REPORT

U2AF³⁵ is encoded by an essential gene clustered in an operon with RRM/cyclophilin in *Caenorhabditis elegans*

DIEGO A.R. ZORIO^{1,2} and THOMAS BLUMENTHAL²

¹Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

²Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

ABSTRACT

In most species the 3' splice site is recognized initially by an interaction between the two-subunit splicing factor U2AF with the polypyrimidine (poly(Y)) tract that results in recruitment of the U2 snRNP to the branch-point consensus just upstream. In contrast, in *Caenorhabditis elegans*, both the poly(Y) tract and the branch-point consensus sequences are missing, apparently replaced by the highly conserved U₄CAG/R 3' splice site consensus. Nevertheless *C. elegans* U2AF⁶⁵ is very similar to its mammalian and fly counterparts and may recognize the 3' splice site consensus. Here we report the cloning of the *C. elegans* U2AF³⁵ gene, *uaf-2*. We show that it lacks an identifiable RS domain, which, in flies, has been shown to play a role in RNA binding, but it contains an extended glycine-rich stretch at its C-terminus. *uaf-2* is in an operon with *cyp-13*, a gene that encodes a cyclophilin with an RRM domain at its N-terminus. We demonstrate by RNA interference that both U2AF genes, *uaf-1* (which encodes U2AF⁶⁵) and *uaf-2*, are required for viability, whereas *cyp-13* is apparently not.

Keywords: pre-mRNA processing; RNA interference; RNA splicing; U2 snRNA auxiliary factor

INTRODUCTION

Nuclear pre-mRNA splicing involves a series of recognition events that occur at both ends of introns resulting in formation of a complex macromolecular particle, the spliceosome (Brody & Abelson, 1985; Frendewey & Keller, 1985; Grabowski et al., 1985; Cheng & Abelson, 1987). From mammals to yeast and plants, introns share little similarity other than the highly conserved sequences at the 5' and 3' ends. These sequences act as targets for splicing factors and serve as cleavage sites. The 5' splice site is a relatively short, 6 nt sequence highly conserved in a variety of organisms including *Caenorhabditis elegans* (Blumenthal & Steward, 1997). The signals at the 3' boundary are defined by three elements: the branch-point site (UACUAAC in yeast, YNYURAC in mammals), the 3' splice site (YAG), and a polypyrimidine tract, typically 15–20 nt long in vertebrates.

The first step of intron recognition involves the 5' splice site base pairing with the 5' end of U1 snRNA (Zhuang & Weiner, 1986; Ruby & Abelson, 1988; Siliaciano & Guthrie, 1988; Rosbash & Seraphin, 1991; Reich et al., 1992). The next step is polypyrimidine-tract recognition by the essential U2 auxiliary splicing factor, U2AF (Ruskin et al., 1988). U2AF binding to the poly(Y) tract facilitates U2 snRNA base pairing with the branch-site consensus (Parker et al., 1987; Wu & Manley, 1989; Zhuang & Weiner, 1989).

U2AF has been extensively studied because of its major role in determining the 3' splice site. Human U2AF is composed of two subunits, U2AF⁶⁵ and U2AF³⁵ (Zamore & Green, 1989). U2AF⁶⁵ has been found in vertebrates, *Drosophila*, *Schizosaccharomyces pombe*, *C. elegans*, and *Caenorhabditis briggsae* (Sailer et al., 1992; Zamore et al., 1992; Kanaar et al., 1993; Potashkin et al., 1993; Zorio et al., 1997). It contains three domains, a polypyrimidine-tract binding domain composed of three RNA recognition motifs (RRMs), an amino-terminal RS domain, and a domain that interacts with U2AF³⁵. U2AF³⁵ contains a pseudo-RRM (Birney et al., 1993), two Zn²⁺ binding motifs (Worthington et al., 1996), a glycine-rich region, and a carboxy-

Reprint requests to: Dr. Thomas Blumenthal, Department of Biochemistry and Molecular Genetics, University of Colorado Health Sciences Center (UCHSC), 4200 E. 9th Avenue, Box B121 Denver, Colorado 80262, USA; e-mail: tom.blumenthal@UCHSC.edu.

terminal RS domain. The latter is present in *Drosophila* and human U2AF³⁵ but not in *S. pombe* (Wentz-Hunter & Potashkin, 1996).

Although *C. elegans* introns share some features with mammalian and yeast introns, there are a number of unusual properties that are inherent to the nematode (Blumenthal & Steward, 1997). *C. elegans* introns tend to be quite short; they have a high A+U content compared with the surrounding exons; although they obey the GU-AG rule and share the mammalian 5' splice site consensus, the 3' end is somewhat different. *C. elegans* introns lack both the poly(Y) tract and the branch-site consensus. However, there is a highly conserved octamer, U₄CAG/R, that forms the 3' splice site (Blumenthal & Steward, 1997). All the bases in this sequence play a critical role in intron recognition (Conrad et al., 1993a; Zhang & Blumenthal 1996).

We have previously reported the cloning of *C. elegans* U2AF⁶⁵, which is quite similar to its mammalian and fly counterparts (Zorio et al., 1997), in spite of the fact that *C. elegans* introns lack a definable poly(Y) tract. To obtain more information about *C. elegans* U2AF, we have now cloned the nematode U2AF³⁵ gene, *uaf-2*. We show that it is clustered in an operon with a novel class of RRM/cyclophilin gene (Mi et al., 1996). *C. elegans* U2AF³⁵ has an extremely short RS sequence, but contains an extended glycine-rich stretch. We demonstrate that both U2AF genes are essential, which is in accord with the requirement for U2AF for viability in *Drosophila* and *S. pombe* (Kanaar et al., 1993; Potashkin et al., 1993; Rudner et al., 1996).

RESULTS AND DISCUSSION

The U2AF³⁵ gene and its chromosomal arrangement

We cloned the *C. elegans* U2AF³⁵ gene, designated *uaf-2*, by reverse transcription of polyadenylated RNA followed by two consecutive PCR reactions with degenerate oligonucleotides that were designed to conserved regions in the U2AF³⁵ genes from humans and flies (Zhang et al., 1992; Rudner et al., 1996). The PCR product was sequenced and found to encode a protein closely related to U2AF³⁵ from flies, humans, and *S. pombe* (data not shown). This cDNA was then used to probe a genomic library (kindly provided by Heidi Browning & Susan Strome). A 7.5-kb genomic clone containing *uaf-2* was sequenced and BLAST analysis was performed against the *C. elegans* genome data bank. The results indicated that *uaf-2*, composed of three exons and two introns, was clustered in the same 5'-to-3' orientation with a novel cyclophilin gene beginning with an RRM domain (Figs. 1A and 4). The intercistronic region between *uaf-2* and *cyp-13* is 107 bp, a distance typical of genes in *C. elegans* operons (Blumenthal & Steward, 1997). *cyp-13* contains three ex-

ons and two introns, one of them unusually long, 4,165 bp (Fig. 1B).

Analysis of *uaf-2* RNA and trans-splicing of *uaf-2* and *cyp-13*

On a Northern blot, only one transcript hybridized to a cDNA probe specific to *uaf-2*, indicating there is only one mRNA species (Fig. 2A). A developmental Northern revealed that the transcript is present in all stages and expressed at similar levels (data not shown). The 1.1-kb transcript contains a 3' untranslated region of ~130 nt verified by RT-PCR and by comparison with cDNA clones from the EST databank (Y. Kohara, pers. comm.).

Approximately one quarter of *C. elegans* genes are organized in operons (Zorio et al., 1994). Genes are clustered together and coexpressed from a single promoter, not unlike bacterial operons, to make polycistronic transcripts (Spieth et al., 1993). These polycistronic transcripts are processed by *trans*-splicing to SL1 or SL2 to give monocistronic messages. SL1 is added to the 5' ends of messages that start with an outtron, an intron-like sequence that lacks a functional 5' splice site (Conrad et al., 1993b). SL2, however, is specialized for *trans*-splicing to genes that are downstream within an operon (Spieth et al., 1993; Zorio et al., 1994). To test whether *uaf-2* and *cyp-13* genes are in an operon, we performed RT-PCR using gene-specific oligonucleotides located near the 5' ends of the cDNAs, combined with primers equivalent to the first 20 nt of SL1 or SL2. The results (Fig. 2B) indicate that *uaf-2* is *trans*-spliced only to SL1 whereas *cyp-13* receives mostly SL2 and a little SL1.

Although in most of the *C. elegans* operons described to date it has not been possible to discern a clear functional relationship between the gene products, in some cases that relationship has been established (Blumenthal, 1998). For example, *lin-15B* and *lin-15A* are genes in an operon that encode unrelated proteins involved in vulva development (Clark et al., 1994; Huang et al., 1994). Another example is the case of the *des-2* and *deg-3* operon (Treinin et al., 1998) in which the genes encode the two subunits of the acetylcholine receptor. In a third operon, expressed only in cuticle collagen-producing cells, the gene products, two unrelated enzymes, are needed for modifications of collagen that allow it to trimerize (Page, 1997).

uaf-2 encodes an essential splicing factor and is clustered with *cyp-13*, which encodes a novel class of cyclophilin (Mi et al., 1996). Cyclophilins are binding proteins that possess PPI (peptidyl-prolyl *cis-trans* isomerase) activity involved in protein folding and protein-protein interactions. *cyp-13* encodes a cyclophilin that contains an N-terminal RRM; its alignment with the human and *Schistosoma japonicum* homologs revealed strong conservation in the regions that confer PPIase

TGTAAAAATTCGATTTTTCACAACGTTATTTGGTAATTATGCGCAATAATTCGTGCTAAAC
 TCAAAATTTTCTGTAGAAAATTTGGCCAAAATATGCGAAATTTTGGAAAAATAACGAAGA
 GAGATAAAAATTTTATGTGATTTGGCCAAAAAAAACAATAGTTCTTCTATGCAATGCAT
 CTCGTTTCAGAGAAGAACGACATGTGAATTTGTTCTAAATTTTATATACATGACACCTACATC
 CAJAAAATTTATTTATTCGATCTCTAAATTTTGCACAAAATTTTGTGTGCACAGGAAGTGGGAC
 GGAGCTACGCAAAAATTTTCTCTCTTTTCGAAATGGGTAAATTTTTCATTCGTGCAAAAATTC
 GAGAAAAATTCAGTAAGAAAATGTAAATTTATTTAGTGGGGAAAACTCTCATGCGGCAGTGTG
 GTGAAGCTACGCTCGCCGCCACTTCTGTGGCGACAAATTTTTCGTGCAAAATTCAGAGGATC
 GATAGCAAAATTTACAGTACTCTGTAGAAATTCCTGAGAAGCTCCAAATTTGAAATGGGGCCC
 TGTGTCACAAATTTGATGTGCGTTGTGATAAAAATAAATCCAAACGCTGCAATTTCTCTCAAGA
 AAGGGCTCGCGGATCTCATGCTCTGTAGGCGGCTTTTCAATTTGGGATAATTCATAGTCTCC
 GGAATTTCCGGCTGTCAGGTTCTGTAGCGGATTAGTGTTATCTCTTCAATGATTTAGTCTCTC
 CACACATCTCAATTTGTGAAAATATAGCTTTTGTAGTACTTTGTGCTGTCTCAAGGAAGAA
 ATATCTTCACTCTCCCTGTGCGTGCATTAACCTAGCTGGCTCCAACTGCATGACATATAGACA
 CTTTGAACTACACCTCTAGATATGCTCAATTTCACTGGTTGTGATTCACATCTCTGCGACCT
 TCACTTCATCTCAATGTCCACCGGACTCGCAACTCTCTCTTCTTCTTCTGGAATTTGTTTTC
 ACCTGTAGAGATCTTCGTTGTGAAATTTGGATACGAAATGAGCAATCTGTGGAACATTTTGC
 AAACTCTGCTCTCAATAAGCTGTGAAATCAATGGAACATCACTGATACAGATCTCCGAGTTG
 TTTTTCACATTTTTCACGAAATCTGTGAACTCTCAATTTCCGGAAGATATCTTGTCTGTGAAT
 TCTCTGTGCTTTTCAGTTTCTTTCTCAATAGCTTTATCTCAAAATTTTGAATAATGAAATCTCT
 TATATGTCGATGTGCTTGTGATCGGCTCTCTGAGTCTGATTTTACGTGTGATCGAGAAATTTCCGG
 ACATCTTTTATGTTCTAGAAAAGTCTTCAGTGTTCATTAACAAAACTTTGATGTGTTT
 AATGATTTGTGAGAGAGTTAGTGTACGAGAAATTTGTTHAACTCTTCTCAAAATGTGAAACGA
 TCGATAACATTTATTCAGGTAAATGTGAAACAAAGTAAAGCTGTAGCGAAAAATTCGACATAA
 ACTTTATCTGGGAGACACAGCGGAGTCCAAAGCATGAGCTATCTCAAGCTCTTTCAGTATCT
 TCTCGAGAGTTTTCATAGTCCAGTGGAAATGATGAAGAATGATGTTCTCAAGTGTTCACAA
 CCACTCAGAAATACAGAGACTTCCGAGCATAGCTGCCAACTCTCTCAAAATTTTCAGGCG
 ATTTTATGTGGGCAATTTGATAGGCTTHAAGCTTTCTGCGAGTGTGTTCAACATTTCTGT
 TTAGCTCTCGAANCTCTCTCCAGCCCAAGGATCTCCGAGCGGTTAGAAATCGACTCGTGA
 AATCTCTTGTGTGATCATTTTTCAGCTTTGTACATCTTTGGAGCTCTCTCTCAGAAATCTG
 TTTCTCTCTTTCGAAAGCTGTGACATTTGTAGATGTCACTGTCTCAAAATTTCTCTCTTCT
 TCACTGTCTTCTAACATATGAATCTTTGTGATCATTTCACTAATCTGAACTTTTCTCGT
 ACGTGGCTTAAACGTGCAATCTTATGATTTGTGTTCTATCTCTCAAGTATTCACAGATTCGT
 ACGCTCGAGACTTTGAGACATTTATCTACTTAACTTTTTCACAAAAATTTGAGGACTCAAT
 GTAGCGCTGGTAAATATGTGAACGGCACATCTCAAGCGCTCTTGACGCCATCTGACAGATTTG
 CGGATGTATCTCAAGCATGTGACATGTGAACATGACGCCATTTGTGATGAGTAAATCTCTTGACA
 TCTAGTCTTAAGTTCAGGAAGGTGTCAAAAACTTATTTTCTCATGTAGGTTAGTCTCAAGTAT
 TTCCCATCAAAATTTCTCAATTTCTCATAGTACGCTTATATCTGATTTGATTTGGCTCTCCACT
 TTTCAAATCTCTTCTCAGATCTGCAATGTATGTAGCTCTCAAGTATGATCAAAATTTGAGA
 ATATTTTTCAGATTTGAAAATTTGGAATAATCAGAAATTTCTATAGCTATCTCAAAATTTCTCAT
 TCTGCAATAGTCTCACTTCATATATATCCAGCTTGAAATTTCAATCTCTTTGAGTGTAGCA
 ACTGTCGAGATACGCTGTGAGATGTAATTTTGAATGGGAAATGACGCTGTGCTGAAATTTTGAAT
 TTTCTCAAGATTTAATTTTGAAGACGCTGTGAAATCGAAACAGGATATTTGTGGCTCAAAA
 TCCACAGAGCTGTATCTTTCTCAGTTTGTATTTGTACAAAAATGAGAACACCGAGAGGTTTATA
 CTCTATGTACAGTAGAGCGCCCTTTTATAGGAAGCTGAGTTTTCGCGCCGCTCTACAGAGCG
 GTTTTATCTACGTGGCGGTAAAAATGAGAAAATTCGTGGGCTCTATACAAATGGGTTGCGA
 ACATTTTCTCAAAAATTCGCCACGTGGAACTATCTATGCGGCAATAAATCTGTGCGAGTT
 TTGATTTTCTTAGGCCACAGTAGTAAAGCGCGCTCTTTCGCGGTGTCTCAAAATTCAGCGTT
 TGTGCAAAAATTTTCTTATTTTGTGTCACTAAATTTTCTTCAAAAATTTTGACAAATTT
 TCACTAAATCTGTACACCGTACGCGGCTTTTGTGTCAGTTTGTAGCCATTTTGTCTTTACT
 TTGCTATTAATTTGTTATCTCGAAAAAATAACAACTTTCTTATAGAAATTTGCCAAATCTCGG
 TTGTCTCGAAATACGTGTGTCTCAAACTATATCCGCAAGTTTGTGCAAAAATTTGCTTATTT
 GTATCTTTCTCCCGAAAAAATCCCAAGTTTCTTATCAACATTTAGAGAAAAAATTAAT
 AAAATTTGTGTGCGAGTACCTGCGAGTACGATGATCTGAAATTTTCCCACTATTTTCCGAAAA
 AACCGGTTATTTGATTTGTTAGAGCTGAAATATCAATATAAAAAATGTGAAATTTGGCGG
 GATGAGTGTGAAATTTGAAATTTGTAGAGCTGAAATATCAATATAAAAAATGTGAAATTTGGCGG
 AAAATAAAAAACCCAGCGATAGAGAAATCAAGAACTATTTTGTAGCCGATTTGAGTATAGC
 TCGAAAAATCGATGTAGTTTGTGGTTTTTGTGAAGTTTCAAAATTTTCAAAAATTTTATTA
 ATTTGGAAATTAAGTTTATTTTTTTTTTTTTAAATGTTTGTAAAAATTTTACAGTGTCC
 CGAGAGTTTAAAAAATTTGGAATTTCTGCTGTCAACTACGACTTAAATTTGATTTTCTGTA
 GTATTTGTGCGCAACAATAGGAAAAATTTTCAAAATGTCAACTATTTTAAAGCAATTTTGTACT
 CAAAAACTCTCTCATCAAAACGCGAAAAAGTCAAAATTTTCTCTTTTAACTCAAAATTT
 TTGTGCAATCTTTATCCAGAGCTGAAAATTTGTCTGTGAGACGAGTTTGTGCACTAAAA
 TCTGGATTTGAAATGCAATTTTATCAGTTTTTGTCAATAAAAACCCATATATTTTCCACTTAA
 AAACTCTTAAATTTTCAAAATTTTCTGGTTTTTAACTAAAAATTTGCAAAATTTTACGACAA
 TTTTCACTCAAAAGCTGCAATTTTCCATCAAAACGCGTAAAAATTCGAAATTTTCTCTTT
 TTTTCTCGGATTTTGTATCTTAAATTTTCAAAATTTTGTGATCTCTCTTTCAAAACGCG
 TGTAAATTTGAAATTTGTCTTTTGTAGACGATTTTGTGATCTTAAATTTCTGATTTTAAAG
 CATTTTTCACATTAAGCTCTTAAAAATTTGAAATTTTCCAGTTCTCGGCAACAATTTGCC
 L R H T M P
 GGCACAGTTTCAATGGCGCACTCGGTCGCAACCAAAATGGCTCACAATCTTCAATTTGC
 G T V S H A N G C A A N T N G S Q F F I C
 ACCGAAAAACCGAATTTGGCTGTGACGAAAAACAGTAGTGTCTGGCTCATGTTGTGGAAGG
 T E K T D W L D G K H V F G H V E G
 ATGAATTTGTATACGACAGTGTGAACAGGACGAAGCACTCGGTGTGGCAACCTTCAGTGTGT
 M N I V R Q V E Q G Q T P S G K R Q M V
 GTGAAATCTGTGGAAGAGCGGGAAAAATGAGCGGAAAAACCGATTTGCGCGCGAAAAATTTG
 V K I E S E G I E P E K R I A A E K L
 GCCCAAGAGCGCATTTGTCTCGGAGCGGAGATTCAGGAGCGCGCTTCCGCGAGATTTGGAG
 A Q K A V P V G A E I Q E P L P Q A M E
 ACTTAATTTATGAAATTTTAAAAATTTATATGTTTATTTTCAAAAAGAAATTTGTTTAATT
 L R H T M P
 GGCACAGTTTCAATGGCGCACTCGGTCGCAACCAAAATGGCTCACAATCTTCAATTTGC
 G T V S H A N G C A A N T N G S Q F F I C
 ACCGAAAAACCGAATTTGGCTGTGACGAAAAACAGTAGTGTCTGGCTCATGTTGTGGAAGG
 T E K T D W L D G K H V F G H V E G
 ATGAATTTGTATACGACAGTGTGAACAGGACGAAGCACTCGGTGTGGCAACCTTCAGTGTGT
 M N I V R Q V E Q G Q T P S G K R Q M V
 GTGAAATCTGTGGAAGAGCGGGAAAAATGAGCGGAAAAACCGATTTGCGCGCGAAAAATTTG
 V K I E S E G I E P E K R I A A E K L
 GCCCAAGAGCGCATTTGTCTCGGAGCGGAGATTCAGGAGCGCGCTTCCGCGAGATTTGGAG
 A Q K A V P V G A E I Q E P L P Q A M E
 ACTTAATTTATGAAATTTTAAAAATTTATATGTTTATTTTCAAAAAGAAATTTGTTTAATT
 L R H T M P

FIGURE 2. Analysis of *uaf-2* RNA. **A:** Northern blot analysis of RNA isolated from *C. elegans*. Poly (A)⁺ RNA was electrophoresed, transferred to a nylon membrane, and probed with *uaf-2* cDNA sequence made by RT-PCR. Only one transcript of about 1.1 kb is seen. Marker size in kilobases. **B:** RT-PCR analysis of *uaf-2* and *cyp-13* mRNA 5' ends. Poly (A)⁺ RNA from *C. elegans* was reverse transcribed with a gene-specific oligonucleotide complementary to sequences near the 5' end of each coding sequence and PCR amplified with these oligonucleotides and either SL1-20 or SL2-20. Products were electrophoresed, transferred to a nylon membrane, and probed with specific internal oligonucleotides complementary to the 5' end of each gene.

activity and in the RRM (Fig. 3). This new class of cyclophilins was first reported in the nucleus of human T-cells and is the first example of RNA-binding and protein-folding domains in a single polypeptide (Mi et al., 1996). One attractive possibility is that the *uaf-2/cyp-13* operon might constitute another example of an operon whose gene products participate in the same pathway. U2AF³⁵ is a known splicing factor that plays a critical role in splicing (Zuo & Maniatis, 1996; Rudner et al., 1998b). *cyp-13* could also be involved in splicing, acting as a chaperone by promoting folding of proteins that participate in splicing or even RNA-protein complexes formed during the process of intron removal. However, the fact that *cyp-13* is not required for viability (see below) suggests that if it does provide a critical function in splicing, there must be a redundant gene elsewhere in the genome capable of supplying *cyp-13* function. Alternatively, it could contribute to splicing efficiency or accuracy without being required for viability.

RNA-mediated interference

It has recently been shown that injection of RNA into *C. elegans* is capable of interfering with expression of the gene to which the RNA is homologous, producing phenotypes similar to a null mutation in that gene (Guo & Kemphues, 1995; Fire et al., 1998). However, the mechanism of how RNA-mediated interference occurs is still unknown. For example, either RNA strand, sense or antisense, is capable of producing a similar effect when injected into *C. elegans* (Guo & Kemphues, 1995; Fire et al., 1998). In addition, it has been reported that when compared to only one strand, a mixture of both RNA strands was substantially more effective at producing interference (Fire et al., 1998). These results suggest

that more than a simple base pairing mechanism must be required. To test whether U2AF is essential in *C. elegans*, we performed RNA interference experiments. In vitro-synthesized RNA using either T7 or SP6 RNA polymerases was used to produce sense or antisense strands of *uaf-1*, *uaf-2*, or *cyp-13*. These RNAs were microinjected into the gonads of young adult *C. elegans* hermaphrodites. Purified sense or antisense RNA fragments covering 871 nt of *uaf-1*, 452 nt of *uaf-2*, or 1,100 nt of *cyp-13* were injected at 1 mg/mL. As shown in Table 1, sense *uaf-1* and *uaf-2* strands were highly effective, causing a severe phenotype, embryonic lethality (88% and 100%, respectively). However the same was not observed when the sense strand of *cyp-13* was injected; it had no apparent effect. To test whether the lack of phenotype with *cyp-13* was due to inefficient interference by injecting single-stranded RNA (Fire et al., 1998), sense and antisense RNA were mixed prior to injection at a final concentration of 200 µg/mL. This double-stranded RNA injection also had no apparent phenotype. These results indicate that the U2AF genes in *C. elegans*, *uaf-1* and *uaf-2*, are required for viability whereas *cyp-13* is not and are consistent with the U2AF requirement for viability found in *Drosophila melanogaster* (Kanaar et al., 1993; Rudner et al., 1996) and *S. pombe* (Potashkin et al., 1993).

Comparison of *C. elegans* U2AF³⁵ with homologs from other organisms

Figure 4 shows an alignment of U2AF³⁵ from *C. elegans* with mammalian, fly, and fission-yeast homologs. The known functional domains of this protein are present and conserved, but with a few differences. The carboxy-terminus of the fly and *C. elegans* protein are

<i>H. sapiens</i>	1	M A T T - - - - K R V L Y V G G L A E E V D D K V L H A A F I P
<i>S. japonicum</i>	1	- -
<i>C. elegans</i>	1	M N T N F P H N R K R T L Y V G G F T E D V T E K V L M A A F I P
<i>H. sapiens</i>	29	F G D I T D I Q I P L D Y E T E K H R G E A E V E E E L A E D A A
<i>S. japonicum</i>	1	- - - - - - - M P M D Y Q T E K H R G E A E V E E E E V E D A M
<i>C. elegans</i>	34	F G D V V A I S I P M D Y E S G K H R G E G E V E F D M A E D A A
<i>H. sapiens</i>	62	A A I D N M N E S E L F G R T I R V N L A K P M R I K E G S S R P
<i>S. japonicum</i>	26	S A I D N M N E S E I F G R T I R V N V A R P V R I R E G W S R P
<i>C. elegans</i>	67	M A I D N M N E S E L F G K T I R V N F A R P P K A T E R S Q K P
<i>H. sapiens</i>	95	V W S D D D W L K K F S G K T L E E N K E E E G S E P P K A E T Q
<i>S. japonicum</i>	59	V W S D E N W L K K Y - G S A P L E G R K L D E P D I - V N P S D
<i>C. elegans</i>	100	V W A D D E W L K K Y - G R G G E A A A E E D G - D A E K A A T S
<i>H. sapiens</i>	128	E G E P I A K - - - - - - - - - - K A R S N P Q V Y M D I K I G N
<i>S. japonicum</i>	90	T S E N V E D L S D E E M R T K K Q K R N L P R V F F D I R I G N
<i>C. elegans</i>	131	S S S A S T K - - - - - - - - - - L P R V Y L G V K I G I
<i>H. sapiens</i>	151	K P A G R I Q M L L R S D V V P M T A E N F R C L C T H E K G F G
<i>S. japonicum</i>	123	G D A G R I V M E L R S D I V P R T A E N F R A L C T G E R G F G
<i>C. elegans</i>	150	R Y I G R I V I E L R T D V T P K T A E N F R C L C T G E R G F G
<i>H. sapiens</i>	184	F K G S S F H R I I P Q F M C Q G G D F T N H N G T G G K S I Y G
<i>S. japonicum</i>	156	Y H N C C F H R V I P Q F M C Q G G D F V K G D G T G G K S I Y G
<i>C. elegans</i>	183	Y E G S I F H R I I P K F M L Q G G D F T K G D G T G G K S I Y G
<i>H. sapiens</i>	217	K K F D D E N F I L K H T G P G L L S M A N S G P N T N G S Q F F
<i>S. japonicum</i>	189	R K F D D E N F Q L R H E G F G V L S M A N S G P N T N G S Q F F
<i>C. elegans</i>	216	T K F D D E N F T L R H T M P G T V S M A N C G A N T N G S Q F F
<i>H. sapiens</i>	250	L T C D K T D W L D G K H V V F G E V T E G L D V L R Q I E A Q G
<i>S. japonicum</i>	222	I C T T K C D W L D G K H V V F G R V V D G Q N V V K K M E S V G
<i>C. elegans</i>	249	I C T E K T D W L D G K H V V F G H V V E G M N I V R Q V E Q Q G
<i>H. sapiens</i>	283	S K D G K P K Q K V - - I I A D C G E Y V
<i>S. japonicum</i>	255	S K S G K V K E P V - - I I S R C G E L I
<i>C. elegans</i>	282	T P S G K P - Q M V V K I V - E S G E I E P E K R I A A E K L A Q
<i>C. elegans</i>	313	K A V V P G A E I Q E P L P Q A M E T

FIGURE 3. Alignment of the predicted amino acid sequence of *C. elegans* CyP-13 with the novel human nuclear cyclophilin (hCyP33) and *S. japonicum* homologs. The multiple alignment was produced by the Clustal program with minor adjustments. The RRM and PPI (Peptidyl-prolyl *cis-trans* isomerase) domains are underlined and double-underlined, respectively. The RNP-1 and RNP-2 domains are within the RRM domain, boxed and boxed plus underlined, respectively. If two amino acids or more at a particular position are identical they are shaded.

somewhat longer than their mammalian and fission-yeast counterparts. There is a glycine-rich stretch that is present in all but the fission-yeast U2AF small subunit. In *C. elegans* this glycine-rich region is quite long;

it contains 42 glycines including one stretch of 27 amino acids that is 85% glycine. In contrast, the human and fly proteins contain glycine-rich domains with only 13 and 22 total glycines respectively (Fig. 4). The highly

TABLE 1. Effects of RNA-mediated interference.

Gene	Injected RNA	Size (nt)	Eggs		Embryonic lethality
			Laid	Hatched	
<i>uaf-1</i>	sense	871	301	36	88%
<i>uaf-2</i>	sense	452	479	0	100%
<i>cyp13</i>	antisense	1,100	446	446	0%
	sense + antisense		1,903	1,892	.006%

degenerate RRM, or pseudo-RRM (Birney et al., 1993), is present and very highly conserved among all genera. The RS domain is not present in the fission-yeast protein and in the worm homolog it is much shorter, with only two RS dipeptides when compared with seven in the fly and human counterparts (Fig. 4). The differences in length of the RS domain could contribute to the differences in 3' splice site selection. It has been reported in *Drosophila* that the presence of the small

<i>C. elegans</i>	1	MS YGDGLSGAEYLASIYGTEKDKVNCSEFFFKTG
<i>H. sapiens</i>	1	- - - - - MAEYLASIFGTEKDKVNCSEFFFKIG
<i>Drosophila</i>	1	- - - - - MAEYLASIFGTEKDKVNCSEFFFKIG
<i>S. pombe</i>	1	- - - - - MASHLASIYGTEQDKVNCSEFFFKIG
<i>C. elegans</i>	34	ACRHGDKCSRAHHTPTFSPITVVLKNFYHNPVVD
<i>H. sapiens</i>	26	ACRHGDRCSRLHNKPTFSOTIALLNLYRNPNONS
<i>Drosophila</i>	26	ACRHGDRCSRIHNKPTFSOTVLLQNLVYNPNONS
<i>S. pombe</i>	26	ACRHGERCSRKHVKNPFSQITILCPNMYKNPIHE
<i>C. elegans</i>	67	VROADAFDKVGKRNDQE - ORYFDDFYEEVFVEM
<i>H. sapiens</i>	59	SQSADGLRC - A - VSDVEMOEHYDEFFEEVFTEM
<i>Drosophila</i>	59	AKSADGSDLVANVSDEEMQEHYDNFFEDVFVEC
<i>S. pombe</i>	59	PNGKK - - - - - FTQRELAEQFDAFYEDMFCF
<i>C. elegans</i>	99	ERKYGEVEEINVCENIGEHMVGNVYVKFMKEED
<i>H. sapiens</i>	90	EKYEVEEEMNVCDNLGDHLVGNVYVKFRREED
<i>Drosophila</i>	92	EDKYGEIEEMNVCDNLGDHLVGNVYIKFRNEAD
<i>S. pombe</i>	85	- SKYGEVEQLVVCNVDGDLVGNVYVRFKYEES
<i>C. elegans</i>	132	AEKAKNDLNNRWFGQPIYAELCPVTDFRESRC
<i>H. sapiens</i>	123	AEKAVIDLNNRWFGQPIHAELSPVTDFREACC
<i>Drosophila</i>	125	AEKAANDLNNRWFGGRPVYSELSPVTDFREACC
<i>S. pombe</i>	117	AONAIDDLNSRWYSORPVYAELSPVTDFREACC
<i>C. elegans</i>	165	ROHEVTTCSKGRFCNFMHLKAISAE LGDRLYGR
<i>H. sapiens</i>	156	ROYEMGECTRGGFCNFMHLKPI SRELRLYGR
<i>Drosophila</i>	158	ROYEMGECTRSGFCNFMHLKPI SRELRRLYGR
<i>S. pombe</i>	150	RQHETSECQRGGLCNFMHAKKPSPQLLRDLVLA
<i>C. elegans</i>	198	RGRRADAAGHYPSORGGSGGGGGGGGGGGYSG
<i>H. sapiens</i>	189	RRKKHRSRSRS - RERR - SRSRDRG - - - -
<i>Drosophila</i>	191	RRR - ARSRSRSPGRRRGSRSRSRSPGRRGGGRG
<i>S. pombe</i>	183	QRKYLALNAAEEMKKEPNSDSTNRWVS VTAERK
<i>C. elegans</i>	231	GGWGGGGGGGKNRDRGGWGGRGGRGYSGGGGG
<i>H. sapiens</i>	211	- - - - - - - - - - - - - - - RGGGGG
<i>Drosophila</i>	223	DGVGGGN YLNNERDNMRGNDRGNDRDRRKGGGG
<i>S. pombe</i>	216	N
<i>C. elegans</i>	264	GYRYGRGRGGKRRSRSRDR - R - - RY
<i>H. sapiens</i>	217	GGGGGGGRERDRR - RSRDRERSGRF
<i>Drosophila</i>	256	GGGGGGGRY

FIGURE 4. Comparison of the predicted amino acid sequence of *C. elegans* U2AF³⁵ with vertebrate, *Drosophila*, and *S. pombe* homologs. The highly degenerate pseudo-RRM, which is reminiscent of β_3 -loop 4- α_2 region of the model RRM, is double underlined (Birney et al., 1993). The multiple alignment shown was produced by the Clustal program with minor adjustments. Identical amino acids are shaded. RS and related dipeptides are boxed.

subunit results in a 20-fold increase in binding-affinity for a pyrimidine tract in vitro (Rudner et al., 1998a). In addition, at least one RS domain in the heterodimeric U2AF is required for viability, suggesting a critical role in splicing in vivo (Rudner et al., 1998b). In *C. elegans*, it is not yet known whether U2AF³⁵ contributes to binding specificity or avidity, but, based on the sequence, it seems unlikely that the U2AF³⁵ RS domain plays a significant role as it does in flies.

CONCLUSIONS

Even though splicing in *C. elegans* differs in several significant respects (e.g., the existence of *trans*-splicing and polycistronic pre-mRNA splicing) from other animals, the basic splicing machinery found in other systems is nevertheless present. However it may perform its functions somewhat differently. Although *C. elegans* 3' splice sites lack consensus sequences for poly(Y) tracts and branch points, they have a highly conserved U₄CAG/R at the site of splicing. Here we have shown that both subunits of the protein responsible for 3' splice site recognition in other animals, U2AF, are present and required for viability in *C. elegans*. The availability of these clones should allow determination of whether U2AF is responsible for recognition of the unusual 3' splice site consensus of nematodes.

MATERIALS AND METHODS

Cloning of *C. elegans* U2AF small subunit gene

Oligonucleotides UAFS-1, 5'-GARAARGAYAARGTYAAYTG-3', complementary to the region encoding amino acids 12 and 18 and UAFS-4, 5'-RTGCATRAARTTRCARAANCC-3' complementary to the region encoding amino acids 167 and 173 of human U2AF³⁵ were used for RT-PCR. Reverse transcription was carried out as described in Zorio et al. (1997) using 1 µg of poly-A(+) RNA from a mixed population and UAFS-4 for the first-strand DNA synthesis. A small aliquot was then used as a template for back-to-back PCR reactions with 30 rounds of amplification each using UAFS-1 and UAFS-4, with an annealing temperature of 45 °C. The PCR product was gel purified, subcloned into pGEMT (Promega) and sequenced. The same fragment was random prime labeled (Prime it-II, Stratagene) and used to screen a *C. elegans* genomic library (kindly provided by Heidi Browning and Susan Strome) constructed in λgt11. Approximately 320,000 plaques were screened to yield one hybridizing plaque that was further purified and analyzed by restriction digests and ABI sequencing. The partial sequence was analyzed by BLAST against the *C. elegans* genome databank and the entire genomic sequence containing *uaf-2* is shown in Figure 1B.

Northern analysis

Poly-A(+) RNA was isolated from wild type (N2) mixed stage worm population using PolyAT tract mRNA Isolation System

IV (Promega). RNA was electrophoresed on an agarose-formaldehyde gel, transferred to Hybond-N membrane (Amersham), hybridized and washed (Zorio et al., 1997).

Trans-splicing specificity of *uaf-2* and *cyp-13*

Reverse transcription was performed with poly-A(+) mRNA as a template using gene-specific oligonucleotides (UAFS-6 5'-GCATTTGTCTCCGTGGCG-3' and dsU2AF35-1(-) 5'-GGCAACGACGTCTCCAAACG-3') for the first-strand DNA synthesis. PCR was performed using the same gene-specific downstream oligos and SL1-20 or SL2-20 as upstream primers. The products were separated on 10% denaturing polyacrylamide gels, transferred by electroblotting to Hybond-N (Amersham), and hybridized with ³²P end-labeled UAFS-5 (5'-CCTGTCTTGAAGAAAAACG-3') for *uaf-2* or dsU2AF35-1(+) (5'-CGAATTTTCCACACAATCGG-3') for *cyp-13* (Zorio et al., 1994).

RNA synthesis and microinjections

RNA was transcribed from plasmids containing *uaf-1*, *uaf-2*, and *cyp-13* by using T7 and SP6 RNA polymerases (Ampliscribe, Epicentre). RNAs were analyzed by gel electrophoresis and their concentration measured spectrophotometrically. Antisense or sense RNAs were injected individually at 1 µg/µL. Double-stranded *cyp-13* RNA was prepared as above; opposite strands were mixed prior to microinjections to a final concentration of 200 ng/µL (100 ng/µL of each strand). Gonads of six to ten wild-type (N2 Bristol) young adult hermaphrodite worms were injected with each RNA sample. After 12 h, injected animals were transferred to fresh plates and embryonic lethality was scored.

ACKNOWLEDGMENTS

We are grateful to Roland Kanaar and Don Rio for sharing the *Drosophila* U2AF³⁵ sequence with us prior to publication. We thank Peg MacMorris, Donald Evans and Ravinder Singh for critical review of this manuscript. The Cancer Center Core Facility at the University of Colorado Health Sciences Center provided excellent technical support in the ABI sequencing. This work was supported by research grant GM42432 from the National Institute of General Medical Sciences.

Received December 10, 1998; returned for revision January 6, 1999; revised manuscript received January 11, 1999

REFERENCES

- Blumenthal T. 1998. Gene clusters and polycistronic transcription in eukaryotes. *BioEssays* 20:480-487.
- Blumenthal T, Steward K. 1997. RNA processing and gene structure. In: Riddle D, Blumenthal T, Meyer B, Priess J, eds. *C. elegans II*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 117-145.
- Birney E, Kumar S, Krainer AR. 1993. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 25:5803-5816.
- Brody E, Abelson J. 1985. The "Spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* 288:963-967.

- Cheng S-C, Abelson J. 1987. Spliceosome assembly in yeast. *Genes & Dev* 1:1014–1027.
- Conrad R, Liou RF, Blumenthal T. 1993a. Functional analysis of a *C. elegans* trans splice site acceptor. *Nucleic Acids Res* 21:913–919.
- Conrad R, Liou RF, Blumenthal T. 1993b. Conversion of a trans-spliced *C. elegans* gene into a conventional gene by introduction of a splice donor site. *EMBO J* 12:1249–1255.
- Clark SG, Lu X, Horvitz HR. 1994. The *Caenorhabditis elegans* locus lin-15, a negative regulator of tyrosine kinase signaling pathway, encodes two different proteins. *Genetics* 137:987–997.
- Fire A, Xu S-Q, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811.
- Frendewey D, Keller W. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* 42:355–367.
- Grabowski PJ, Seiler SR, Sharp PA. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursor. *Cell* 42:345–353.
- Guo S, Kempthues K. 1995. *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* 81:611–620.
- Huang LS, Tzou P, Sternberg PW. 1994. The lin-15 locus encodes two negative regulators of *Caenorhabditis elegans* vulval development. *Mol Biol Cell* 5:395–412.
- Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. 1993. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: Requirement for viability. *Science* 262:569–573.
- Mi H, Kops O, Zimmermann E, Jaschke A, Tropschug M. 1996. A nuclear RNA-binding cyclophilin in human T cells. *FEBS Lett* 398:210–205.
- Page AP. 1997. Cyclophilin and protein isomerase genes are cotranscribed in a functionally related manner in *Caenorhabditis elegans*. *DNA Cell Biol* 16:1335–1343.
- Parker R, Siliciano PG, Guthrie C. 1987. Recognition of the TATCTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49:229–239.
- Potashkin J, Naik K, Wentz-Hunter K. 1993. U2AF homolog required for splicing in vivo. *Science* 262:573–575.
- Reich CI, VanHoy RW, Porter GL, Wise JA. 1992. Mutations at the 3' splice site can be suppressed by compensatory mutations in U1 snRNA in fission yeast. *Cell* 69:1159–1169.
- Rosbash M, Seraphin B. 1991. Who's on first? The U1 snRNP-5' splice site interaction and splicing. *Trends Biochem Sci* 16:187–190.
- Ruby SW, Abelson J. 1988. An early hierarchic role of U1 small ribonucleoprotein in spliceosome assembly. *Science* 242:1028–1035.
- Rudner DZ, Kanaar R, Breger KS, Rio DC. 1996. Mutations in the small subunit of the *Drosophila* U2AF splicing factor cause lethality and developmental defects. *Proc Natl Acad Sci USA* 93:10333–10337.
- Rudner DZ, Breger KS, Kanaar R, Adams MD, Rio DC. 1998a. RNA binding activity of heterodimeric splicing factor U2AF: At least one RS domain is required for high affinity binding. *Mol Cell Biol* 18:4004–4011.
- Rudner DZ, Kanaar R, Breger KS, Rio DC. 1998b. Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo. *Mol Cell Biol* 18:1765–1773.
- Ruskin B, Zamore PD, Green MR. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52:207–219.
- Sailer A, MacDonald NJ, Weissmann C. 1992. Cloning and sequencing of the murine homologue of the human splicing factor U2AF⁶⁵. *Nucleic Acids Res* 20:2374.
- Siliciano PG, Guthrie C. 1988. 5' splice site selection in yeast: Genetic alterations in base pairing with U1 reveal additional requirements. *Genes & Dev* 2:1258–1267.
- Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T. 1993. Operons in *C. elegans*: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. *Cell* 73:5121–5132.
- Treinin M, Gillo B, Liebman L, Chalfie M. 1998. Two functionally dependent acetylcholine subunits are encoded in a single *C. elegans* operon. *Proc Natl Acad Sci USA* 95:15492–15495.
- Wentz-Hunter K, Potashkin J. 1996. The small subunit of the splicing factor U2AF is conserved in fission yeast. *Nucleic Acids Res* 24:1849–1854.
- Worthington MT, Amann BT, Nathans D, Berg JM. 1996. Metal binding properties and secondary structure of the zinc-binding domain of Nup475. *Proc Natl Acad Sci USA* 93:13754–13759.
- Wu J, Manley JL. 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base-pairing. *Genes & Dev* 3:1553–1561.
- Zamore PD, Green MR. 1989. Identification, purification and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 86:9243–9247.
- Zamore PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355:609–614.
- Zhang H-B, Blumenthal T. 1996. Functional analysis of an intron 3' splice site in *Caenorhabditis elegans*. *RNA* 2:380–388.
- Zhang M, Zamore PD, Carmo-Fonseca M, Lamond AI, Green MR. 1992. Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci USA* 89:8769–8773.
- Zhuang Y, Weiner AM. 1986. A compensatory base change in U1snRNA suppresses a 5' splice site mutation. *Cell* 46:827–835.
- Zhuang Y, Weiner AM. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes & Dev* 3:1545–1552.
- Zorio DAR, Cheng NN, Blumenthal T, Spieth J. 1994. Operons as a common form of chromosomal organization in *C. elegans*. *Nature* 372:270–272.
- Zorio DAR, Lea K, Blumenthal T. 1997. Cloning of *Caenorhabditis* U2AF⁶⁵: An alternatively spliced RNA containing a novel exon. *Mol Cell Biol* 17:946–953.
- Zuo P, Maniatis T. 1996. The splicing factor U2AF³⁵ mediates protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes & Dev* 10:1356–1368.