THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 26, Issue of June 25, pp. 27039–27049, 2004
Printed in U.S.A.

Diversity of Vertebrate Splicing Factor U2AF^{35}

IDENTIFICATION OF ALTERNATIVELY SPLICED *U2AF1* mRNAs*

Received for publication, February 26, 2004, and in revised form, April 16, 2004
Published, JBC Papers in Press, April 19, 2004, DOI 10.1074/jbc.M402136200

Teresa R. Pacheco‡§, Anita Q. Gomes‡¶, Nuno L. Barbosa-Morais‡‖, Vladimir Benes, Wilhelm Ansorge**, Matthew Wollerton‡‡, Christopher W. Smith‡‡, Juan Valcárcel**§§, and Maria Carmo-Fonseca‡¶¶**

*From the ‡Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, 1649-028 Lisbon, Portugal, **EMBL, D-69112 Heidelberg, Germany, and the ‡‡Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, United Kingdom*

U2 small nuclear ribonucleoprotein auxiliary factor small subunit (U2AF^{35}) is encoded by a conserved gene designated *U2AF1*. Here we provide evidence for the existence of alternative vertebrate transcripts encoding different U2AF^{35} isoforms. Three mRNA isoforms (termed U2AF^{35}a–c) were produced by alternative splicing of the human *U2AF1* gene. U2AF^{35}c contains a premature stop codon that targets the resulting mRNA to nonsense-mediated mRNA decay. U2AF^{35}b differs from the previously described U2AF^{35}a isoform in 7 amino acids located at the atypical RNA Recognition Motif involved in dimerization with U2AF^{65}. Biochemical experiments indicate that isoform U2AF^{35}b, which has been highly conserved from fish to man, maintains the ability to interact with U2AF^{65}, stimulates U2AF^{65} binding to a pre-mRNA, and promotes U2AF splicing activity *in vitro*. Real time, quantitative PCR analysis indicates that U2AF^{35}a is the most abundant isoform expressed in murine tissues, although the ratio between U2AF^{35}a and U2AF^{35}b varies from 10-fold in the brain to 20-fold in skeletal muscle. We propose that post-transcriptional regulation of *U2AF1* gene expression may provide a mechanism by which the relative cellular concentration and availability of U2AF^{35} protein isoforms are modulated, thus contributing to the finely tuned control of splicing events in different tissues.

In higher eukaryotes, most protein-coding genes contain sequences that are spliced from the nascent transcripts (pre-mRNAs) in the nucleus. Intron excision is carried out by an assembly of small nuclear RNAs and proteins that are collectively recruited to pre-mRNAs forming the spliceosome (reviewed in Ref. 1). Although introns are excised with a high degree of precision, for many pre-mRNAs there is flexibility in the choice of alternative splice sites, often in response to tissue-specific, physiologically or developmentally regulated states. Alternative splicing produces multiple mRNAs encoding distinct proteins, thus expanding the coding capacity of genes and contributing to the proteomic complexity of higher organisms (2–4). Moreover, alternative splicing may contribute to regulate protein expression by generating premature termination codons that target the transcript to nonsense-mediated mRNA decay (5).

In metazoans, pre-mRNA sequences implicated in splicing are only weakly conserved. Multiple, relatively weak protein-protein and protein-RNA interactions involving these sequences and additional regulatory sequence elements, which can positively or negatively affect spliceosome assembly at nearby splice sites, constitute the basis to control alternative splicing (for recent reviews see Refs. 4, 6, and 7). Proteins that bind to pre-mRNA and affect splicing regulation include SR proteins, hnRNPs,[1] and tissue- or developmental stage-specific factors. Individual SR proteins interact only weakly with enhancer elements, but their binding to pre-mRNA is highly cooperative and can lead to recognition of distinct RNA sequence motifs, thus contributing to the selection of regulated splice sites. A distinct role is played by hnRNP proteins, which in general antagonize the stimulatory activity of SR proteins, and cell type-specific proteins, which may either inhibit or promote splicing. Thus, control of alternative splicing is achieved through the combinatorial interplay of both regulatory sequence signals and trans-acting protein factors (4, 8, 9).

The spliceosome is a multicomponent RNA-protein machine containing five uracil-rich small nuclear ribonucleoproteins (U snRNPs) and many non-snRNP protein splicing factors. In the late 1990s ~100 splicing factors were identified (10), and since then the number has nearly doubled (1). Initiation of spliceosome recruitment to a pre-mRNA involves recognition of the 5′ splice site by the U1 snRNP, whereas the U2 snRNP associates with the 3′ region of the intron. The establishment of a stable interaction between U2 snRNP and pre-mRNA requires an auxiliary factor, U2AF (11). The U2

* This work was supported in part by Grant POCTI/MGI/36547/2000 from Fundação para a Ciência e Tecnologia, Portugal, and by Grant RG0300/2000-M from the Human Frontier Science Program Organization. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequences for chicken U2AF1 gene and cDNA (classes I–III) sequences and human U2AF^{35}b cDNA sequence have been deposited in the EMBL database (European Bioinformatics Institute, Cambridge, UK) under the accession numbers AJ291762, AJ291763, AJ291764, AJ291765, and AJ627978, respectively.

§ Supported by Fundação para a Ciência e Tecnologia Fellowship PRAXIS XXI/BD/18044/98.

¶ Supported by Fundação para a Ciência e Tecnologia Fellowship POCTI SFRH/BPD/9388/2002.

‖ Supported by Fundação para a Ciência e Tecnologia Fellowship PRAXIS XXI SFRH/BD/2914/2000.

§§ Present address: ICREA and Centre de Regulacio Genomica, 08003 Barcelona, Spain.

¶¶ To whom correspondence should be addressed: Institute of Molecular Medicine, Faculty of Medicine, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal. Tel.: 351-21-7940157; Fax: 351-21-7951780; E-mail: carmo.fonseca@fm.ul.pt.

[1] The abbreviations used are: hnRNP, heterogeneous nuclear; snRNP, small nuclear ribonucleoprotein; RRM, RNA recognition motif; NMD, nonsense-mediated mRNA decay; UTR, untranslated region; NTA, nitrilotriacetic acid; GST, glutathione *S*-transferase; siRNA, small interfering RNA; GFP, green fluorescent protein; AdML, adenovirus major late.

snRNP auxiliary factor (U2AF) consists of two subunits, U2AF^{65} and U2AF^{35}, that interact to form a stable heterodimer (12). U2AF^{65} binds directly to the polypyrimidine tract of pre-mRNA and is essential for splicing (13). U2AF^{35} recognizes the conserved 3′ splice site dinucleotide AG (14–16) and is required for splicing of a subset of primary transcripts, the so-called AG-dependent pre-mRNAs (17).

Human U2AF^{35} is 240 residues long, with an atypical RNA recognition motif (RRM) involved in dimerization with U2AF^{65} (18). Additionally, the protein contains a C-terminal arginine/serine (RS)-rich domain interrupted by glycines (19). The RS region of U2AF^{35} has been shown to establish protein-protein interactions with splicing factors of the SR family, and it was proposed that SR proteins bound to purine-rich exonic splicing enhancers facilitate recruitment of U2AF^{65} to the polypyrimidine tract via bridging interactions mediated by U2AF^{35} (20–22). Other results, however, argue that the RS domain-mediated interactions with SR proteins bound to the exonic splicing enhancers is dispensable for the function of U2AF^{35} in AG-dependent pre-mRNA splicing (23). Although interaction of U2AF^{35} with the 3′ splice site AG can stabilize U2AF^{65} binding, not all the activities of U2AF^{35} correlate with increased cross-linking of U2AF^{65} to the polypyrimidine tract, suggesting an additional, yet unknown function for U2AF^{35} in pre-mRNA splicing (23).

In this work we show that primary transcripts encoding U2AF^{35} splicing factor in higher vertebrates can be alternatively spliced and polyadenylated. Alternative splicing of U2AF^{35} pre-mRNAs may either introduce a premature stop codon that targets the resulting mRNA to nonsense-mediated mRNA decay (NMD) or alter 7 amino acid residues within the RRM domain. This alteration gives rise to a protein isoform that maintains the ability to bind U2AF^{65} and stimulates splicing activity *in vitro*. Most interestingly, the two U2AF^{35} protein isoforms have been under high selective pressure during evolution, suggesting that they play specific functions in vertebrate organisms. Our results further show that the relative abundance of mRNAs encoding each U2AF^{35} isoform differs between cell types, arguing that expression of the *U2AF1* gene is controlled in a tissue-specific manner.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of Chicken U2AF1 cDNA—Degenerate oligonucleotides were designed to regions of U2AF^{35} cDNA conserved among *Homo sapiens*, *Drosophila melanogaster*, and *Schizosaccharomyces pombe*. The forward PCR primer (5′-AAGATHGGIGCITGYCGICAYGC-3′) was designed from amino acids 23 to 30 of human U2AF^{35}. The reverse PCR primer (5′-TCRTAYTGICGRCAIGCYTC-3′) was designed to be complementary to the region encoding from amino acids 152 to 160. PCR was carried out with a pool of random chicken embryo cDNA molecules as template. A Biometra® UNO II thermocycler (Germany) was used, and the conditions for amplification were as follows: 94 °C for 5 min, followed by 40 cycles of denaturation at 94 °C for 30 s, annealing at 54 °C for 2 min, and extension at 72 °C for 1 min, with a time increment of 1 s per cycle, and finally 72 °C for 30 min. The PCR product was gel-purified using QIAEX II Gel Extraction Kit (Qiagen) and subcloned into pBluescript II KS (+/−) by "TA cloning" (24). This 392-bp chicken U2AF^{35} cDNA fragment (PCR35) was random prime-labeled with [α-^{32}P]dCTP using the Prime-it® II Random Primer Labeling kit (Stratagene®) and used to screen a chicken embryo cDNA library constructed in λZAP II (Stratagene®). Approximately 1.5×10^6 plaques were screened, and five positive plaques were isolated. cDNA inserts in pBluescript II KS phagemid were excised *in vivo* using the ExAssist/SOLR System (Stratagene®), purified, analyzed by restriction digest, and completely sequenced.

Cloning of Chicken U2AF1 Gene—The chicken U2AF^{35} cDNA fragment PCR35 was random prime labeled and used to screen a DT40 cell line genomic library constructed in LambdaGEM®-11 (Promega), using standard procedures (25). From 1.2×10^6 plaques, six positive clones were isolated. Phage DNAs were purified and analyzed by restriction endonuclease mapping and Southern blotting (26). Genomic fragments were subcloned into pBluescript II KS, and an 18-kb region covering the entire chicken *U2AF1* gene was sequenced.

Sequencing—Cycle sequencing of plasmid DNA was performed with the AmpliTaqFS Dye Primer Core kit (Applied Biosystems), using 1 μg of plasmid DNA and 3 pmol of each standard forward and reverse primers labeled with fluorescein isothiocyanate or CY5. An MJ Research (Waltham, MA) PT-200 cycler was used for 35 cycles (97 °C, 15 s; 55 °C, 30 s; 68 °C, 30 s). Reactions were loaded "off-gel" on 72-clone porous-membrane combs, applied to 60-cm long polyacrylamide gels (4.5% Hydrolink Long Ranger gel solution, FMC BioProducts), and analyzed on the ARAKIS sequencing system with array detectors, developed at EMBL (27). Raw sequencing data were evaluated and analyzed, and the consensus sequence was assembled by using the software packages (Lane Tracker and Gene Skipper) developed at EMBL. Remaining sequencing gaps were covered by primer walking (28).

Cell Culture—DT40 cells (ATCC CRL-2111) were grown in Dulbecco's modified Eagle's medium (Invitrogen) supplemented as recommended by the American Type Culture Collection. HeLa cells (ECACC 93021013) were grown in minimum essential medium with Earle's salts supplemented with 10% (v/v) fetal calf serum and 1% (v/v) nonessential amino acids (Invitrogen).

Northern Blotting—Total RNA was extracted from either DT40 cells or chicken embryos using the TRIzol® reagent (Invitrogen). Poly(A) RNA was isolated using oligo(dT)-cellulose (Amersham Biosciences), electrophoresed on an agarose-formaldehyde gel, and transferred to an Hybond-N+ membrane (Amersham Biosciences) (26). Hybridizations were carried out with the following probes: (*a*) chicken U2AF^{35} cDNA fragment PCR35; (*b*) a 1175-bp XbaI/XbaI (2403–3578) fragment obtained by restriction digest of pBluescript-chU2AF^{35}III (a class III clone); additionally, a probe complementary to chicken β-actin was used as an internal control. Probes were gel-purified and labeled with [α-^{32}P]dCTP by random priming. Pre-hybridization (30 min) and hybridization (1 h) were carried out at 68 °C in ExpressHyb Solution (Clontech). Following hybridization, washing was performed in 2× SSC, 0.05% SDS (three times 10 min at room temperature), and 0.1× SSC, 0.1% SDS (two times at 50 °C). Filters were exposed to a Kodak Biomax-MR film (Eastman Kodak Co.), with intensifying screens, at −70 °C. Autoradiograms were analyzed using the software package ONE-Dscan™, version 1.0 (Scanalytics, a Division of CSPI).

NMD Inhibition and Reverse Transcriptase-PCR—RNA interference of Upf1 was carried out as described previously (29). Briefly, for a typical siRNA transfection, 24-well dishes were seeded with 10^5 cells prior to transfection. The following day (day 2) 120 pmol of siRNA duplex was transfected using 3 μl of LipofectAMINE® 2000 (Invitrogen) as described previously (30). On day 3 cells were trypsinized and split to a well of a 6-well plate. On day 4 cells were retransfected with siRNA duplex as on day 2, and finally, cells were harvested on day 6. mRNA targets for gene specific knockdown were UPF1: AAGAUGCAGUUCCGCUCCAUU (31) and control C2: AAGGUCCGGCUCCCCCAAAUG (30). Gene silencing was monitored by immunoblotting. Total RNA was isolated with the TRIzol® reagent (Invitrogen) and treated with RQ1 RNase-free DNase (Promega). Reverse transcriptase reactions were primed with oligo(dT) by using Superscript II RT enzyme (Invitrogen). The resulting cDNA was amplified with primers BothFor, 5′-GCACAATAAACCGACGTTTAGCCAG-3′, and BothRev, 5′-TGGATCGGCTGTCCATTAAACCAAC-3′ (exons 2–4, Fig. 4). The amplification products were digested with HinfI, separated by gel electrophoresis, and detected by ethidium bromide staining.

Expression and Purification of Recombinant Proteins—Full-length human U2AF^{35}a cDNA was used to screen a Uni-ZAP®XR Human Fetal Spleen λ cDNA library (Stratagene®). Positive clones were analyzed by restriction digestion with HinfI and sequenced in order to isolate U2AF^{35}b cDNA clones. U2AF^{35}b cDNA was subsequently cloned into NcoI/KpnI sites of pFastBac™ HT plasmid, and recombinant baculoviruses were generated by using the Bac-to-Bac® Baculovirus Expression System (Invitrogen). His$_6$-tagged U2AF^{35}a and U2AF^{35}b proteins were purified from baculovirus-infected cells as described previously (32) and dialyzed against 100 mM KCl buffer D (20 mM HEPES (pH 8.0), 0.5 mM EDTA, 20% glycerol, 1 mM dithiothreitol, 0.05% Nonidet P-40). U2AF^{65} and U2AF^{65}Δ35 were expressed as glutathione *S*-transferase (GST) fusion proteins in *Escherichia coli* and purified as described previously (33). The plasmids used for protein expression were described previously (13, 34). The purified proteins were dialyzed against 100 mM KCl buffer D. Protein concentrations were estimated by comparing dilutions of the preparations to serial dilutions of a bovine serum albumin standard in SDS-containing denaturing gels.

Construction and Expression of GFP Fusion Proteins—Green fluorescent protein (GFP) fusion constructs were obtained by restriction

TABLE I
Sequence of primers and TaqMan® MGB probes used in the quantitative real time reverse transcriptase-PCR
FAM, 6-carboxyfluoresceine; MGB, Minor Groove Binder.

Gene	Primers and probes sequences	Concentration in reaction	Amplicon size
		nm	*bp*
U2AF^{35}a	Forward, 5′-TAGCCAGACCATTGCCTCTT-3′	900	
	Reverse, 5′-CGCAAACCGTCAGCAGACT3′	900	72
	Probe, FAM-5′-AACATTTACCGTAACCCT3′-MGB	150	
U2AF^{35}b	Forward, 5′-TCCCCAAAACAGTGCACAGA3′	900	
	Reverse, 5′-TCATCATAGTGCTCCTGCATCTC3′	900	78
	Probe, FAM-5′-GGCTCACACTGTGCTG3′-MGB	150	
U2AF^{65}	Forward, 5′-ACATCACCCCAATGCAGTACAA3′	300	
	Reverse, 5′-AGGGCAGTGGCTGGAATCT3′	300	61
	Probe, FAM-5′-ACCCGCAGCTTGCATG3′-MGB	200	

digestion and subcloning into the appropriate pEGFP-C vector (Clontech). GFP-U2AF^{35}a was described previously (35). GFP-U2AF^{35}b was constructed by digestion of pBS-U2AF^{35}b with BglII followed by fill-in and PstI digestion. The insert was then cloned into PstI/SmaI sites of pEGFP-C3. HeLa cells were transfected and analyzed by confocal microscopy as described previously (35).

In Vitro Binding Assays—GST pull-down assays were carried out using 1 μg of GST-U2AF^{65} or GST-U2AF$^{65}\Delta$35 fusion proteins immobilized on 10 μl of glutathione *S*-Sepharose 4B beads (Amersham Biosciences) and incubating with 5 μl of a standard [^{35}S]methionine-labeled rabbit reticulocyte RNase-A-treated lysate reaction (TNT T7 kit; Promega). Incubation was carried out in 20 mM HEPES (pH 7.5), 100 mM NaCl, 0.5% Nonidet P-40, 5 mM $MgCl_2$, 0.1 mM EDTA, 100 μg/ml bovine serum albumin, 1 mM dithiothreitol, for 2 h at 4 °C. After extensive washing of the beads, proteins were eluted by boiling in SDS-loading dye, electrophoresed on SDS-10% polyacrylamide gels, and visualized by autoradiography. Alternatively, 0.5 μg of GST-U2AF^{65} or GST-U2AF$^{65}\Delta$35 was incubated with 0.5 μg of His-U2AF^{35}a or His-U2AF^{35}b in 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 0.5% Nonidet P-40, on ice, for 15 min. Then 10 μl of Ni^{2+}-NTA-agarose beads (Qiagen) was added, and incubation was continued at 4 °C for 1 h. Beads were washed three times with 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 15 mM imidazole and resuspended in 15 μl of SDS-loading dye. Proteins were resolved on SDS-PAGE, transferred to nitrocellulose, and probed by Western blotting with anti-U2AF^{65} monoclonal antibody MC3 (36) and anti-U2AF^{35} polyclonal serum. Immunoblots were developed using horseradish peroxidase-coupled secondary antibodies and detected by enhanced chemiluminescence (ECL; Amersham Biosciences).

UV Cross-linking—U_{15}CAG oligo RNA was 5′ end-labeled with [γ-^{32}P]ATP (Amersham Biosciences) and T_4 polynucleotide kinase (New England Biolabs). RNA-protein binding reactions were assembled in buffer D, with 2 mg/ml tRNA, 0.5 μM purified GST-U2AF^{65}, and His-tagged U2AF^{35}a or U2AF^{35}b, and ~50,000 cpm ^{32}P-labeled RNA in a final volume of 20 μl. After a 15-min incubation on ice, samples were UV cross-linked (Stratalinker; 254 nm, 0.6 J, 4-cm distance to light source) and then treated with RNase A (final concentration, 1 mg/ml) at 37 °C for 20 min. Samples were mixed with SDS-loading dye, boiled for 5 min, and loaded on an SDS-10% polyacrylamide gel. The gel was fixed and dried, and cross-linked proteins were detected by autoradiography.

In Vitro Splicing Assays—HeLa nuclear extract was prepared as described by Dignam *et al.* (37) and depleted of U2AF exactly as described (38) by passing the extract over an oligo(dT)-cellulose column at 1 M KCl. The column flow-through of this procedure yielded the depleted nuclear extract (odTΔNE). Transcription templates were generated by PCR using plasmids harboring the sequence of AdML (39, 40) preceded by an SP6 promoter. Full-length substrates were transcribed in the presence of a CAP analogue (m7G(5′)ppp(5′)G, New England Biolabs) and [α-^{32}P]UTP (Amersham Biosciences) as described (17). After a 2-h incubation at 37 °C, the transcripts were gel-purified, ethanol-precipitated, and resuspended in water. Splicing reactions and splicing complementation assays were performed as described previously (17). Spliced products were resolved on 13% denaturing polyacrylamide gels in Tris/borate/EDTA buffer (TBE) and analyzed by autoradiography.

Real Time Quantitative PCR—Total RNA was extracted from a variety of murine tissues (brain, heart, lung, and skeletal muscle) using the TRIzol® reagent (Invitrogen) and treated with RNase-free DNase I (Roche Diagnostics). The concentration of RNA was determined by spectrophotometry, and RNA quality was assessed by gel electrophoresis. Only samples yielding distinct 28 S and 18 S bands and A_{260}/A_{280} ratios between 1.8 and 2.1 were further used. Production of cDNA was carried out using Superscript II reverse transcriptase (Invitrogen). TaqMan® Minor Groove Binder probes and primers (Table I) were designed by using the computer software Primer Express (Applied Biosystems, Foster City, CA). Primer and probe concentrations were optimized for each amplicon using Mouse Universal RNA (Invitrogen) as a template. The values obtained for each set of primers and probe are listed in Table I. For 18 S quantification we used a pre-developed assay (D64333760, Applied Biosystems, Foster City, CA). All reactions were performed in the ABI7000 Sequence Detector (Applied Biosystems, Foster City, CA). The relative expression of each gene in murine tissues was calculated using a derivative of the $2(-\Delta\Delta C(t))$ method, as described (41). To validate the use of the $2^{-\Delta Ct}$ method, serial dilutions of cDNA prepared from Mouse Universal RNA were amplified by real time PCR using gene-specific primers and fluorogenic probes for U2AF^{35}a, U2AF^{35}b, U2AF^{65}, and 18 S rRNA. All the analyzed transcripts exhibited high linearity amplification plots ($r^2 > 0.98$) (data not shown). The slopes of the *Ct versus* the log RNA quantity were −3.36, −3.27, −3.28, and −3.27 for U2AF^{35}a, U2AF^{35}b, U2AF^{65}, and 18 S rRNA, respectively. The similarity of these slopes clearly shows that the PCR efficiencies for these genes are similar. Also when doing ΔCt plots *versus* log RNA quantities for each pair of genes, the slope was approximately equal to 0, further confirming that the expression of each of these genes can be directly compared with one another.

Sequence Analysis—BLAST searches (42) were carried out on NCBI data bases (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/blast) and on the Mouse, *Fugu,* and *Danio rerio* Ensembl genomic sequence data bases (43) (www.ensembl.org/). Genomic structures were further analyzed using Wise2 software (44) from EMBL-EBI (European Bioinformatic Institute; www.ebi.ac.uk/). Analysis of synonymous and nonsynonymous codon positions was carried out with SNAP (45). Protein sequences were obtained from SPTREMBL, SWISSPROT, and GenBank™ and predicted from ESTs sequences. Multiple sequence alignments were generated with Clustal X (1.83) (46), with no adjustment of the default parameters, and shaded using BOXSHADE 3.21 (www.ch.embnet.org/) according to conservation and similarity of residues

RESULTS

Cloning of Chicken U2AF1 Gene and Identification of Three Alternatively Spliced and Polyadenylated mRNA Isoforms—Chicken (*Gallus gallus*) *U2AF1* cDNA was cloned using degenerate primers complementary to previously defined conserved regions in the *U2AF1* sequence from *H. sapiens*, *D. melanogaster,* and *S. pombe* (19, 47, 48). These primers were used to amplify by PCR a chicken U2AF^{35} cDNA fragment, which was subsequently cloned and sequenced. The sequence (data not shown) revealed a cDNA closely related to those of other organisms. This cDNA fragment was then used as a probe to screen a chicken embryonic cDNA library. Restriction digestion and sequence analysis of five isolated cDNA clones revealed the presence of three distinct classes of *U2AF1* transcripts (Fig. 1*A*). The first class consists of a 915-bp cDNA that is 86% identical to the human *U2AF1* cDNA reported previously (19). The open reading frame of this clone encodes a putative protein of 237 amino acid residues with a predicted M_r of 27,684. An alignment of the amino acid sequences encoded by chicken and human *U2AF1* genes shows that the two proteins are 99%

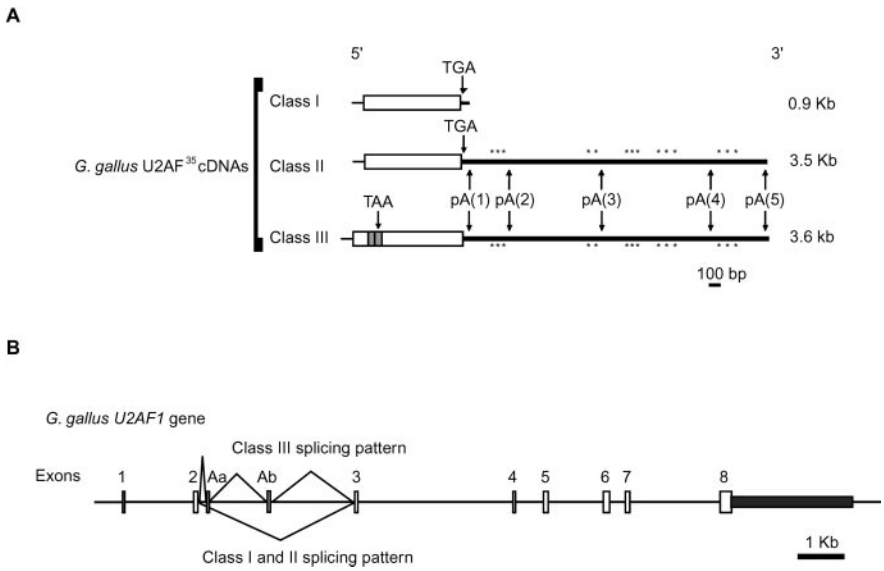

FIG. 1. **Structure of chicken *U2AF1* gene.** *A*, schematic representation of the three classes of cDNA isolated from an embryonic library. Classes II and III cDNAs contain a long 3′-UTR (*solid line*). This 3′-UTR has a high content of A and T residues and multiple copies of the ATTTA motif (*asterisks*). Potential polyadenylation sites in the 3′-UTR are indicated (pA). Class III cDNAs contain an in-frame premature stop codon (TAA). *B*, chicken *U2AF1* gene spans ~13 kb and contains 10 exons (*boxes*). Exons Aa and Ab can be alternatively spliced and the two splicing patterns identified are indicated; cDNAs from classes I and II arise by splicing out exons Aa and Ab, whereas class III transcripts result from an alternative splice site selection that causes inclusion of these two exons.

identical and differ only at the glycine tract located at the C-terminal part of the protein. The U2AF^{35} glycine-stretch is 12 residues long in the human protein, whereas in chicken it contains only 9 residues.

The remaining two classes have not been reported previously. Compared with the first class, the cDNAs of classes II and III are much larger (3.5 and 3.6 kb, respectively). Both contain a long 2.6-kb 3′-untranslated region that arises by use of an alternative polyadenylation site located 2570 bp downstream of the translation stop codon. This long 3′-UTR has a high content of A and T residues and contains multiple copies of the pentanucleotide ATTTA (Fig. 1*A*, *asterisks*), which is thought to function as an RNA-destabilizing element (49). Within the long 3′-UTR, four additional potential polyadenylation signals were identified (Fig. 1*A*, *pA*). Similarly to class I cDNA, the transcripts of class II encode a full-length U2AF^{35} protein. In contrast, class III cDNAs contain a novel 124-bp sequence that introduces a premature termination codon (Fig. 1*A, TAA*).

From the screening of a genomic library, six clones were isolated and sequenced. These clones span the entire genomic sequence of chicken *U2AF1*, thus allowing the determination of the complete exon-intron organization of this gene (Fig. 1*B* and Table II). In addition to the eight exons identical in structure to the previously described *U2AF1* gene from man and *Fugu rubripes* (50), two novel exons termed Aa and Ab were identified. The sequence of these two new exons corresponds precisely to the 124-bp insert observed in class III cDNAs. Thus, we conclude that cDNAs from classes I and II arise by splicing out exons Aa and Ab, whereas class III transcripts result from an alternative splice site selection that causes inclusion of these two exons.

By having established that chicken *U2AF1* transcripts can be alternatively spliced and polyadenylated, we next analyzed the expression pattern of this gene by Northern blotting. Hybridization of poly(A) RNA from DT40 cells (51) with a chicken *U2AF1* cDNA fragment reveals three major mRNA species of ~1.1, 1.5, and 3.8 kb (Fig. 2*A*, *lane 1*). In contrast, a probe complementary to the last 1175 nucleotides of the 3′-UTR hybridizes only to the 3.8-kb species (Fig. 2*A*, *lane 2*). From these results we infer that the 1.1-kb mRNA corresponds to class I cDNA, whereas the largest form (3.8 kb) may correspond to cDNAs of classes II or III (because class II and III transcripts differ by only 124 nucleotides, they cannot be resolved in these gels). The 1.5-kb mRNA may result from the use of the second polyadenylation signal (located 553 nucleotides downstream of the translation stop codon), but further experiments are needed to confirm this interpretation. The same three species of mRNAs are expressed in chicken embryos and pre-B lymphoid cells (Fig. 2*B*). However, densitometer scanning of the autoradiograms indicated that the relative proportion of the 1.1-kb species varied from ~46% in embryonic cells to over 70% in DT40 cells. In summary, we conclude that in chicken there are multiple forms of *U2AF1* mRNAs, which arise by alternative processing of the primary transcripts, and that there might be cell type variations in the relative amounts of *U2AF1* mRNA isoforms.

Three Different mRNA Isoforms Are Produced by Alternative Splicing of Human U2AF1 Gene—In an attempt to identify sequences corresponding to exons Aa and Ab in the human *U2AF1* gene, we performed BLAST searches against the human chromosome 21 genomic contig (accession number NT_030188). This allowed us to identify a sequence homologue to chicken exon Ab in the second intron of *hU2AF1* (Fig. 3*A*), but we could not find any sequence homologue to chicken exon Aa in either the *hU2AF1* gene or ESTs data base. Sequence analysis of human ESTs led us to group mRNAs according to three different alternative splicing patterns (Fig. 3, *A* and *B*). Pattern a corresponds to the mRNA previously described (19). In splicing pattern b, exon Ab substitutes exon 3 with no alteration of the reading frame. Translation of this type of transcript gives rise to a protein product very similar to the known U2AF^{35}, with only 7 different amino acid residues (Fig. 3*C*). Finally, as result of splicing pattern c, exon Ab is included between exons 2 and 3 and induces a frameshift that introduces a premature termination codon (TAA), similarly to what is observed in chicken class III cDNAs (see Fig. 2*B*). We shall hereafter refer to the isoform produced by each alternative splicing pattern as U2AF^{35}a, U2AF^{35}b, and U2AF^{35}c, respectively.

U2AF^{35}c Transcripts Are Degraded by NMD—In eukaryotes, mRNAs that contain premature stop codons tend to be selectively degraded by a critical quality control mechanism termed nonsense-mediated decay (NMD) (reviewed in Refs. 52–54). Because the *hU2AF1* gene transcripts that undergo splicing pattern c carry a premature termination codon, we asked whether the mRNA isoform U2AF^{35}c is targeted to NMD. Because a first round of translation is required to trigger NMD, we treated HeLa cells with the translational inhibitor cycloheximide (20 μg/ml). Reverse transcriptase-PCR was then per-

TABLE II
Exon-intron organization of chicken U2AF1 gene

Intronic sequences are in lowercase letters; exons are in capital letters.

Exon	Size	Intron	5′ splice site	Size	3′ splice site
	bp			*bp*	
1	>44	1–2	CAA:gtgagtacccca	1,490	ttatacttgcag:AGT
2	88	2-Aa	CAG:gtttgttttcct	216	taatggctgcag:GAA
Aa	57	Aa-Ab	CAG:gtgagtgtgcct	1,241	ttttccccgcag:ACC
Ab	67	Ab-3	ACT:gtaagtcccaca	1,792	tttgtatttcag:ACC
3	67	3–4	GCT:gtaagttcaata	3,300	tggtttttacag:GTG
4	50	4–5	GAG:gtatgcagaaaa	611	gtttgtgtttag:GAG
5	99	5–6	AAG:gtaggataaaaa	1,200	ttttgcttttag:TTT
6	134	6–7	GGG:gtaaaaagcaac	329	attgtgtttcag:AGA
7	93	7–8	AAA:gtgagtttgctt	1,935	ctttctttgaag:GCA
8	242/2,828				

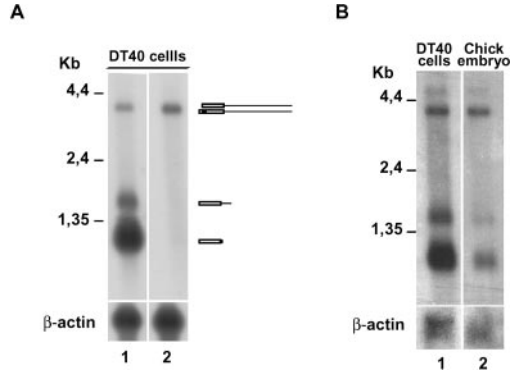

FIG. 2. **Northern blot analysis of chicken *U2AF1* mRNAs.** *A,* poly(A) RNA was isolated from chicken DT40 cells and probed with either PCR35 chicken U2AF^{35} cDNA fragment (*lane 1*) or with a probe complementary to the last 1175 nucleotides of the 3′-UTR (*lane 2*). The structure of each mRNA species is illustrated on the *right. B,* poly(A) RNA isolated from DT40 cells (*lane 1*) and chicken embryos (*lane 2*) probed with PCR35 fragment. A probe complementary to chicken β-actin was also used as an internal control.

formed on total cellular RNA to determine the abundance of each U2AF^{35} isoform. Prior to treatment, the isoform U2AF^{35}c was nearly undetectable, while after 2–10 h of treatment with cycloheximide it became progressively more abundant (data not shown). A key factor in NMD is hUpf1, an RNA helicase that is part of the mRNA surveillance complex (55). We therefore used siRNAs to down-regulate hUpf1 protein in HeLa cells, as described previously (29, 31). PCR was carried out using primers flanking the alternative regions, and the amplified DNA was then digested with HinfI to allow the discrimination of the three isoforms. As depicted in Fig. 4, 48 h after transfection of HeLa cells with siRNAs against hUpf1, there is a specific increase in the steady-state levels of the U2AF^{35}c isoform (Fig. 4, *lane 3*), compared with control transfected cells (Fig. 4, *lane 2*). Western blot analysis confirmed that the level of Upf1 protein was significant reduced in the transfected cells (not shown). From this we conclude that *U2AF1* pre-mRNAs that undergo splicing pattern c, which includes exon Ab between exons 2 and 3 introducing a premature termination codon, are subject to NMD.

U2AF^{35}b Protein Isoform Binds to U2AF^{65} and Enhances in Vitro U2AF Splicing Activity—Full-length human U2AF^{35}a cDNA was used as a probe to screen a human fetal spleen cDNA library, and positive clones were analyzed by restriction digestion and sequenced in order to isolate U2AF^{35}b cDNA clones. U2AF^{35}b cDNA encodes a protein product 97% identical to the known U2AF^{35}a, with 7 amino acid substitutions at positions 47–49, 59, 61, 65, and 66 (Fig. 3*C*). This region is part of the noncanonical RRM domain of U2AF^{35} involved in dimerization with U2AF^{65}, and although the residues that directly interact with U2AF^{65} are identical between the two isoforms (18), we asked whether the substitution of exon 3 (in U2AF^{35}a) by exon Ab (U2AF^{35}b) would affect the physical interaction with U2AF^{65}. For these experiments we used recombinant U2AF^{65} and U2AF^{65}Δ35 (a U2AF^{65} deletion mutant lacking amino acid residues 95–138, unable to interact with U2AF^{35} (34)) expressed as N-terminal GST fusion proteins in *E. coli*. Recombinant proteins were purified by affinity chromatography on glutathione-Sepharose 4B beads and incubated with either U2AF^{35}a or U2AF^{35}b transcribed and translated *in vitro* in the presence of [^{35}S]methionine. Fig. 5*A* shows that GST-U2AF^{65} immobilized on glutathione-Sepharose beads interacts with both U2AF^{35}a (*lane 1*) and U2AF^{35}b (*lane 2*), whereas the signal detected with GST-U2AF^{65}Δ35 (*lane 3*) is close to background level (*lane 4*). This result was confirmed by incubating His-tagged recombinant U2AF^{35}a and U2AF^{35}b with either GST-U2AF^{65} or GST-U2AF^{65}Δ35 (Fig. 5*B*). Proteins linked to His-tagged U2AF^{35} were purified using Ni^{2+}-NTA-agarose beads, fractionated by SDS-PAGE, and analyzed by Western blotting with anti-U2AF^{65} monoclonal antibody. As can be seen in Fig. 5*B*, both His-U2AF^{35}a and His-U2AF^{35}b bind to GST-U2AF^{65} (*lanes 1* and *3*), whereas no His-U2AF^{35}b binds to either GST-U2AF^{65}Δ35 (*lane 4*) or uncoated beads (*lane 5*).

By having established that U2AF^{35}b binds U2AF^{65}, we wished to determine whether the heterodimer U2AF^{65}/U2AF^{35}b recognizes a pre-mRNA substrate as described for the heterodimer U2AF^{65}/U2AF^{35}a (14–16). For these experiments, purified recombinant GST-U2AF^{65} protein was used in RNA-binding assays either alone or together with recombinant His-U2AF^{35}a and His-U2AF^{35}b (Fig. 5*C*). Similar amounts of proteins were exposed to UV light in the presence of a ^{32}P-labeled RNA substrate. The RNA substrate is an oligonucleotide (U_{15}CAG) that mimics a U-rich polypyrimidine tract followed by the consensus 3′ splice site AG. Although GST-U2AF^{65} alone binds the RNA substrate (Fig. 5*C, lane 1*), addition of either U2AF^{35}a (*lane 2*) or U2AF^{35}b (*lane 3*) results in a stronger signal. This suggests that U2AF^{35}b can replace U2AF^{35}a in stabilizing U2AF^{65} binding to pre-mRNA substrates.

Next, we asked whether U2AF^{35}b can functionally replace U2AF^{35}a in an *in vitro* splicing reaction. Oligo(dT) chromatography was performed to deplete U2AF from HeLa nuclear extracts, as described previously (38). The extract obtained from the column odTΔNE was unable to support *in vitro* splicing assays unless complemented with U2AF activity. For some splicing substrates that contain strong 3′ splice site signals, such as AdML, complementation can be achieved with U2AF^{65} alone (23) (Fig. 5*D*, compare *lanes 2* and *3*). However, simultaneous addition of U2AF^{65} and U2AF^{35}a significantly enhances the efficiency of AdML pre-mRNA splicing (Fig. 5*D*, compare *lanes 3* and *4*). Although U2AF^{35}b can also enhance splicing activity, this isoform appears to be less efficient than U2AF^{35}a (compare *lanes 4* and *5*).

Collectively, these data suggest that the alternative inclu-

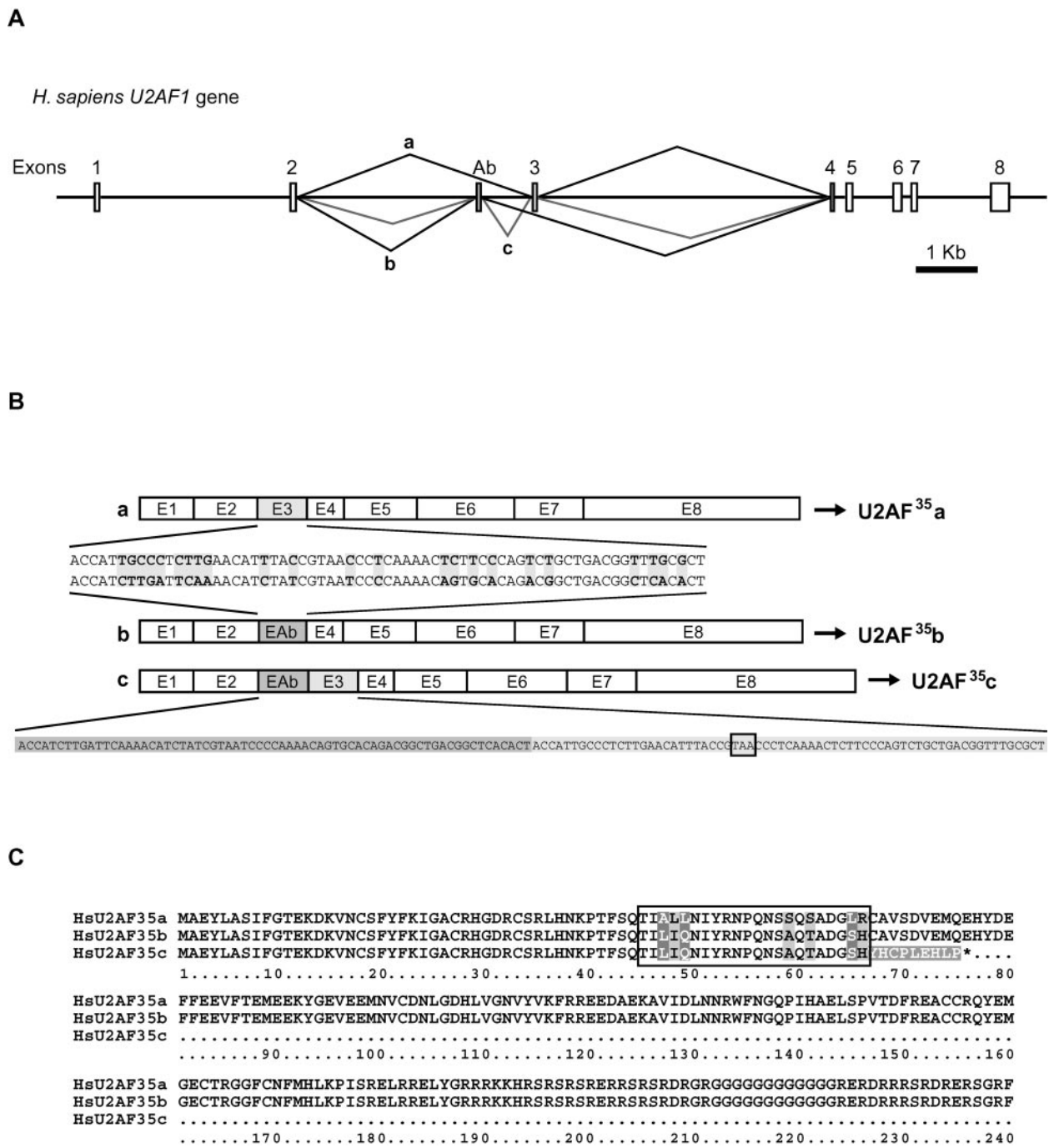

FIG. 3. **Alternative splicing of human *U2AF1* gene.** *A*, structure of human *U2AF1* gene; exons are represented by *boxes* and introns by *lines*. A homologue of chicken exon Ab is present in the sequence of human *U2AF1* second intron. The alternative splicing patterns are indicated (*a–c*); *B*, representation of the three distinct alternative splicing products of human *U2AF1* gene; form a corresponds to the cDNA reported by Zhang *et al.* (19); in form b exon Ab substitutes exon 3; and in form c exon Ab is included between exons 2 and 3, introducing of a premature stop codon (TAA). Sequences of exons 3 and Ab are shown, and the different nucleotides between both are highlighted in *gray*. *C*, comparison of the amino acid sequences translated from the three alternative splicing products. The peptide sequences coded by exons 3 and Ab are *boxed* and differ in 7 amino acid residues. Nonsimilar amino acids are depicted as *dark gray*, and similar amino acids are depicted as *light gray*. The *asterisk* indicates a stop codon.

sion of exon Ab maintains the ability of the resulting U2AF^{35} protein isoform to interact with U2AF^{65}, stabilizes U2AF^{65} binding to a pre-mRNA substrates, and enhances U2AF splicing activity *in vitro*.

U2AF^{35}a and U2AF^{35}b Protein Isoforms Have Similar Cellular Localization—A GFP-tagging approach combined with confocal microscopy was used to compare the cellular localization of U2AF^{35}a and U2AF^{35}b proteins. Plasmids encoding each U2AF^{35} protein fused to GFP at the N terminus were transfected into HeLa cells. As shown in Fig. 6, GFP-U2AF^{35}a (*A*) and GFP-U2AF^{35}b (*B*) are similarly distributed. The two proteins are exclusively detected in the nucleus, excluding the nucleoli. Both isoforms are detected diffuse in the nucleoplasm with additional concentration in speckles, as described previously (35) for GFP-U2AF^{65}.

U2AF^{35}a and U2AF^{35}b mRNA Isoforms Are Differentially Expressed in a Tissue-specific Manner—A real time, quantitative PCR assay was performed to compare the expression of U2AF^{35}a, U2AF^{35}b, and U2AF^{65} mRNAs in a variety of mouse tissues (brain, heart, lung, and skeletal muscle). The probes

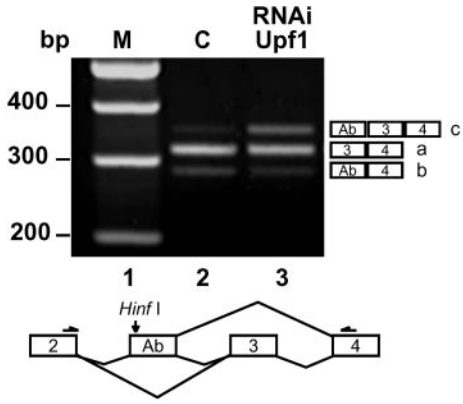

FIG. 4. **U2AF^{35}c mRNA isoform is targeted to NMD.** Depletion of cellular Upf1 was achieved by RNA interference by using specifically designed 21-nucleotide double-stranded RNAs. At 48 h after the second transfection of HeLa cells with the siRNA, total RNA was extracted from control transfected cells (*C, lane 2*) or cells transfected with siRNA against Upf1 (*lane 3*). Reverse transcriptase-PCR amplification was carried out on total cellular mRNA using primers flanking the alternatively spliced regions. The amplification products were digested with HinfI, which specifically cleaves the PCR products inside exon Ab, separated by agarose gel electrophoresis, and detected by ethidium bromide staining. The structure of each product is illustrated on the *right.* Molecular weight markers are indicated on the *left* (*lane 1*).

and primers are schematically depicted in Fig. 7*A*. To test that the probes were specific, we used as template plasmid DNA containing U2AF^{35}a, U2AF^{35}b, or U2AF^{65} cDNA inserts. We observed amplification only when the specific probe for the template plasmid was used (data not shown).

First, we compared the relative expression of U2AF^{35}a and U2AF^{35}b mRNA isoforms (Fig. 7*B*). The results clearly show that U2AF^{35}a is the major isoform expressed in all tissues examined. However, the ratio between U2AF^{35}a and U2AF^{35}b differs from tissue to tissue, being lowest in brain (where U2AF^{35}a is 10-fold more abundant than U2AF^{35}b) and highest in lung and skeletal muscle (where U2AF^{35}a is ~20-fold more abundant than U2AF^{35}b). These observations suggest that alternative splicing of primary transcripts from the *U2AF1* gene is regulated in a tissue-specific manner.

Next, we compared the expression of the major U2AF^{35} mRNA isoform (U2AF^{35}a) relative to U2AF^{65} mRNA (Fig. 7*C*). The results show that in skeletal muscle U2AF^{35}a and U2AF^{65} mRNAs are expressed at similar levels. However, in heart the level of U2AF^{35}a mRNA is lower than that of U2AF^{65}. An even lower expression ratio was detected in brain, where the level of U2AF^{35}a mRNA is less than half of U2AF^{65} mRNA. This shows for the first time that in mammals there is a tissue-specific regulation of U2AF gene expression.

Duplicate U2AF1 Genes in F. rubripes—The identification of previously unknown exons that can be alternatively spliced in chicken and human *U2AF1* transcripts prompted us to investigate the evolutionary history of this gene. A search of data bases of expressed sequence tags revealed several U2AF^{35} ESTs containing a sequence homologue to chicken and human exon Ab. These include ESTs from *Mus musculus* (*e.g.* GenBank^{TM} accession number CB203010), *Rattus norvegicus* (*e.g.* GenBank^{TM} accession number CB797022), *Bos taurus* (*e.g.* GenBank^{TM} accession number CB418879), *Sus scrofa* (*e.g.* GenBank^{TM} accession number BE235127), *Oryzia latipes* (*e.g.* GenBank^{TM} accession number BJ527861), and *Ictalurus punctatus* (*e.g.* GenBank^{TM} accession number CB940474). We could not find any sequence homologue of chicken exon Aa.

Looking specifically to the mouse *U2AF1* gene located on chromosome 17 (www.ensembl.org/; accession ENSMUSG00000024040) we could also identify the exon Ab homologue on the second intron of Mm*U2AF1* gene, confirming the alternative splicing patterns identified on the ESTs data bases. We then used human exon Ab sequence to search the draft sequence generated by the Japanese pufferfish *F. rubripes* genome project, and surprisingly, the only result with high score obtained was a sequence on scaffold_79 (length 361,291 bp). The *Fugu U2AF1* gene is annotated on scaffold_3186 (length 17,872 bp; accession SINFRUG00000150274). Although there is a second annotation on scaffold_79 (accession SINFRUG00000124936), it shows a very short and unlikely sequence, and we decided to analyze it in greater detail. By using GENEWISE software (44) the amino acid sequence from human U2AF^{35}a was aligned with the nucleotide sequence on scaffold_79 and that in fact revealed the existence of a second putative *U2AF1* gene comprising 8 exons. The new prediction of exon sizes and intron placement match the gene structure of *U2AF1* gene on scaff._3186 (Fig. 8 *A*; Table III), and all the intronic sequences identified are in agreement with the GT/AG splice site rule. Because duplicated gene copies without evolutionary constraints can frequently evolve as pseudogenes, we compared the number of synonymous *versus* nonsynonymous substitutions in the *Fugu U2AF1* genes. The nucleotide identity between exonic sequences is about 70% with a 10-fold predominance of synonymous codon substitutions (0.95 and 0.07 substitutions per synonymous and nonsynonymous position, respectively), which strongly suggests that both putative *U2AF1* genes evolved under selective conditions. Most interestingly, the third exon of the predicted *U2AF1* gene located on scaffold_3186 is a homologue of human exon 3, whereas the third exon of the predicted *U2AF1* gene located on scaffold_79 is a homologue of exon Ab. These data suggest that the pufferfish *F. rubripes* contains two separate genes encoding homologues of the U2AF^{35}a and U2AF^{35}b protein isoforms that are generated in higher vertebrates using alternative exons from a single gene. We propose to refer to each *Fugu* gene as *U2AF1-a* and *U2AF1-b,* respectively.

Based on the ancient fish-specific genome duplication hypothesis (56), we extended our search looking for duplicated *U2AF1* orthologues in other teleost species. Searching the Ensembl genomic sequence data base of *D. rerio,* we only found one putative *U2AF1* gene on chromosome fragment ctg30223, containing exon 3 (accession ENSDARG00000015325). However, it cannot exclude the hypothesis that *U2AF1* duplication occurred before the divergence of zebrafish and pufferfish but with one copy being lost or not sequenced in zebrafish. Additionally, it is worthwhile to mention that the protein products predicted from the teleosts *O. latipes* and *I. punctatus* ESTs most probably correspond to the U2AF^{35}b isoform (Fig. 8*B*).

Taken together these data argue that the genomic sequences that encode both U2AF^{35}a and U2AF^{35}b protein isoforms have been highly conserved in vertebrates, from fish to mammals.

DISCUSSION

This work describes the cloning and sequencing of the chicken *U2AF1* gene coding for the small subunit of the splicing factor U2AF. This work also provides the first evidence that vertebrate *U2AF1* transcripts can be alternatively spliced and polyadenylated.

Variations in the relative concentration of general splicing factors and heterogeneous nuclear RNPs can affect splice site choice and therefore regulate alternative splicing decisions (57). However, the mechanisms that control the ratio of the relevant splicing factors in any given cell type remain largely unknown. One possibility is that the transcripts coding for the splicing factors are themselves targets for post-transcriptional regulation. In fact, a tissue-specific expression of alternatively processed transcripts has been described for the following three

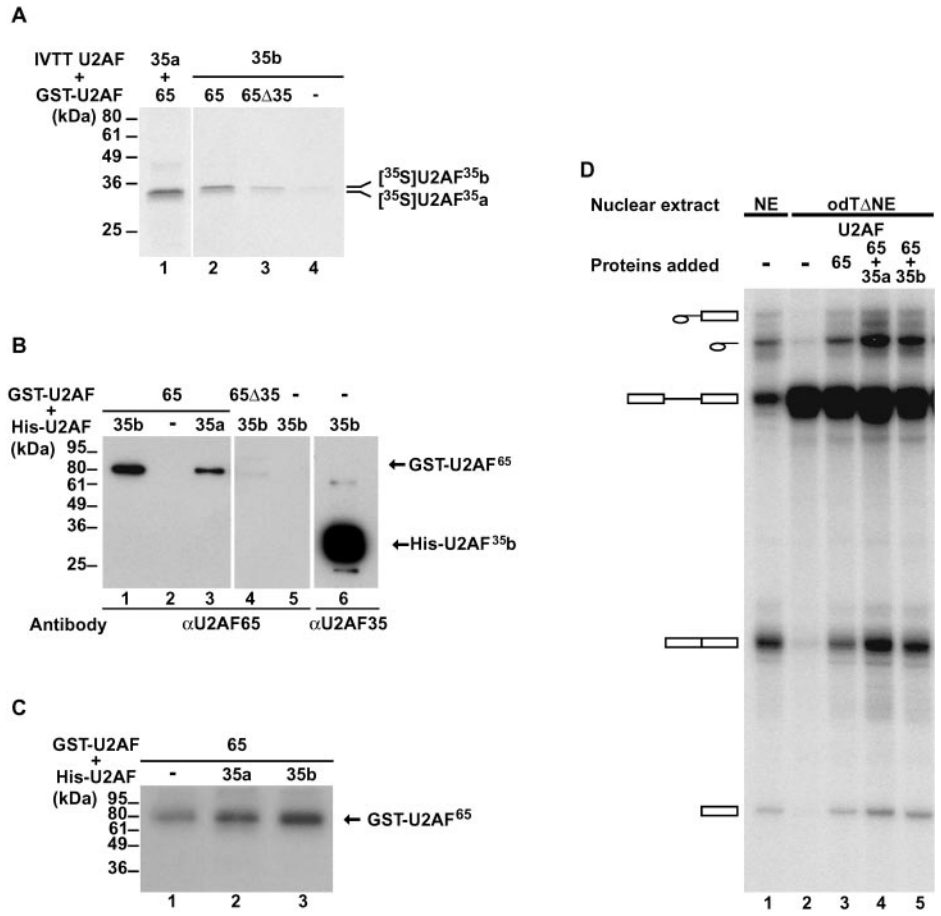

FIG. 5. **U2AF^{35}b protein isoform interaction with U2AF^{65} and functional analysis of U2AF^{65}/U2AF^{35}b heterodimer.** *A*, U2AF^{35}a (*lane 1*) and U2AF^{35}b (*lanes 2–4*) protein isoforms were *in vitro* transcribed-translated (*IVTT*) and [^{35}S]methionine-labeled. After incubation with immobilized GST-U2AF^{65} (*lanes 1* and *2*), GST-U2AF^{65}Δ35 (*lane 3*), or Sepharose glutathione beads alone (*lane 4*), proteins bound to the beads were fractionated by SDS-PAGE and visualized by autoradiography. Molecular weight markers are indicated on the *left*. *B*, GST-U2AF^{65} (*lanes 1–3*) and GST-U2AF^{65}Δ35 (*lane 4*) were incubated with His-tagged U2AF^{35}b (*lanes 1, 4,* and *5*) or U2AF^{35}a (*lane 3*). After purification with Ni^{2+}-NTA-agarose beads, proteins were fractionated by SDS-PAGE and analyzed by Western blotting using either a monoclonal antibody against U2AF^{65} (*lanes 1–5*) or a rabbit serum against U2AF^{35} (*lane 6*). In control *lane 2,* no His-tagged U2AF^{35} was added, and in control *lanes 5* and *6*, no GST-U2AF^{65} was added. Molecular weight markers are indicated on the *left*. *C*, the radioactively labeled U_{15}CAG RNA oligonucleotide was incubated with GST-U2AF^{65} alone (*lane 1*), or GST-U2AF^{65} together with either His-U2AF^{35}a (*lane 2*) or His-U2AF^{35}b (*lane 3*). After UV irradiation, the samples were subjected to electrophoresis and autoradiography. Molecular weight markers are indicated on the *left*. *D*, *in vitro* splicing reconstitution assay. Radioactively labeled AdML pre-mRNA substrate was incubated in HeLa nuclear extract (*NE*, *lane 1*) or extract chromatographically depleted of U2AF (*odTΔNE*, *lane 2*). The reconstitution assay was performed by adding to odTΔNE recombinant GST-U2AF^{65} alone (*lane 3*) or GST-U2AF^{65} in the presence of either His-U2AF^{35}a (*lane 4*) or His-U2AF^{35}b (*lane 5*). Splicing substrate and products are indicated schematically on the *left*.

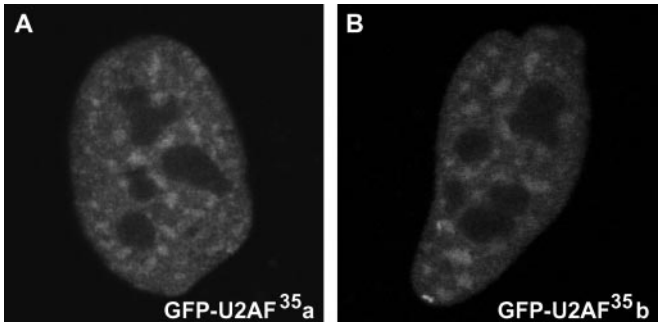

FIG. 6. **U2AF^{35}a and U2AF^{35}b protein isoforms show similar subcellular localization.** HeLa cells were transfected with pEGFP plasmid derivatives expressing GFP fusions of U2AF^{35}a (*A*) and U2AF^{35}b (*B*). Approximately 20 h after transfection, cells were fixed in formaldehyde and directly observed with a confocal microscope.

members of the SR family of splicing factors: SC35/PR264 (58), 9G8 (59), and SRp20 (60, 61). In the case of SC35/PR264, several mRNAs with variable 3′-untranslated regions and different stability are produced by alternative splicing and polyadenylation (58). Similarly, distinct mRNAs coding for splicing factor 9G8 can be generated by alternative splicing of intron 3 and use of two alternative polyadenylation sites (59). Expression of SRp20 mRNAs is also regulated during the cell cycle (61), and the SRp20 protein has been shown to auto-regulate the alternative splicing of its own pre-mRNA, whereas another SR protein (ASF/SF2) antagonizes this activity (60). More recently, polypyrimidine tract binding protein and the SR-like protein HTRA2-BETA1 were added to the growing list of splicing factors that can auto-regulate their own expression at the post-transcriptional level (29, 62).

In the chicken, human, and mouse *U2AF1* genes, we identified a novel exon, referred to as exon Ab. The simultaneous inclusion of exon Ab and exon 3 by alternative splicing introduces an in-frame premature termination codon, and we have classified this type of mRNA as isoform U2AF^{35}c (Fig. 3*B*). Our experimental data further indicate that human U2AF^{35}c mRNAs are targeted to NMD, a mechanism by which cells selectively recognize and degrade mRNAs that contain premature termination codons (for recent reviews see Refs. 52–54). Although NMD was classically considered to represent a major

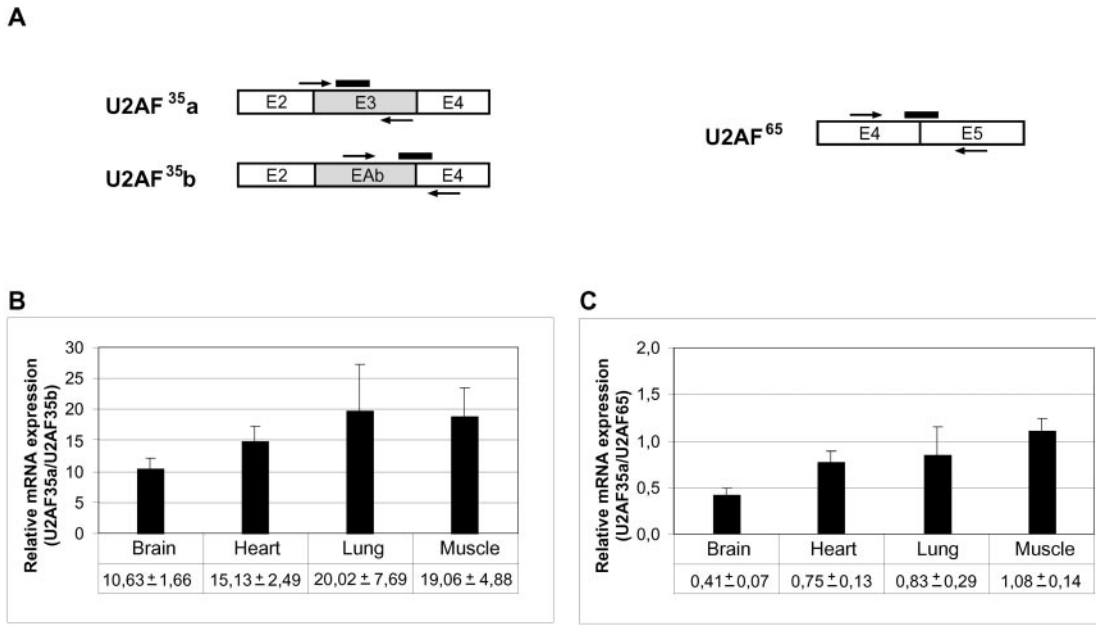

FIG. 7. **Relative expression of U2AF^{35}a, U2AF^{35}b, and U2AF^{65} mRNAs in mouse tissues determined by quantitative real time PCR.** *A*, schematic representation of gene-specific PCR primers (*arrows*) and TaqMan probes (*thick lines*) used to amplify and detect each cDNA; *boxes* represent exons. *B*, U2AF^{35}a mRNA level relative to U2AF^{35}b. *C*, U2AF^{35}a mRNA level relative to U2AF^{65}. The ratios between mRNA levels were calculated from the $2^{-\Delta Ct}$ formula as described under "Experimental Procedures," and results are expressed as mean ± S.D. (n = 12).

surveillance mechanism that destroys aberrant mRNAs generated as a result of biosynthetic errors or gene mutations, it was recently proposed that NMD additionally participates in a pathway that controls expression of a large number of genes (5). Based on a bioinformatic approach, Lewis *et al.* (5) predicted that 35% of human EST-suggested alternative isoforms contain premature termination codons and are potential targets for NMD. The same authors claim that alternative splicing of NMD-targeted transcripts may be even more prevalent than their data suggest. Additional experimental work will be necessary to characterize further the physiological role of coupled alternative splicing and NMD in the expression of U2AF^{35} splicing factor. However, the existence of alternative mRNA isoforms carrying premature termination codons and a possible link to NMD have been suggested for the splicing factors SC35 (63), TIA-1, and TIAR (64). More recently, it has been shown that regulated unproductive splicing and translation plays an important role in quantitative control of polypyrimidine tract binding protein gene expression (29). Thus, regulation of splicing factors by alternative splicing leading to NMD might prove to be a common mechanism of cellular regulation of their gene expression.

As the result of a distinct alternative splicing pattern of *U2AF1* transcripts, exon 3 may be replaced by exon Ab producing U2AF^{35} isoform b (here referred to as U2AF^{35}b). U2AF^{35}b mRNA gives rise to a protein 97% identical to the known U2AF^{35}, with only 7 different amino acid residues located in the RRM domain. Because this domain mediates dimerization with U2AF^{65} (18), we analyzed the ability of U2AF^{35}b protein isoform to interact with U2AF^{65}, to stabilize the binding of U2AF heterodimer to pre-mRNA, and to promote splicing activity *in vitro*. Although U2AF^{35}b was able to replace U2AF^{35}a in all experiments, the results suggest that it may have slightly less activity in splicing of the AdML pre-mRNA (Fig. 5*D*). From our experiments we cannot rule out that this apparent difference in specific activity is caused by variations in preparation of U2AF^{35}a and U2AF^{35}b recombinant proteins. However, three of the seven different amino acids of U2AF^{35}b are within the RNP-2 motif of the RRM, a change from alanine to leucine at position 47, a change from leucine to isoleucine at position 48, and a change from leucine to glutamine at position 49. Most importantly, based on modeling of previously published RRM/RNA co-crystal structures, it was proposed that Ala-47 of U2AF^{35} may directly contact RNA (18). Thus, it is compelling to think that U2AF heterodimers containing a U2AF^{35}a or U2AF^{35}b isoform may have slightly different specificity concerning the sequence context at the 3′ splice site. Therefore, differences in the expression levels of the two isoforms would influence alternative splicing by the recognition and selection of distinct 3′ splice sites.

BLAST analysis revealed that the genomic sequence homologues to chicken exon Ab are present in *U2AF1* genes from man, mouse, and pufferfish. Most interesting, two separate putative *U2AF1* genes are present in the genome of *F. rubripes*, one containing a homologue of human exon 3 (which we refer to as *U2AF1-a*) and the other containing a homologue of human exon Ab (which we refer to as *U2AF1-b*) (Fig. 8). This duplication of the *U2AF1* locus observed in *Fugu* could be part of the whole genome duplication proposed to have occurred in the lineage leading to modern day teleosts, when they separated from the last common ancestor they shared with amphibians, reptiles, birds, and mammals (56, 65–68).

The classical model for the fate of duplicate genes predicts that one member of the duplicate pair is generally inactivated through accumulation of degenerative mutations ("defunctionalization") or is occasionally preserved because of beneficial mutations that confer a novel function or an improvement of an existing function to one copy, with the other retaining the original function ("neofunctionalization"). However, contrary to the predictions of the classical model, vertebrates, particularly teleosts, seem to have retained a large proportion of gene duplicates (66, 69). Recently, a third alternative mechanism by which duplicate genes may be preserved was proposed. According to the duplication/degeneration/complementation model (70), when a gene duplicate has independently mutable subfunctions, the possibility exists that the two members of the

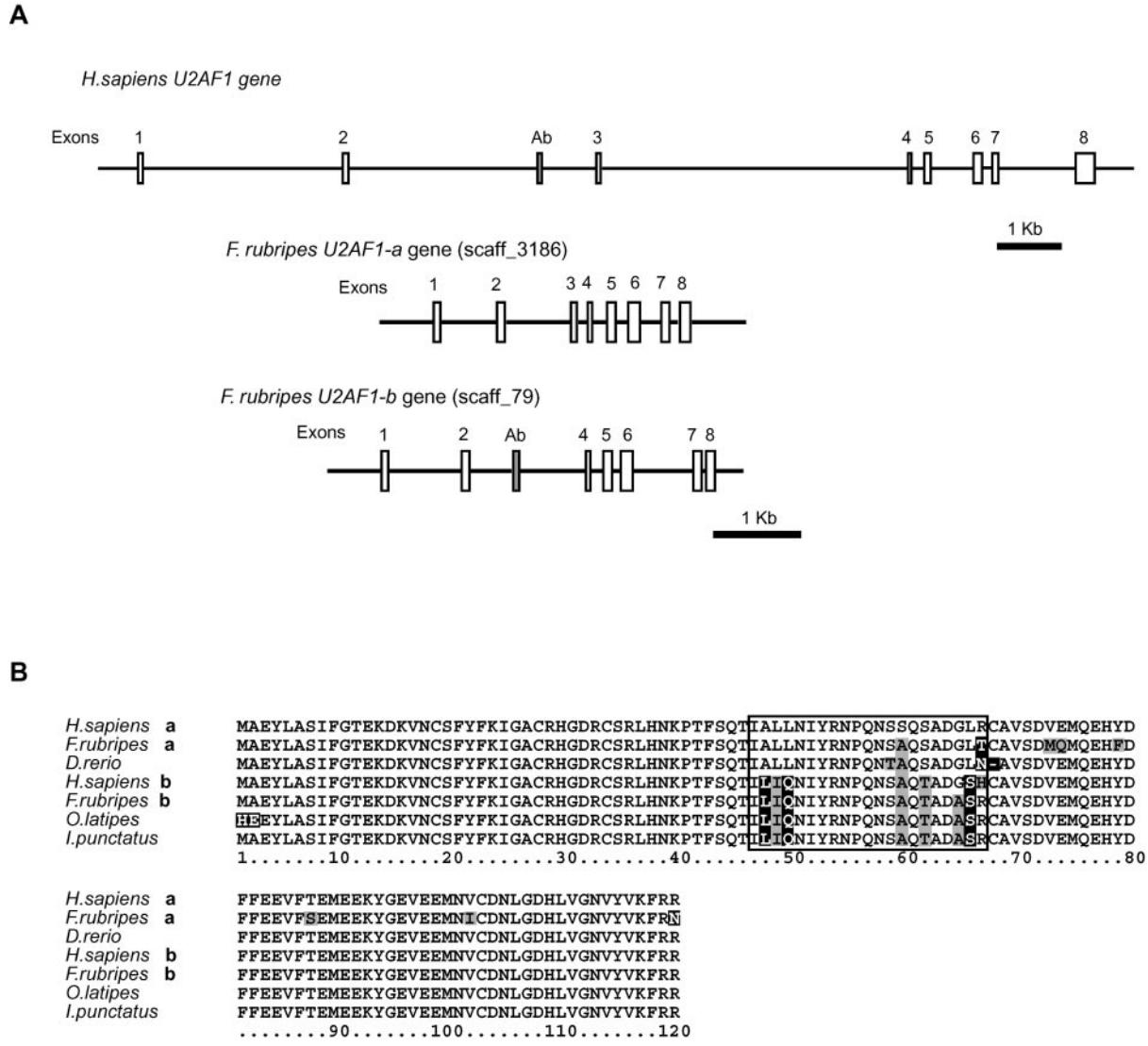

FIG. 8. **Structure of *F. rubripes U2AF1* genes and comparison of human and fish U2AF^{35} orthologues.** *A*, schematic representation of *Fugu U2AF1-a* and *U2AF1-b* genes. Introns are represented by *lines,* and exons are represented by *boxes* and numbered according to the homologous regions of the human *U2AF1* gene. *Fugu U2AF1-a* gene contains a homologue of human exon 3, and *U2AF1-b* contains a homologue of human exon Ab. *B*, alignment of human and fish U2AF^{35} orthologues (amino acids 1–120). The accession numbers are as follows: *H. sapiens*, Q01081; *F. rubripes*, AAD13394; and *D. rerio*, NP_803432. U2AF^{35} sequences from *O. latipes* and *I. punctatus* were predicted from ESTs BJ527861 and CB940474, respectively; *F. rubripes* U2AF^{35}b was derived from the genomic sequence on scaff._79. Identical residues are *black type* on *white*, similar amino acids are highlighted in *gray,* and nonsimilar amino acids are *white type* on *black.* Peptide sequences coded by exons 3 and Ab are *boxed*.

TABLE III
Exon-intron organization of F. rubripes putative U2AF1 duplicate genes

U2AF1-a (scaffold_3186)				*U2AF1-b* (scaffold_79)			
Exon	Size	Intron	Size	Exon	Size	Intron	Size
	bp		*bp*		*bp*		*bp*
1	>44	1–2	620	1	>44	1–2	824
2	88	2–3	734	2	88	2-Ab	482
3	67	3–4	103	Ab	67	Ab-4	747
4	50	4–5	183	4	50	4–5	161
5	99	5–6	127	5	99	5–6	120
6	134	6–7	255	6	134	6–7	710
7	93	7–8	81	7	93	7–8	68
8	>118			8	>118		

pair may experience fixation of complementary loss-of-subfunction mutations, leading to gene preservations rather than gene loss (71). In this situation, both members may accumulate degenerative mutations in a way such that the function of the ancestral gene is partitioned between the duplicates ("subfunctionalization") (70, 71). One example is the gene that codes for the microphthalmia-associated transcription factor, MITF. In mammals and birds, at least four isoforms of *MITF* are generated by using alternative exons and promoters from a single gene, whereas in teleost fish two separate genes are present (72, 73). Each fish gene is thought to encode proteins corresponding to a particular bird/mammalian MITF isoform, be-

cause of the degeneration of alternative exons and regulatory elements (73). Another example is the duplicated *Fugu Syn2* gene in which the two isoforms generated by alternative splicing of the single human *SYN2* gene are encoded separately by the two duplicate *Fugu* genes (74).

Most probably, a mechanism of subfunctionalization has occurred in the two separate *Fugu U2AF1* genes, leading to the degeneration of the alternative exons. In fact, we were unable to detect either exon Ab sequences within the intron upstream of exon 3 in the *U2AF1-a* gene or exon 3 sequences within the intron downstream of exon Ab in the *U2AF1-b* gene.

The finding that both U2AF^{35}a and U2AF^{35}b isoforms have been under high selective pressure during evolution strongly suggests that each protein has important specific functions that are essential for vertebrates. This, together with the observation that the relative abundance of U2AF^{35}a and U2AF^{35}b mRNAs varies from organ to organ, raises the possibility that different concentrations of U2AF^{35} protein isoforms are available in each tissue. Considering that the function of U2AF^{35} is critical for achieving splicing of a subset of introns (17), we propose that tissue-specific regulation of *U2AF1* gene expression contributes to the finely tuned control of specific pre-mRNA splicing events in vertebrate organisms.

Acknowledgments—We thank Prof. M. R. Green (University of Massachusetts, Worcester, MA) for pGEX-hU2AF^{35} and pBluescript-βglobinU2AF^{65}; Dr. D. Henrique (University of Lisbon, Portugal) for the chicken embryo cDNA and the chicken β-actin probe; Prof. D. Wilkinson (National Institute for Medical Research, London, UK) for the chicken embryo cDNA library; and C. Martinez (EMBL, Heidelberg, Germany) for technical help.

REFERENCES

1. Jurica, M. S., and Moore, M. J. (2003) *Mol. Cell* **12,** 5–14
2. Maniatis, T., and Tasic, B. (2002) *Nature* **418,** 236–243
3. Brett, D., Pospisil, H., Valcarcel, J., Reich, J., and Bork, P. (2002) *Nat. Genet.* **30,** 29–30
4. Black, D. L. (2003) *Annu. Rev. Biochem.* **72,** 291–336
5. Lewis, B. P., Green, R. E., and Brenner, S. E. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 189–192
6. Graveley, B. R. (2002) *Cell* **109,** 409–412
7. Caceres, J. F., and Kornblihtt, A. R. (2002) *Trends Genet.* **18,** 186–193
8. Smith, C. W., and Valcarcel, J. (2000) *Trends Biochem. Sci.* **25,** 381–388
9. Roberts, G. C., and Smith, C. W. (2002) *Curr. Opin. Chem. Biol.* **6,** 375–383
10. Burge, C. B., Tuschl, T. H., and Sharp, P. A. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., eds) pp. 525–560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
11. Ruskin, B., Zamore, P. D., and Green, M. R. (1988) *Cell* **52,** 207–219
12. Zamore, P. D., and Green, M. R. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 9243–9247
13. Zamore, P. D., Patton, J. G., and Green, M. R. (1992) *Nature* **355,** 609–614
14. Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999) *Nature* **402,** 838–841
15. Zorio, D. A., and Blumenthal, T. (1999) *Nature* **402,** 835–838
16. Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999) *Nature* **402,** 832–835
17. Guth, S., Martinez, C., Gaur, R. K., and Valcarcel, J. (1999) *Mol. Cell. Biol.* **19,** 8263–8271
18. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) *Cell* **106,** 595–605
19. Zhang, M., Zamore, P. D., Carmo-Fonseca, M., Lamond, A. I., and Green, M. R. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 8769–8773
20. Blencowe, B. J. (2000) *Trends Biochem. Sci.* **25,** 106–110
21. Graveley, B. R. (2000) *RNA* (*New York*) **6,** 1197–1211
22. Graveley, B. R., Hertel, K. J., and Maniatis, T. (2001) *RNA* (*New York*) **7,** 806–818
23. Guth, S., Tange, T. O., Kellenberger, E., and Valcarcel, J. (2001) *Mol. Cell. Biol.* **21,** 7673–7681
24. Marchuk, D., Drumm, M., Saulino, A., and Collins, F. S. (1991) *Nucleic Acids Res.* **19,** 1154
25. Kaiser, K., Murray, N. E., and Whittaker, P. A. (1995) in *DNA Cloning: A Practical Approach* (Glover, D. M., and Hames, B. D., eds) Vol. I, pp. 37–82, IRL Press at Oxford University Press, Oxford
26. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, pp. 7.39–7.52, 9.31–9.59, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
27. Erfle, H., Ventzki, R., Voss, H., Rechmann, S., Benes, V., Stegemann, J., and Ansorge, W. (1997) *Nucleic Acids Res.* **25,** 2229–2230
28. Voss, H., Wiemann, S., Grothues, D., Sensen, C., Zimmermann, J., Schwager, C., Stegemann, J., Erfle, H., Rupp, T., and Ansorge, W. (1993) *BioTechniques* **15,** 714–721
29. Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A., and Smith, C. W. (2004) *Mol. Cell* **13,** 91–100
30. Wagner, E. J., and Garcia-Blanco, M. A. (2002) *Mol. Cell* **10,** 943–949
31. Mendell, J. T., ap Rhys, C. M., and Dietz, H. C. (2002) *Science* **298,** 419–422
32. Zuo, P., and Maniatis, T. (1996) *Genes Dev.* **10,** 1356–1368
33. Lin, Y.-S., and Green, M. R. (1991) *Cell* **64,** 971–981
34. Fleckner, J., Zhang, M., Valcarcel, J., and Green, M. R. (1997) *Genes Dev.* **11,** 1864–1872
35. Gama-Carvalho, M., Carvalho, M. P., Kehlenbach, A., Valcarcel, J., and Carmo-Fonseca, M. (2001) *J. Biol. Chem.* **276,** 13104–13112
36. Gama-Carvalho, M., Krauss, R. D., Chiang, L., Valcarcel, J., Green, M. R., and Carmo-Fonseca, M. (1997) *J. Cell Biol.* **137,** 975–987
37. Dignam, J., Lebovitz, R., and Roeder, R. (1983) *Nucleic Acids Res.* **11,** 1475–1489
38. Valcarcel, J., Martinez, C., and Green, M. R. (1997) in *mRNA Formation and Function* (Richter, J. D., ed) pp. 31–53, Academic Press, New York
39. Watakabe, A., Tanaka, K., and Shimura, Y. (1993) *Genes Dev.* **7,** 407–418
40. Zillmann, M., Zapp, M., and Berget, S. (1988) *Mol. Cell. Biol.* **8,** 814–821
41. Schmittgen, T. D., Teske, S., Vessela, R. L., True, L. D., and Zakrajsek, B. A. (2003) *Int. J. Cancer* **107,** 323–329
42. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402
43. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002) *Nucleic Acids Res.* **30,** 38–41
44. Birney, E., Thompson, J. D., and Gibson, T. J. (1996) *Nucleic Acids Res.* **24,** 2730–2739
45. Nei, M., and Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426
46. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882
47. Rudner, D. Z., Kanaar, R., Breger, K. S., and Rio, D. C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10333–10337
48. Wentz-Hunter, K., and Potashkin, J. (1996) *Nucleic Acids Res.* **24,** 1849–1854
49. Chen, C. Y., and Shyu, A. B. (1995) *Trends Biochem. Sci.* **20,** 465–470
50. Tassone, F., Villard, L., Clancy, K., and Gardiner, K. (1999) *Gene* (*Amst.*) **226,** 211–223
51. Buerstedde, J. M., and Takeda, S. (1991) *Cell* **67,** 179–188
52. Maquat, L. E. (2002) *Curr. Biol.* **12,** R196–R197
53. Schell, T., Kulozik, A. E., and Hentze, M. W. (2002) *Genome Biol.* **3,** Reviews 1006. Epub 2002 Feb. 26
54. Wilkinson, M. F., and Shyu, A. B. (2002) *Nat. Cell Biol.* **4,** E144–E147
55. Ohnishi, T., Yamashita, A., Kashima, I., Schell, T., Anders, K. R., Grimson, A., Hachiya, T., Hentze, M. W., Anderson, P., and Ohno, S. (2003) *Mol. Cell* **12,** 1187–1200
56. Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003) *Genome Res.* **13,** 382–390
57. Caceres, J. F., and Krainer, A. R. (1997) in *Eukaryotic mRNA Processing* (Krainer, A. R., ed) pp. 174–212, Oxford University Press, Oxford
58. Sureau, A., and Perbal, B. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 932–936
59. Popielarz, M., Cavaloc, Y., Mattei, M. G., Gattoni, R., and Stevenin, J. (1995) *J. Biol. Chem.* **270,** 17830–17835
60. Jumaa, H., and Nielsen, P. J. (1997) *EMBO J.* **16,** 5077–5085
61. Jumaa, H., Guenet, J. L., and Nielsen, P. J. (1997) *Mol. Cell. Biol.* **17,** 3116–3124
62. Stoilov, P., Daoud, R., Nayler, O., and Stamm, S. (2004) *Hum. Mol. Genet.* **13,** 509–524
63. Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. (2001) *EMBO J.* **20,** 1785–1796
64. Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stevenin, J., and Del Gatto-Konczak, F. (2001) *J. Biol. Chem.* **76,** 40638–40646
65. Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., Ho, R. K., Langeland, J., Prince, V., Wang, Y. L., Westerfield, M., Ekker, M., and Postlethwait, J. H. (1998) *Science* **282,** 1711–1714
66. Meyer, A., and Schartl, M. (1999) *Curr. Opin. Cell Biol.* **11,** 699–704
67. Taylor, J. S., Van de Peer, Y., Braasch, I., and Meyer, A. (2001) *Philos. Trans. R. Soc. Lond-Biol. Sci.* **356,** 1661–16679
68. Van de Peer, Y., Taylor, J. S., Braasch, I., and Meyer, A. (2001) *J. Mol. Evol.* **53,** 436–446
69. Wittbrodt, J., Meyer, A., and Schartl, M. (1998) *BioEssays* **20,** 511–512
70. Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., and Postlethwait, J. H. (1999) *Genetics* **151,** 1531–1545
71. Lynch, M., and Force, A. (2000) *Genetics* **154,** 459–473
72. Lister, J. A., Close, J., and Raible, D. W. (2001) *Dev. Biol.* **237,** 333–344
73. Altschmied, J., Delfgaauw, J., Wilde, B., Duschl, J., Bouneau, L., Volff, J.-N., and Schartl, M. (2002) *Genetics* **161,** 259–267
74. Yu, W.-P., Brenner, S., and Venkatesh, B. (2003) *Trends Genet.* **19,** 180–183