4510–4519 Nucleic Acids Research, 1994, Vol. 22, No. 21

SWAP pre-mRNA splicing regulators are a novel, ancient protein family sharing a highly conserved sequence motif with the *prp21* family of constitutive splicing proteins

Deborah A.Spikes, Joseph Kramer, Paul M.Bingham* and Kevin Van Doren[1]
Department of Biochemistry and Cell Biology, 450 Life Sciences Building, University of New York, Stony Brook, NY 11794 and [1]Department of Biology, Syracuse University, Syracuse, NY 13244, USA

Received May 12, 1994; Revised and Accepted August 22, 1994 EMBL accession nos U06933 and U09415

ABSTRACT

Regulators responsible for the pervasive, nonsex-specific alternative pre-mRNA splicing characteristic of metazoans are almost entirely unknown or uncertain. We describe here a novel family of splicing regulators present throughout metazoans. Specifically, we analyze two nematode (*Caenorhabditis elegans*) genes. One, *CeSWAP*, is a cognate of the *suppressor-of-white-apricot* (*DmSWAP*) splicing regulator from the arthropod *Drosophila*. Our results define the ancient, conserved SWAP protein family whose members share a colinearly arrayed series of novel sequence motifs. Further, we describe evidence that the *CeSWAP* protein autoregulates its levels by feedback control of splicing of its own pre-mRNA analogously to the *DmSWAP* protein and as expected of a splicing regulator. The second nematode gene, *Ceprp21*, encodes an abundant nuclear cognate of the constitutive yeast splicing protein, *prp21*, on the basis of several lines of evidence. Our analysis defines *prp21* as a second novel, ancient protein family. One of the motifs conserved in *prp21* proteins—designated *surp*—is shared with SWAP proteins. Several lines of evidence indicate that both new families of *surp*-containing proteins act at the same (or very similar) step in early prespliceosome assembly. We discuss implications of our results for regulated metazoan pre-mRNA splicing.

INTRODUCTION

Regulated alternative pre-mRNA splicing is well documented in eukaryotes (8,16,34,35,37,43,44 and references therein). This process is ancient and profoundly pervasive, particularly in multicellular animals (metazoans). However, understanding of regulated splicing is severely limited and the number of currently known or suspected regulators of such processes is very small.

Several constitutive metazoan splicing proteins — the SR family and hnRNPA1 — can affect splice site selection *in vitro* under appropriate conditions (see, for example, references 17,20,29,36 and 53). While these proteins are candidates for splicing regulators, direct evidence for this role *in vivo* is lacking.

Four currently known metazoan proteins behave as expected of dedicated splicing regulators — all from the arthropod, *Drosophila*. Three of these participate in sex determination (2,4,6,9,35,45,47 and references therein). Analysis of this specialized sex-determination circuitry has produced important insight into possibly general mechanisms for splicing regulation (ibid.); however, none of these regulatory proteins is implicated in the control of generalized, nonsex-specific alternative splicing.

The fourth of these potential dedicated regulators is encoded by the *suppressor-of-white-apricot* (*DmSWAP*) gene (see Materials and Methods for naming convention). *DmSWAP* modulates a specific set of somatic, sex-independent pre-mRNA processing events — including autoregulation of splicing of its own pre-mRNA (8,34,35,38,41,50–52).

Phylogenetic analysis of gene and protein structure is powerful and efficient. Studies of this sort are fundamentally independent of other approaches — including *in vitro* biochemical analysis — thereby providing unique insight not attainable in other ways. For example, the power of this approach is well illustrated by the analysis of the RRM RNA binding module. This module was originally identified and partially characterized largely on the basis of phylogenetic analyses (reviewed in references 15 and 27).

Nematodes and arthropods are distinct phyla comparably divergent to arthropods and vertebrates (ca. 600 MYrs). We report studies exploiting this extreme divergence to define a novel, ancient family of splicing regulators (SWAP) related to *DmSWAP*. Our studies further define an ancient family of constitutive splicing proteins (*prp21*) sharing a novel conserved sequence motif with SWAP family proteins. We discuss the implications of our results for mechanisms of regulated metazoan pre-mRNA splicing.

MATERIALS AND METHODS

SWAP family naming convention

The historically defined gene name, *suppressor-of-white-apricot*, and its corresponding abbreviation, *su(w^a)*, are cumbersome. In light of the results described here and together with R.Lafyatis who has recently cloned a mammalian SWAP gene (personal communication; Discussion), we suggest adoption of SWAP as

*To whom correspondence should be addressed

the new name/abbreviation for this gene family with genus/species initials for its individual members. On this nomenclature *Drosophila suppressor-of-white-apricot* is designated *DmSWAP*. New *Drosophila* SWAP family members would be designated *DmSWAP2*, *DmSWAP3* and so on.

Molecular biology

Northern analysis was carried out as described in Zachar *et al.* (52). mRNA size measurements used a series of radiolabelled RNA standards generated by *in vitro* transcription (T7) of fully characterized (completely sequenced) templates linearized at specific sites — resulting in RNAs whose size is known to within a few bases. (Segments of the *Escherichia coli* β-galactosidase gene were used.) In this case these RNAs were 3380, 2502 and 1343 bases in size. In addition, these RNA standards were supplemented by an end-labeled 1 kb 'ladder' of standard DNA fragments (BRL). Standards were run in flanking channels and transferred to a nylon membrane with the polyadenylated nematode RNA sample. After Northern hybridization, the nematode mRNA and contiguous standards were visualized by autoradiography. We estimate that this procedure allows size measurements for the various *CeSWAP* and *Ceprp21* RNAs with uncertainties of less than 10%.

RT/PCR was carried out as described (19). The primer used for *CeSWAP* reverse transcription (bottom right panel, Figure 1) is GGGATGTATTTGTTTTCACG. PCR primers used for *CeSWAP* transcript analysis (bottom right panel, Figure 1; diagrammed in Figure 2) were as follows: PCR-1 is CACACAGTGCCATGTCTCGG, PCR-2 is ATGCCGGTCTTCGATGATTC and PCR-3 is GCAACTCGAGACCAACGGGG. Oligo-dT cellulose fractionated RNA (52) isolated from a nematode population in which all developmental stages were represented was used throughout for Northern and RT/PCR experiments.

DNA sequence determination was done as described (24).

The structure of the alternatively spliced regions of *CeSWAP* transcripts were defined as follows. Exon1−exon 2 and exon1−exon1a−exon2 spliced forms were generated by RT/PCR as shown in the bottom right panel of Figure 1, cloned (Bluescript; Strategene) from the first exon PCR-1 primer through the *Xho*I site at DNA sequence coordinate 1285 in the third exon (top panel, Figure 2) and sequenced. (Also see legend to Figure 1.)

Transcripts containing unspliced first *CeSWAP* introns were characterized by Northern and RT/PCR analysis (Figure 1) and by S1 protection as described (52) (results not shown).

Retrieval of cloned nematode segments

We retrieved the inserts from cDNA phage clones serial numbers *cm2h4* and *cm2h5* (49) by recloning into sequencing-ready plasmid vectors (Bluescript; Stratagene). In the case of *cm2h4*, these plasmid clones were used to retrieve chromosomal segments from lambda replacement libraries as described (42). *cm2h4* corresponds to *CeSWAP* and *cm2h5* to *Ceprp21*.

Data base searches and sequence alignments

Sequence similarity searches were done in variety of ways. The following selected subset of these searches illustrates the high statistical significance of the sequence matches defining the major motifs of the SWAP and *prp21* families. (All searches and alignments used default parameters.) A BLAST search (tblastn; 1) of the dbEST data base (10) using the amino acids 217 through 293 of the *DmSWAP* protein (includes the first *surp* module; Figure 3) retrieved the *cm2h4* (accession # gnl|dbest|5246 or gb# U06933) and *cm2h5* (accession # gnl|dbest|5247) expressed cDNA tags as the best matches.

A BLAST (blastp) search of the GenBank protein data base (before inclusion of *CeSWAP* and *Ceprp21*) with amino acids 7 through 82 of the *Ceprp21* protein (first *surp*; see Results; Figures 4 and 5) yields *DmSWAP* (*P* value 1.9×10^{-5}) and yeast *prp21/spp91* (accession # sp P32524; *P* value 8.5×10^{-4}) as the best two matches. The next best matches did not include the conserved features of the *surp* motif and had *P* values greater than 0.3. Moreover, a BLAST search of the GenBank nucleotide sequence data base with this protein segment using tblastn (that is, a search of the translation products of all six reading frames of the DNA sequence data base) yielded no additional examples of high quality matches to the *surp* motif.

A BLAST (blastp) search of the GenBank protein data base with amino acids 1 through 690 of *CeSWAP* (eliminating the RS module to prevent retrieval of the many non-SWAP RS domain-containing proteins; see Results) yields *DmSWAP* as the highest quality heterologous match (*P* value 8.8×10^{-16}).

A BLAST (blastp) search of the GenBank protein data base (after inclusion of *CeSWAP*) with amino acids 51 through 182 of the *DmSWAP* protein (DRY CEEERYL motif; see Results; Figure 3) yields *CeSWAP* (*P* value 1.1×10^{-11}) as the highest quality heterologous match. The next highest quality heterologous match does not contain the conserved features characteristic of DRY CEEERYL and has a *P* value ca. 108-fold higher. Moreover, a BLAST search of the GenBank nucleotide sequence data base with this protein segment using tblastn yielded no additional examples of high quality matches to the DRY CEEERYL motif.

Search of the GenBank protein data base with amino acids 582 through 653 of the *Ceprp21* protein (the ubiquitin-like segment; see Results; Figures 4 and 5) yields exclusively ubiquitins and ubiquitin-like segments as the first 100 best matches (*P* values ranging from 1.4×10^{-9} to 5.1×10^{-7}).

Pairwise sequence comparisons (see figure legends) were generated using either the GAP or PILEUP programs of the Genetics Computing Group analysis package (version 7.0; 14).

Antibody production and immunolocalization

A segment encoding *Ceprp21* amino acids 1 through 211 was cloned into the pQE-30 bacterial expression vector (Diagen, Inc.) resulting in fusion of the *Ceprp21* segment to a short additional peptide including a polyhistidine tag. This fusion protein was purified under denaturing conditions by binding to nickel-agarose according to the manufacturer's instructions (Diagen, Inc.).

Polyclonal mouse antibodies were produced as follows. Fifty micrograms of *Ceprp21* fusion protein (above) dissolved in 17 μl of 8 M urea was suspended in 500 μl of MPL+TDM Emulsion [RIBI Adjuvant System (RAS); RIBI ImmunoChem Research, Inc., Hamilton, MT, USA] and injected intraperitoneally. After a series of four such injections at 2 week intervals, blood samples were recovered by suborbital bleeds and cleared after clotting to produce sera. (See reference 23 for additional technical details.)

Immunolocalizations on cryosectioned nematode samples was carried out as described (52). Sera were used at 1:300 dilution for images shown in Figure 6. The population of nematodes sectioned contained all ages and all showed similar nuclear labeling. The anti-histone antibody used as a positive control in Figure 6 was a mouse monoclonal (MAB052) purchased from Chemicon (Temecula, CA, USA).

Western analysis of the *Ceprp21* protein (Figure 6) was done using conventional procedures (23). Nematode nuclear extracts were prepared from a culture containing all developmental stages using the recipe previously described for *Drosophila* embryos (44). The control, bacterially expressed *Ceprp21* protein consisted of amino acids 1 through 659 fused to a 40 amino acid peptide including a polyhistidine tag. This fusion protein was purified by nickel-agarose binding as above.

RESULTS

Structure of the major gene product of a nematode DmSWAP cognate, *CeSWAP*

Waterston *et al.* (49) reported a large collection of short (ca. 350 bp) segments of single-stranded DNA sequence from the 5′ regions of a collection of nematode cDNAs. Two of these segments encoded peptides with statistically significant sequence similarity to small portions of the *Drosophila DmSWAP* gene. We have analyzed the two genes containing these segments.

The first is designated *CeSWAP* (Materials and Methods) and encodes a ca. 2.5 kb major polyadenylated RNA (designated mRNA; Figure 1). We characterized the structure of this transcript as follows. (Other, alternatively spliced *CeSWAP* RNAs — the 'blocked' and 'frameshifted' RNAs in Figure 1 — are discussed below.) A cDNA clone of a portion of the transcript was sequenced (Materials and Methods). This yielded a 1981 base segment extending through the polyadenylated 3′ terminus (top panel, Figure 2) — a segment significantly shorter than the major 2.5 kb mRNA. To analyze the remainder of the gene and to facilitate characterization of alternatively processed transcripts we retrieved and sequenced most of the chromosomal gene (top panel, Figure 2). Comparison of this additional sequence with the *Drosophila DmSWAP* gene strongly suggested the existence of two additional exons (designated 1 and 2) 5′ to those present in the truncated cDNA. Northern analysis demonstrated homology to these two predicted exons in the 2.5 kb RNA (top panel, Figure 1).

We confirmed this structure for the major 2.5 kb mRNA by sequencing additional cloned cDNA segments produced by reverse transcription/PCR (RT/PCR)(see legend to Figure 1 and Materials and Methods for technical details; see top panel, Figure 2 for details of transcript structure).

The major 2.5 kb *CeSWAP* mRNA encodes a 775 amino acid conceptual translation product (top panel, Figure 3) from a presumptive initiator AUG in the first exon (top panel, Figure 2). This *CeSWAP* protein has extensive, colinearly arrayed blocks of similarity to the *Drosophila DmSWAP* protein beginning with

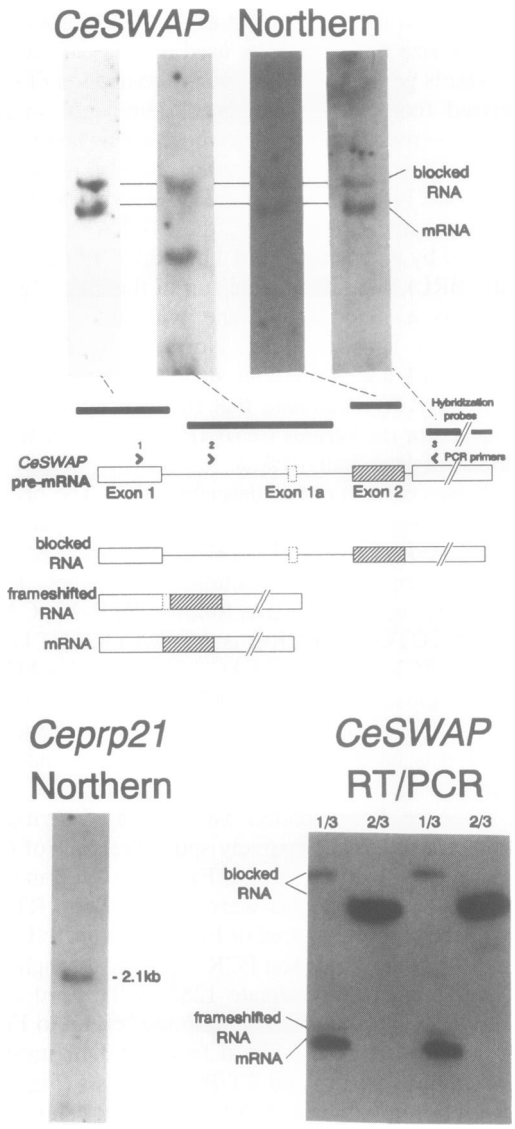


Figure 1. Analysis of structures of the *CeSWAP* and *Ceprp21* transcripts. **Top:** Northern analysis of *CeSWAP* transcripts. Probes used for each filter are indicated by lines connecting to transcript structure diagram. [Probes extend from DNA sequence coordinates 1–510, 630–978, 1008–1144 and 1235–3356 (top panel, Figure 2), respectively.] Positions of mRNA (2.5 kb) and blocked RNA (3.1 kb) are indicated. The frameshifted RNA (below) comigrates with the mRNA under the conditions of this experiment. Low abundance and very short homology to the probe segment prevents detection of the frameshifted RNA with the first intron probe. The same filter was probed with the first exon probe, stripped and reprobed with the first intron probe and, last, stripped and reprobed with the second exon probe. The position of the possible free first intron species (text) detected with the first intron probe is indicated by the dot. **Middle:** Structural diagram of the *CeSWAP* pre-mRNA and its alternatively processed forms (also see Figure 2). Exons 1, 1a and 2 and introns 1 and 2 are indicated. Several small introns in the 3′ portion of the gene (top panel, Figure 2) are omitted for simplicity. Also indicated are the positions of PCR primers and hybridization probes used in the analyses of *CeSWAP* RNAs. We emphasize that we have not precisely mapped extreme 5′ termini of *CeSWAP* transcripts and that more complex structures for these termini than those diagrammed are possible. Note, however, that the sizes of *CeSWAP* transcripts exclude any large 5′ extensions beyond the fully characterized portions of these RNAs. **Bottom right:** Two fully independent runs of an RT/PCR analysis of *CeSWAP* RNAs. Polyadenylated RNA was reverse transcribed from a primer immediately 3′ to PCR primer 3 (RT in top panel of Figure 2; Materials and Methods). PCR amplifications were performed using primer pairs indicated above each gel channel (precise positions of primers shown in the top panel of Figure 2). RT/PCR products were analyzed by Southern gel analysis using the second exon segment diagrammed at top as sequence probe. Positions of the PCR products corresponding to mRNA, blocked RNA and frameshifted RNA are indicated. Direct sequencing of primer 1/3 PCR products with a primer in the second exon directed across the first intron demonstrates that the majority PCR product corresponds to the exon 1–exon 2 spliced form (mRNA) (results not shown). Note that the stoichiometry of the blocked PCR product is slightly reduced in the 1/3 amplifications due to modest selection against large products during RT/PCR. **Bottom left:** Northern analysis of *Ceprp21* RNAs. The *Ceprp21* cDNA shown in Figure 5 (top panel) was used as sequence probe. Note the presence of a single polyadenylated transcript form at ca. 2.1 kb.

CeSWAP gene

```
   1  GGATCCATCG CTTAAGCTGT TGTTTTTGTT GTTACAAATG AATATTAATT CCGTTGAACT GTTTGCAAGT TGAATAAAAC AGAATGGTGT TCGGGTCATT

 101  AGAATATTTC ATTTAAGTTT ATTATTTTTA ACGTTACTTT CCTATTCTGT TTAGACAATG GGAAGTTGGA ATCGTCGAAA CCAAGTCGGC AATAATGACG

 201  ATCAGAATGT GAGTAACATT TTCAATTCAA GCAATCAGTT TCATAACGTT ACTTATGCCT TGCAACTGTT TTAAAAGTAA TGTGCATGTT CAGGAATACA
                                                          ^ presumptive initiation codon
                                                                                    /- - - -PCR-1- - - - >
 301  AAGATCTGCT AGTTTTCGGA TATGCTTCTA CAATATTTCC AAATGATTAT CAGTCGGAGC ACATTGCAGA GGAACGACAC ACAGTGCCAT GTCTCGGAGA

 401  TCCAGAAAAT CGCGTGGATA GGTATGGTTT TTTTCATTAT TGATGAATCT CTAACCATTA AAAATCGGCA CAACGAAAGA GAAGTCGTTT GGCTTTTTTG

 501  CTGAAAGAGT AACTGATTCA CTGGAATATA TGGATTTTAT AAGTGGTGAG TTTACTTTCC TAATCAATTT TATATTAGGT GTATTGATTT TTCTGCTAAA

                                          /- - - -PCR-2 - - - - >
 601  AAACTTACAG AATCAAGGTT TTGACAGAAA GCTTTCATGC CGGTCTTCGA TGATTCCGAT ATTTCCTCAC AATTTACCAT ATAATGTTCA ATTAGTTTTA

 701  AATTTTTCTT CAATCCAATT AATTGAAACA GTTCACCAAT CGTTTGCTGA TGACACAATT TTTAGAAAGC TCAAGTATTA AAACAAATAT AAACCACAAA

 801  AGTTGCTTCG TTTGTACGGG TCTATATAAT TTAAAAATTA TTAATTGATT TAATTAATTT TGAATATTCA GGTCGATCGA TTGGCTGGCA CACCGAACAT
                                                    alternative exon (1a) >  GTCGATCGA TTGGCTGGCA CACCGAACAT

 901  GTCAAATAAC CGTGAGTATT TCGAAAAAAA AGTGCGGGAA ATAGTCAAAA CTACGATTTG AATAAAGAAA ATTTCGCCAC TAATAAATGA TTTTTCAGGT
      GTCAAATAAC C

1001  ATGACTGTCG ACTTCTGTTG CCATCAATCG ACGTGGCAAT CAAAAGAAAT GGTTCGCCGT CTGAACAATG TCCTACAGAA GCAATGGAAG AGGATATGTG

1101  TGAAGAGGAA AGATATCTTG ATATGTATAA AGATATTCAA AGTATGAAAC AGCTTTATTA GCCAAAAACA GTTATTTTTA ATTTCAGGAG AGCAAGAAAA

1201  AGAAGAAGAG GAGAAGCGAA GGAATGACCA ACGAAATGCC ATTGGATTCG ATTACGGAAC AGGAAAAGTA AAAGCTCGAG AGAGTGATAG TGAGGATGAA
                              ^5' end of cDNA                                          ^XhoI
                                 < - - - -PCR-3 - - - -\
1301  CCATTTGAGC CGCCAGAAGG AATAAAATTC CCCGTTGGTC TCGAGTTGCC TTCGAATATG AAACTTCATC ATATTATCGA GAAGACAGCC TCATTTATAG

1401  TGGCAAATGG TACACAAATG GAGATTGTTA TCAAGGCGAA GCAAAGGAAT AATGCTGAAC AATTCGGATT CTTGGAATTC GATCATCGAT TGAATCCATT
                                 < - - - RT- - - - - -\
1501  TTATAAGTAT CTTCAAAAGC TTATTCGTGA AAAGAAATAC ATCCCAGATC TCAATAAAAG GCCAAAAAAG CTAACGAAAA CGTCAAGAGC TTCTACTTCA

1601  AAACCTGCAA TTTCTAGCTC CCTTGCTGCA ATTGCAGCTG CTCATGGATC AGATTCAGAA GATTCAGATT CAGACTACGA GCTTCACCCA TCTCTGTTGT

1701  CCGGCGGCGC GAAACGTCCT GTTACTCCAG AGAAGCCAGG AGCTATTGGT CCACGGAAGA AGCCTGTTGA GCCGGAGAAA CCACCAGATT TCACCCTCAA

1801  GCCAGTCGGT GATATTTCGC AGAGAAATGA TGTTTATGCG GCGCTTTTCA AGAATCTGGC GCACGTAACG AGGCAAGCTG CAGGTGTAGA AGAAGTTAAG

1901  ATGAATGTTG AAGAAGCTAA GAAAGAGAAA GAAAATGATC ATCTCGACGA TCCAGAATAC CGGGAGTGGT ACGAAAACTT CTATGGACGT CCGTGCCCAT

2001  GGATTGGGCC TCGTCCCATG ATTCCAGCAA CTCCAGATCT TGAGCCCATC CTTAATAGCT ATGCGGAACA CGTGGCTCAA CGTGGATTAG AGGCAGAGGC

2101  GTCTCTTGCA GCCCGAGAAG ATCTTCAATT GCATTTTATG GAACCTAAAA GTCCTTATTA TTCATATTAT CATCACAAAG TGAGTTTTTT GTAATTCAAG

2201  AATATAAGTT TATGCTCCGT TTCAGGTTCG TATGCATCAA TGGAGAATGT ACCAACCCAT TGAACAAAAT CTATCACCAC TTGTTCTCAA CTCACCAGCT

2301  CCACCATCGG CTGTCAGTTC ACCAGGACCT TCCAGTCTTA TGAGCCTGAA TCTATCGACC CCGGAGCCGC CACTCAATCG AAGGCAGAGA CGGCGTCTCC

2401  TAGATTCCAG CCGTCTTGAC GAGTCTATCA CTGAACCAGG AGTCATAGAT CCAATCACAA TGGTAAGGAA TTCTTGCAAA ATTTATGTAT CATCTCTTCA

2501  TAGTTACAGA TTCCAAAAAG TGTATCGACT CCTGCAAATC TAGATATTCT GAAGACACCT ATATCATTTT CCCTTAGAAA CGACGAGCCA CGGGATGAAT

2601  CAAGTTTCCG ATTTgatcCG GATCTGGATG AAACTGCAGG ACCTTCAGAC ACAACTGCAA ACTTCAGTGA TATTAGTGGC CTATTCCCTC CACCAACACC

2701  TCCTGTAATC CCACCATCCA CTCAAATGCA AGTCGATCGG AAGGAAAAAG CGAGAATTTT CATGGAGAAG TTGCTTCAAG AGAAGAAAGC AAAGAAATTA

2801  CAAGAAGAAG AAGAACGATC AAAATTAGAG GAAGAGACAC GGAAGAAGGC TGAAAAGATA TCAGAATCAT TGTCAGAACG GAAAAATACT GGTAGATCGG

2901  ATCGGAGAGA AGAAGCACCC AAAGGGGCGA GATCTCTCGA TGAAATAATT AATAATAGGA TCAACAGTTT GCTATCCGAA TCTGGTTTTG AACCCGTTGA

3001  GGAGATGAAG AGAACAGACG AGGATAGGGA GAGAAAAAGA CATCGAAAAC GAAGCCGTTC ACGACGACGA TCTCGCTCCT GTAGCCCCAG AGACCGATCA

3101  CGGGAGCACA AAAAATCCCG AAAATCCGGT AGACATCATC GATCTCGCTC TCGTTCCTCA TCCAGAGATC GTCATCGTCG AAATCGCAGT AGAAGTCGGG

3201  ATAGACGGCG GTGATTTCTT GTGGAATAAT TAATTTTTTA AAGCAATTTT CTCATAATTT ATATATCATT TATTTTTGTT TGTCGAATAA ATTAACAACA
                ^termination codon                                          polyadenylation signal^
3301  ATAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA
```

Splicing signals for alternative Exon 1a

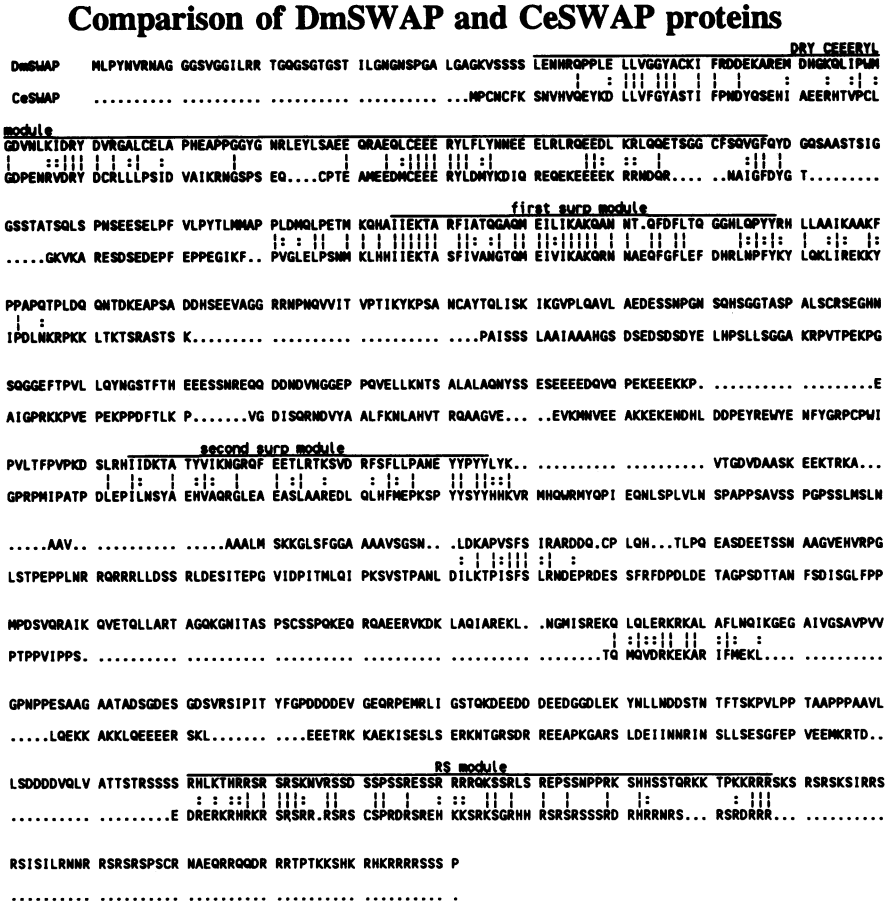
consensus TTTCAG AG GTAAGT

TAATTTTGAATATTCAG[GTCGA..Exon 1a..TGTCAAATAACCC]GTGAGTAT

Figure 2. DNA sequence of the *CeSWAP* gene. **Top:** Shown is the sequence of the *CeSWAP* chromosomal gene region beginning at a *Bam*HI cleavage site and proceeding through the point of polyadenylation of the *CeSWAP* RNAs. Chromosomal sequence extends to coordinate 2617 (lower case *Sau*3A site). Sequence from 2617 through 3359 is from cDNA (Materials and Methods). Introns removed to produce the major 2.5 kb mRNA are underlined. Positions of presumptive translation initiation and termination codons, of alternative exon 1a, of the 5′ extent of the original *CeSWAP* cDNA and of the PCR primers used for analysis are indicated. Also see legend to Figure 1. **Bottom:** Shown are the splice sites surrounding alternative *CeSWAP* exon 1a. Note that these splice sites deviate from optimal consensus and these deviations are underlined.

first exon-coded sequences and extending through sequence encoded by the 3′ terminus of the gene (top panel, Figure 3; below). Moreover, placement of the first introns of the *Drosophila DmSWAP* gene and the *Caenorhabditis elegans CeSWAP* genes relative to their encoded proteins is precisely conserved (bottom panel, Figure 3; see below for additional discussion of this particularly important intron). [Second introns are at slightly different positions in the two genes (see reference 13 for *DmSWAP* exon/intron structure) and the remaining introns are not similarly placed. This likely represents intron movement and differential intron loss/acquisition in the two lines of descent — see reference 21 and references therein for discussion.] Collectively, these results indicate that *DmSWAP* and *CeSWAP* descended from a gene present in a common ancestor of contemporary arthropods and nematodes and possessing the numerous shared features of these two contemporary genes (also see Discussion).

Segments corresponding approximately to amino acids 235 through 277 and 484 through 524 of the fly protein and 166 through 209 and 391 through 431 of the nematode protein represent copies (two per protein) of a conserved sequence motif (top panel, Figure 3; Figure 4). [This and several other protein



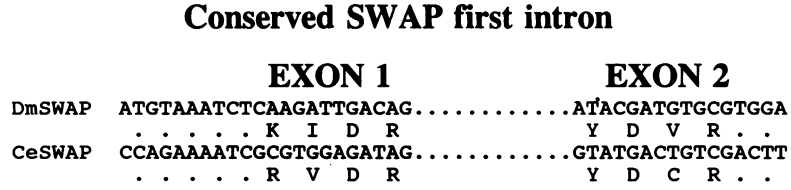


Figure 3. Conserved features of two SWAP genes. **Top:** Shown is a comparison of the *DmSWAP* and *CeSWAP* protein sequences. Identities (¦) and conservative substitutions (:) *within* conserved modules are indicated. The extended conserved segments discussed in the text are labeled and overlined. Alignment was constructed with the GAP program from GCG package using all and various subfragments of the proteins (Materials and Methods). Various nearly equally probable alignments of the RS modules are possible given their internally repetitive structures. Moreover, the alignments of the short segments of similarity on the interval between second *surp* and the RS module were chosen by eye from among several possible alignments (as determined by aligning different subfragments of the two proteins) to maximize alignment of nonrepetitive amino acid segments. **Bottom:** The conserved positions of the first introns in *DmSWAP* and *CeSWAP* are shown. Exon DNA sequences are written and intron sequences are indicated by dotted lines. The amino acid sequence resulting from translation of the spliced mRNA is indicated below each exon.

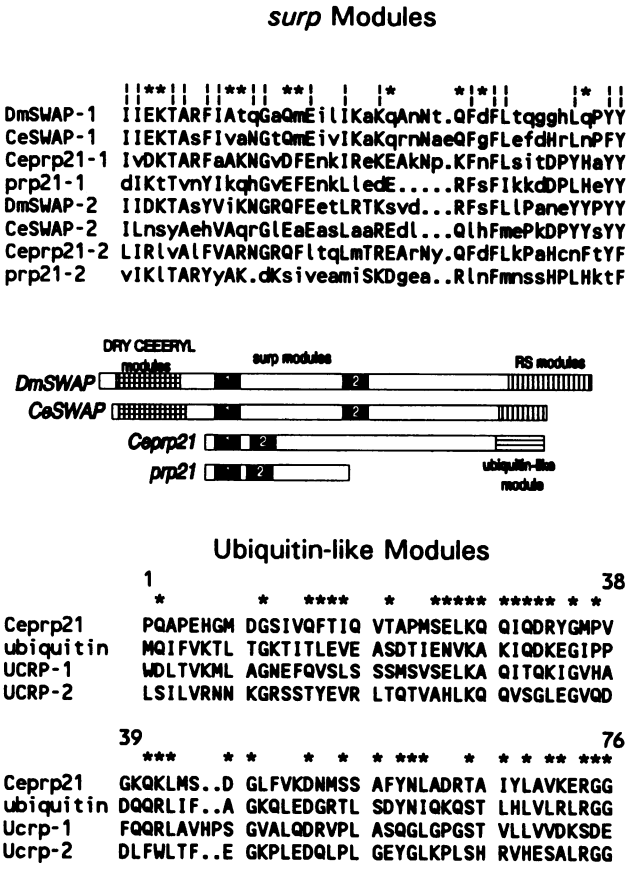


Figure 4. Structural features of *surp* superfamily proteins. **Top:** Alignment of *surp* modules from *DmSWAP*, *CeSWAP*, *Ceprp21* and *prp21*. The top four lines represent the first *surp* module in each gene and the bottom four the second. Residues shared by three or more *surp* modules (including members of conservative pairs ED, RK, IL, ST and FY) are capitalized. The most conserved features are indicated by vertical hatches above the sequence stack. Positions with less extensive conservation of residue type are indicated by asterisks. Several additional positions show conservation of general residue type. Individual pairs of *surp* modules show some limited additional sequence similarity extending into the regions immediately flanking the 'core' *surp* modules diagrammed here (see top panel, Figure 3 and bottom panel, Figure 5); however, the functional *surp* module is unlikely to be larger than ca. 60–90 residues based on various observations including module spacing in *Ceprp21* and *prp21* (Results). The second *surp* modules of *CeSWAP* and *prp21* show less extensive similarity to the *surp* consensus; however, we believe these are likely to be authentic in view of their positions in the corresponding proteins and of conservation of key residues. This alignment was generated using the PILEUP program of the GCG package (Materials and Methods). **Middle:** Organization of four *surp* superfamily proteins. DRY CEEERYL modules (top panel, Figure 3) are cross-hatched, *surp* modules are solid portions of each bar, RS modules (text; top panel, Figure 3) are vertically hatched and the ubiquitin-like module of *Ceprp21* (top panel, Figure 5) is horizontally hatched. *surp* modules are amino acids 235–277 and 484–524 of *DmSWAP* (13), 166–209 and 391–431 of *CeSWAP*, 37–79 and 134–176 of *Ceprp21* (cDNA fragment translation product) and 11–49 and 95–135 of *prp21* (3 and 12). **Bottom:** Alignment of the ubiquitin-like domain of *Ceprp21* with ubiquitin (nematode; 18 and references therein) and the two ubiquitin-like modules from the human UCRP gene (22). Sequence similarities between the *Ceprp21* module and any of the other three modules (including members of conservative pairs ED, RK, IL, ST and FY as identities) are indicated by asterisks. [Note that a number of additional positions show substantial conservation of general residue type.] Sequences are numbered relative to the 76 amino acid mature ubiquitin module. No gapping is required in the alignment of the *Ceprp21* module, ubiquitin and UCRP-2 module. Amino acids C-terminal to the final GG (or equivalent) have been removed (see top panel, Figure 5). Additional ubiquitin-like domains (5,46) show comparable similarity with the *Ceprp21* module. This alignment was generated using the PILEUP program of the GCG package (Materials and Methods).

motifs are referred to as 'modules' in Figures for reasons described below (Discussion).] This motif shows extensive similarity to only one other protein in the current sequence data base—the obligate, constitutive yeast splicing factor, *prp21*/spp91 (top panel, Figure 4; references 3,12). [See Materials and Methods for discussion of statistical significance of these and other matches.] This motif is also found in a nematode *prp21* cognate, *Ceprp21*, described below.

We designate this new motif *surp* after the first two cloned genes containing it [*suppressor-of-white-apricot* and *prp21/spp91*]. The conserved features of the *surp* motif include aliphatic, aromatic and basic residues (top panel, Figure 4). *surp* is not related in detailed sequence to any module previously defined in RNA binding and splicing proteins (8,15,27,39,48 and references therein).

At the C-terminus of *CeSWAP* is a 68 amino acid motif very rich in arginine and serine (top panel, Figure 3; Figure 4). RS motifs were originally recognized in the *Drosophila DmSWAP* and *tra* splicing proteins (8) and have subsequently been found in a large number of vertebrate and arthropod splicing-associated proteins (see, for example, references 20,29,33,47 and 53). We believe this to be the first documented report of an RS motif in nematodes.

None of the remaining segments of sequence similarity between *DmSWAP* and *CeSWAP* show obvious relationship to other proteins in the current data base (Materials and Methods). We designate the largest of these unique motifs DRY CEEERYL (pronounced 'dry cereal') after two conserved amino acid segments within it (top panel, Figure 3).

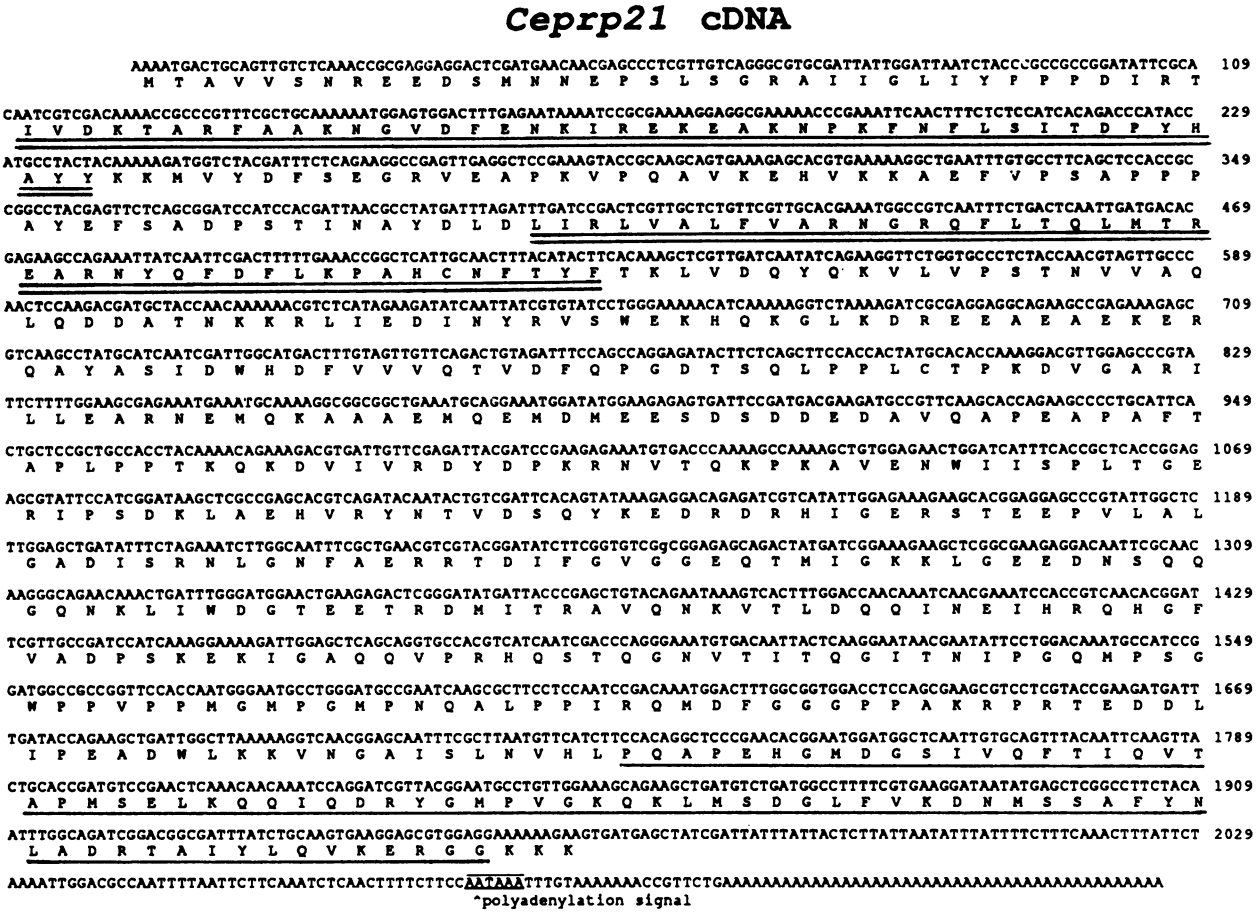
Our attempts to date to generate antisera against the *CeSWAP* protein suitable for reliable *in situ* immunolocalization have been unsuccessful. [Based on our experience with the *DmSWAP* protein this likely reflects relatively low abundance of the *CeSWAP* protein necessitating production of more sensitive antibody probes than are currently available (our unpublished results).] We note that the arthropod *DmSWAP* protein immunolocalizes to nuclei as expected of a splicing regulator (51).

Evidence for on/off regulation of *CeSWAP* expression at the level of pre-mRNA splicing

In addition to the major 2.5 kb mRNA, *CeSWAP* produces substantial levels of two other polyadenylated RNAs. One of these migrates at ca. 3.1 kb and results from the failure to remove the first intron from the *CeSWAP* pre-mRNA as assessed by Northern and S1 protection analyses (top panel, Figure 1 and results not shown). We therefore refer to this as the 'blocked' or incompletely spliced *CeSWAP* RNA. Translation of the blocked RNA initiated in the first exon would generate a short fusion protein containing the first 56 amino acids of the 775 amino acid *CeSWAP* protein (Figure 3) and would terminate at a UGA codon early in the first intron (DNA sequence coordinate 441; top panel, Figure 2). This RNA is thus unlikely to encode a functional protein.

Such high levels of incompletely spliced, polyadenylated pre-mRNA are quite uncommon in general. However, strikingly similar high levels of incompletely spliced pre-mRNAs are observed from the *Drosophila DmSWAP* gene. These incompletely spliced *DmSWAP* transcripts result from repression of first (and second) intron removal by the *DmSWAP* protein itself in an autoregulatory feedback circuit (13,50–52; Discussion).

The third *CeSWAP* polyadenylated transcript is similar in structure to the major 2.5 kb mRNA but contains an additional



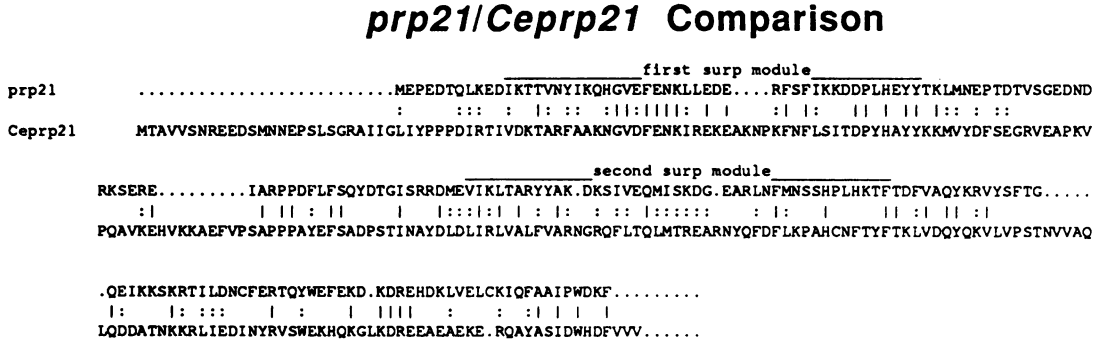


Figure 5. Analysis of the *Ceprp21* cDNA and protein. **Top:** Shown are cDNA and conceptual translation product sequences. The two *surp* modules near the amino terminus are double underlined and the ubiquitin-like module near the C-terminus is single underlined (see also bottom panel, Figure 4). Northern analysis (bottom left panel, Figure 1) indicates that this is likely a full length cDNA. **Bottom:** At bottom is comparison of yeast *prp21* and nematode *Ceprp21* beginning at the amino terminus of each protein and extending through regions with significant similarity. Identities (|) and similarities (:) *within* conserved sequence segments are indicated. The conserved *surp* modules (text) are overlined and labeled. This alignment was generated using the GAP program of the GCG package (Materials and Methods).

exon between exons 1 and 2 (middle and lower right panels, Figure 1; top panel, Figure 2). This additional exon (designated 1a) is very small (40 bases) and the RNA containing it is not resolvable on Northern gels from the major 2.5 kb RNA. This species was originally detected as a secondary band (ca. 10–20% as abundant as the major 2.5 kb RNA) in RT/PCR experiments and its structure was determined by sequencing the corresponding cloned cDNA product (lower right panel, Figure 1; Materials and Methods).

Inclusion of the 40 base exon 1a results in a translation reading frameshift between the presumptive initiation site in the first exon and the remainder of the gene. We therefore refer to this splice variant as the 'frameshifted' *CeSWAP* transcript (middle panel, Figure 1; top panel, Figure 2). Translation of the frameshifted

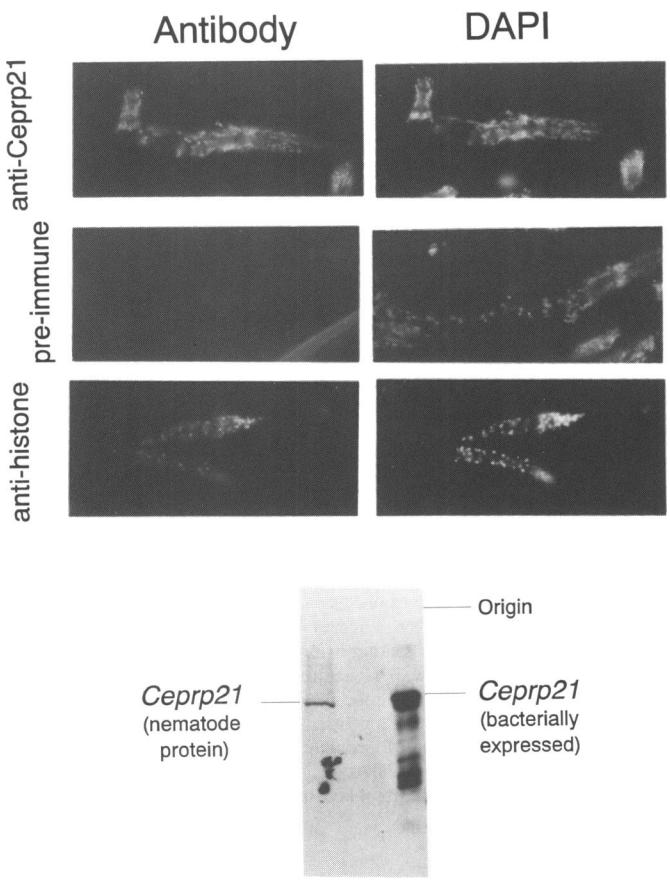


Figure 6. Immunolocalization of *Ceprp21* protein. **Top:** Shown are paired images from double label *in situ* localization. The left-hand panel is an immunolocalization with the indicated antibody. The right-hand panel is a DAPI stain of nuclear DNA in the same section. A single pair of pre- and post-immune anti-*Ceprp21* sera is shown from four independent pairs producing similar results. **Bottom:** Western analysis of *Ceprp21* protein. The left-hand channel contains a nematode nuclear extract and the right-hand panel bacterially expressed *Ceprp21* protein (Materials and Methods). Positions of full length *Ceprp21* protein are indicated. Bacterially expressed material also contains significant amounts of smaller products resulting from degradation and/or premature translation termination. Bacterially expressed protein moves slightly more slowly than the authentic worm protein as a result of the 40 amino acid N-terminal extension used for purification (Materials and Methods).

transcript would generate a short fusion protein containing only the first 56 amino acids of the 775 amino acid *CeSWAP* protein (top panel, Figure 3) and would terminate at a UGA codon early in exon 2 (DNA sequence coordinate 1002; top panel, Figure 2). This RNA is thus unlikely to encode a functional protein product. The presence of the 1a exon within the first intron strongly suggests that exon 1a inclusion and delayed or blocked first intron splicing are mechanistically related (Figure 7; Discussion).

In addition to the 3.1 kb blocked RNA, the first intron probe hybridizes to a second, small RNA species. This species does not have detectable sequence homology to any exons of the *CeSWAP* mRNA (top panel, Figure 1). There are various possible origins for this RNA — including a relatively stable, excised first intron lariat (reviewed in reference 28) — and we have not investigated it further.

Structure and nuclear expression of the protein product of the nematode *prp21* cognate, *Ceprp21*

We have analyzed a second nematode gene — designated *Ceprp21* — retrieved on the basis of similarity to *DmSWAP* (top panel, Figure 5; Materials and Methods). The 2145 base *Ceprp21* cDNA we have sequenced is homologous to a single 2.1 kb mRNA as assessed by Northern analysis (bottom left panel, Figure 1) and is, thus, indistinguishable from a full length copy of the corresponding mRNA. This cDNA encodes a 659 amino acid conceptual translation product assuming the 5′-most methionine to be the translation initiator (top panel, Figure 5). *Ceprp21* contains two *surp* motifs—one well conserved and a second less well conserved (top panel, Figure 4).

Several lines of evidence indicate that *Ceprp21* is the nematode cognate of the constitutive *prp21* protein. First, the spacing of *surp* motifs in *Ceprp21* and *prp21* is very similar and distinct from SWAP family *surp* spacing (middle panel, Figure 4). Second, *Ceprp21* shows sequence similarity to *prp21* beyond the basic *surp* repeats in contrast to SWAP family proteins (bottom panel, Figure 5). Third, *Ceprp21* shows very extensive sequence similarity (47% identity, 65% similarity over a 450 amino acid segment) to a recently sequenced fragment of a presumptive mammalian *prp21* cognate (A.Kraemer, personal communication). Fourth, yeast *prp21* is apparently present in stoichiometric amounts with two additional proteins — *prp9* and *prp11* — whose metazoan cognates have recently been identified as abundant nuclear proteins (7,11,30). Thus, we anticipate that *Ceprp21* should be a very abundant nuclear protein if it is, in fact, a *prp21* cognate. We raised four independent, polyclonal antisera against bacterially expressed *Ceprp21* protein in mice showing no detectable preimmune reactivity with nematode tissues under standard conditions (Materials and Methods). All four immune sera produce intense, specific nuclear labeling in sectioned nematode tissue samples (Figure 6 and results not shown). Moreover, all four of these antisera label a single prominent protein in nematode nuclear extracts with the expected migration relative to bacterially expressed *Ceprp21* protein in Western transfers (Figure 6 and results not shown). Thus, *Ceprp21* is an abundant nuclear protein as predicted.

The nematode *Ceprp21* protein is substantially larger than yeast *prp21* (Figure 4). We note, however, that this is not inconsistent with these being cognate proteins. Some previously characterized metazoan cognates of other yeast splicing proteins have proven to be substantially larger — likewise containing extra motifs not present in the corresponding yeast proteins (see, for example, reference 7).

The C-terminal 78 amino acids of *Ceprp21* show significant sequence similarity to ubiquitin and to ubiquitin-like modules previously identified in several other metazoan proteins (bottom panel, Figure 4; top panel, Figure 5; Materials and Methods). It is particularly notable that this C-terminal motif has a three amino acid extension beyond the ubiquitin-like segment itself. A very similar structure is seen at the C-terminus of conventional ubiquitin genes (reviewed in reference 18).

DISCUSSION

Regulated metazoan pre-mRNA splicing is poorly understood. In particular, the regulators involved are virtually entirely unknown or uncertain (Introduction). We report here the identification and initial characterization of a novel, ancient family

of presumptive metazoan splicing regulators (SWAP) that—together with the *prp21* family of constitutive splicing proteins—make up a new superfamily of splicing-associated proteins.

Evidence that the ancient, conserved SWAP protein family consists of a colinearly arrayed set of protein modules

Our results show that *DmSWAP* and *CeSWAP* descended from a SWAP gene present in a shared ancestor of arthropods and nematodes. A SWAP gene recently isolated from a third phylum (vertebrates) shows similar conservation of protein motifs and precise first intron placement (R.Lafyatis, personal communication). Collectively, these results demonstrate that SWAP genes were present in the common ancestor of all contemporary metazoan phyla and have been highly conserved during the ca. 600 MYr since the metazoan radiation.

Several observations indicate that the conserved sequence motifs making up the large SWAP proteins are functionally distinct protein elements or modules. First, three of these motifs (the two *surp* motifs and the RS motif) occur in other proteins in different contexts (Results). Second, the DRY CEERYL, *surp* and RS motifs are flanked by segments with hinge-like properties as expected if they represent functionally independent modules. That is, these flanking segments are poorly conserved in precise sequence, relatively hydrophilic and rich in glycine and proline residues (Figure 3).

Implications of the newly identified *surp* module for regulated and constitutive splicing

Our results define the new *surp* protein module. This module existed in the common ancestor of all eukaryotes and is conserved, apparently exclusively, in the contemporary *surp* superfamily consisting of SWAP splicing regulators and *prp21* constitutive splicing factors (Results).

This limited occurrence and substantial conservation indicate that *surp* module function is both specialized and important. Combining this observation with two earlier studies provides additional insight as follows. First, *prp21* is implicated in an early step irreversibly committing the pre-mRNA to the splicing pathway and to nuclear retention until splicing is executed (31,32). Second, *DmSWAP* represses splicing of its own pre-mRNA at a step before first covalent modification of the target intron and after commitment of unspliced material to nuclear retention (51,52) — a step indistinguishable from that involving *prp21*. Thus, we propose that *surp* modules function at a specific step early in prespliceosome assembly.

Implications of regulation of *CeSWAP* pre-mRNA splicing

Splicing of the first *DmSWAP* intron is subject to autogenous regulation by the *DmSWAP* protein resulting in accumulation of incompletely spliced pre-mRNAs. This feedback circuit represents a homeostatic device to control levels of the *DmSWAP* protein (reviewed in references 51 and 52). The first *CeSWAP* intron is subject to complex, alternative splicing including delayed or blocked splicing (Results). It is not currently feasible to directly test whether this alternative *CeSWAP* splicing represents the action of a homeostatic autoregulatory circuit analogously to the *DmSWAP* case. However, as follows, several of our results provide strong indirect support for this hypothesis indicating, in turn, that SWAP proteins are universally involved in splicing regulation.

First, the positions of the first introns of the arthropod, vertebrate and nematode SWAP genes are precisely identical

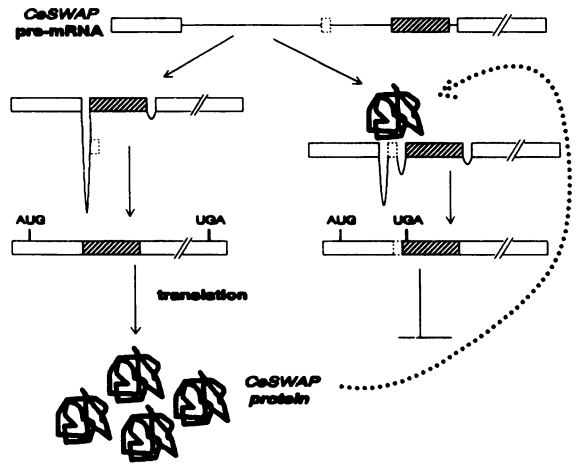


Figure 7. A model for autoregulation of *CeSWAP* at the level of splicing. *CeSWAP* exons are boxes and introns are lines. The model proposed is as follows: High levels of *CeSWAP* protein activate inclusion of exon 1a (dashed box) to produce the frameshifted *CeSWAP* RNA (Figures 1 and 2; text). This likely reflects *CeSWAP* protein-directed recognition of exon 1a or of one of the two introns flanking it. (Analogy with *DmSWAP* autoregulation implicates the small intron between exons 1a and 2; see text.) Splicing events involving exon 1a are proposed to be slow and *CeSWAP*-directed exon 1a inclusion thus also leads to accumulation of substantial steady-state levels of incompletely spliced first intron. The position of the presumptive initiator AUG codon and first inframe nonsense codon terminating translation are indicated for both the mRNA and the frameshifted RNA.

(above). This demonstrates beyond significant ambiguity that these three SWAP introns are descended from a common ancestral intron. Thus, shared properties — including regulated splicing — also very likely have common ancestry and mechanism. Second, *DmSWAP* autoregulation results in production of noncoding, alternatively spliced RNAs (reviewed in references 51 and 52). Structures of alternatively spliced *CeSWAP* RNAs indicate that they likewise do not encode functional protein (Results) as expected if they are byproducts of autoregulation analogous to *DmSWAP*. Third, structural similarities between the homologous first *CeSWAP* and *DmSWAP* introns strongly suggest a similar underlying mechanism of autoregulation. Specifically, extensive reverse genetic analysis of *DmSWAP* autoregulation indicates that this process involves recognition of an intron-like segment nested within — and sharing a 3′ splice site with — the larger first intron (I.P.Mims and P.M.Bingham, unpublished). An attractive, simple interpretation of our results is that *CeSWAP* autoregulation involves recognition of the small intron between alternatively included exon 1a and exon 2 (see Figure 7 and its legend; Results). The position of this *CeSWAP* intron is precisely homologous to the intron-like target for *DmSWAP* autoregulation.

Implications of additional protein modules of the *prp21* family

In addition to *surp* modules, *Ceprp21* and *prp21* show a third interval of sequence similarity (*Ceprp21* amino acids 202 through 251, *prp21* amino acids 151 through 200; bottom panel, Figure 5). This segment of yeast *prp21* contains the site of the spp91-1 mutation which apparently affects the *prp21*–*prp9* interaction (12,30). Thus, our results suggest that the interaction between *prp9* and *prp21* is likely to be conserved in the metazoan cognates of these yeast proteins.

The C-terminus of *Ceprp21* consists of a ubiquitin-like module and the details of its sequence strongly suggest ubiquitin-like structure and/or function (Results). Such modules have been observed in several other metazoan proteins but their significance remains unclear (4,22,25,26,46). The presence of a ubiquitin-like module in *Ceprp21* — a component of a potentially well understood, multiprotein splicing complex — may provide an unusually valuable opportunity to investigate the role of these modules.

ACKNOWLEDGEMENTS

We are grateful to Pierre Legrain, Michael Rosbash, Jaime Arenas and Martin Chalfie for helpful discussions. We are grateful to Robert Lafyatis and Angela Kraemer for sharing results before publication. We are grateful to Chris Martin and Bob Waterston for providing the *cm2h4* and *cm2h5* cDNA clones. We are especially grateful to colleagues Edwin Smith and Zuzana Zachar for ongoing discussions while this work was in progress. K.V.D. is very grateful to Valarie Engelder for technical assistance. This study was supported by NIH grant GM32003 to P.M.B. and NSF grant DMB9102833 to K.V.D. J.K. is a Hoffman-La Roche predoctoral fellow of the Institute for Cell and Developmental Biology.

REFERENCES

1. Altschol, S.F., Gish, W., Miller, W., Meyers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.* **215:** 410–415.
2. Amrein, H., Gorman, M. and Nothinger, R. (1988) *Cell* **55:** 1025–1035.
3. Arenas, J.E. and Abelson, J.N. (1993) *Proc. Natl Acad. Sci. USA* **90:** 6771–6775.
4. Baker, B.S. (1989) *Nature* **340:** 521–524.
5. Banerji, J., Sands, J., Strominger, J.L. and Spies, T. (1990) *Proc. Natl Acad. Sci. USA* **87:** 2374–2378.
6. Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988) *Cell* **55:** 1037–1046.
7. Bennett, M. and Reed, R. (1993) *Science* **262:** 105–108.
8. Bingham, P.M., Chou, T-B., Mims, I. and Zachar, Z.. (1988) *Trends Genet.* **4:** 134–138.
9. Boggs, R.T., Gregor, P., Idriss, S., Belote, J. and McKeown, M. (1987) *Cell* **50:** 739–747.
10. Boguski , M.S., Lowe, T.M.J. and Tolstoshev, C.M. (1993). *Nature Genet.* **4:** 332–333.
11. Brosi, R., Gronig, K., Behrens, M., Luhrmann, R. and Kraemer, A. (1993) *Science* **262:** 102–105.
12. Chapon, C. and Legrain, P. (1992) *EMBO J.* **11:** 3279–3288.
13. Chou, T-B., Zachar, Z. and Bingham, P.M. (1987) *EMBO J.* **6:** 4095–4104.
14. Devereux, A., Hueberly, F. and Smithies, O. (1984). *Nucleic Acids Res.* **12:** 387–395.
15. Dreyfuss, G., Swanson, M.S. and Pinol-Roma, S. (1988) *Trends Biochem. Sci.* **13:** 86–91.
16. Eng, F.J. and Warner, J.R. (1991). *Cell* **65:** 797–804.
17. Eperon, I.C., Ireland, D.C., Smith, A.R., Mayeda, A. and Krainer, A.R. (1993) *EMBO J.* **12:** 3607–3617.
18. Finley, D., Bartel, B. and Varshavsky, A. (1989) *Nature* **338:** 394–401.
19. Frohman, M.A. (1993) *Methods Enzymol.* **218:** 341–348.
20. Ge, H., Zuo, P. and Manley, J.L. (1991) *Cell* **66:** 373–382.
21. Gilbert, W., Marcionni, M. and McKnight, G. (1986) *Cell* **46:** 151–154.
22. Haas, A.L., Ahrens, P., Bright, P.M. and Ankel, H. (1987) *J. Biol. Chem.* **262:** 11315–11323.
23. Harlow, E. and Lane, D. (1988). *Antibodies: A laboratory manual.* Cold Spring Harbor Press. Cold Spring Harbor, NY.
24. Henikoff, S. (1987) *Methods Enzymol.* **155:** 156–165.
25. Jones, D. and Candido, E.P.M. (1993) *J. Biol. Chem.* **268:** 19545–19551.
26. Kas, K., Michiels, L. and Merregaert, J. (1992) *Biochem. Biophys. Res. Commun.* **187:** 927–933.
27. Kenan, D.J., Query, C.C. and Keene, J.D. (1991) *Trends Biochem. Sci.* **16:** 214–220.
28. Kopczynski, C.C. and Muskavitch, M.A.T. (1992) *J. Cell Biol.* **119:** 503–512.
29. Krainer, A.R., Mayeda, A., Kozak, D. and Binns, G. (1991) *Cell* **66:** 383–394.
30. Legrain, P. and Chapon, C. (1993) *Science* **262:** 108–110.
31. Legrain, P., Chapon, C. and Galisson, P. (1993) *Genes Dev.* **7:** 1390–1399.
32. Legrain, P. and Rosbash, M. (1989) *Cell* **57:** 573–583.
33. Li, H. and Bingham, P.M. (1991) *Cell* **67:** 335–342.
34. Maniatis, T. (1991) *Science* **251:** 33–34.
35. Mattox, W., Ryner, L. and Baker, B.S. (1992) *J. Biol. Chem.* **267:** 19032–19026.
36. Mayeda, A. and Krainer, A.R. (1992) *Cell* **68:** 365–375.
37. McKeown, M. (1992) *Annu. Rev. Cell Biol.* **8:**133–155.
38. Mohler, J. and Vani, K. (1992) *Development* **115:** 957–971.
39. Ruby, S.W. and Abelson, J. (1988) *Trends Genet.* **7:** 79–85.
40. Ruby, S.W., Chang, T-H. and Abelson, J. (1993) *Genes Dev.* **7:** 1909–1925.
41. Rutledge, B.J., Mortin, M.A., Schwarz, E., Thierry-Mieg, D. and Meselson, M. (1988) *Genetics* **119:** 391–397.
42. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory manual,* 2nd edn. Cold Spring Harbor Press, Cold Spring Harbor, NY.
43. Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1989) *Annu. Rev. Genet.* **32:** 527–577.
44. Spikes, D. and Bingham, P.M. (1992) *Nucleic Acids. Res.* **20:** 5719–5727.
45. Tian, M. and Maniatis, T. (1993) *Cell* **74:** 105–114.
46. Toniolo, D., Persico, M. and Alcalay, M. (1988) *Proc. Natl Acad. Sci. USA* **85:** 851–855.
47. Valcarcel, J., Singh, R., Zamore, P.D. and Green, M.R. (1993) *Nature* **362:** 171–175.
48. Warner, J.R. (1987) *Genes Dev.* **1:** 1–3.
49. Waterston, R., Martin, C., Craxton, M., Huynh,C., Coulson,A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, R., Wilson, R., Berks, M., Du, Z., Thomas, J., Thierry-Mieg, J. and Sulston, J. (1992) *Nature Genet.* **1:** 114–123.
50. Zachar, Z., Chou, T-B. and Bingham, P.M. (1987) *EMBO J.* **6:** 4105–4111.
51. Zachar, Z., Chou, T-B., Kramer, J., Mims, I.P. and Bingham, P.M. (1994). *Drosophila. Genetics* **136**, 139–150.
52. Zachar, Z., Kramer, J., Mims, I. and Bingham, P.M. (1993) *J. Cell Biol.* **121:** 729–742.
53. Zahler, A.M., Neugebauer, K.M., Stolk, J.A. and Roth, M.B. (1993) *Science* **260:** 219–222.