*, **6**, 283–291.
39 Gibson,T.J. and Thompson,J.D. (1994) *Nucleic Acids Res.*, **22**, 2552–2556.
40 Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.
41 Pause,A. and Sonenberg,N. (1992) *EMBO J.*, **11**, 2643–2654.
42 Weng,Y., Czaplinski,K. and Peltz,S.W. (1996) *Mol. Cell. Biol.*, **16**, 5477–5490.
43 Burd,C.G. and Dreyfuss,G. (1994) *Science*, **265**, 615–621.
44 Klug,A. and Rhodes,D. (1987) *Trends Biochem. Sci.*, **12**, 464–469.
45 Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
46 Theunissen,O., Rudt,F., Guddat,U., Mentzel,H. and Pieler,T. (1992) *Cell*, **71**, 679–690.
47 Joho,K.E., Darby,M.K., Crawford,E.T. and Brown,D.D. (1990) *Cell*, **61**, 293–300.
48 Vallee,B.L., Coleman,J.E. and Auld,D.S. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 999–1003.
49 Donahue,T.F., Cigan,A.M., Pabich,E.K. and Valavicius,B.C. (1987) *Cell*, **54**, 621–632.
50 Muhlrad,D. and Parker,R. (1994) *Nature*, **370**, 578–581.
51 Struhl,K. (1989) *Annu. Rev. Biochem.*, **58**, 1051–1077.
52 Steinert,P.M., Mack,J.W., Korge,B.P., Gan,S.-Q., Haynes,S.R. and Steven,A.C. (1990) *Int. J. Biol. Macromol.*, **13**, 130–139.
53 Siomi,H. and Dreyfuss,G. (1995) *J. Cell Biol.*, **129**, 551–560.
54 Weighardt,F., Biamonti,G. and Riva,S. (1995) *J. Cell Sci.*, **108**, 545–555.
55 He,F., Peltz,S.W., Donahue,J.L., Rosbash,M. and Jacobson,A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 7034–7038.
56 Kessler,O. and Chasin,L.A. (1996) *Mol. Cell. Biol.*, **16**, 4426–4435.
57 Cheng,J. and Maquat,L.E. (1993) *Mol. Cell. Biol.*, **13**, 1892–1902.
58 Belgrader,P., Cheng,J. and Maquat,L.E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 482–486
59 Aoufouchi,S., Yelamos,J. and Milstein,C. (1996) *Cell*, **85**, 415–422.
60 Naeger,L.K., Schoborg,R.V., Zhao,Q., Tullis,G.E. and Pintel,D. J. (1992) *Genes Dev.*, **6**, 1107.
61 Dietz,H.C., Vall,D., Francomano,C.A., Kendzior,R.J., Pyeritz,R.E.and Cutting,G.R. (1993) *Science*, **259**, 680–683.59.
62 Rajewsky,K. (1996) *Nature*, **381**, 751–758.