Autoregulation at the level of mRNA 3′ end formation of the *suppressor of forked* gene of *Drosophila melanogaster* is conserved in *Drosophila virilis*

AGNÈS AUDIBERT* AND MARTINE SIMONELIG†‡

Dynamique du Génome et Evolution, Institut Jacques Monod, Université Denis Diderot, 2, place Jussieu, 75005 Paris, France

ABSTRACT The *Drosophila melanogaster* Suppressor of forked [Su(f)] protein shares homology with the yeast RNA14 protein and the 77-kDa subunit of human cleavage stimulation factor, which are proteins involved in mRNA 3′ end formation. This suggests a role for Su(f) in mRNA 3′ end formation in *Drosophila*. The *su(f)* gene produces three transcripts; two of them are polyadenylated at the end of the transcription unit, and one is a truncated transcript, polyadenylated in intron 4. Using temperature-sensitive *su(f)* mutants, we show that accumulation of the truncated transcript requires wild-type Su(f) protein. This suggests that the Su(f) protein autoregulates negatively its accumulation by stimulating 3′ end formation of the truncated *su(f)* RNA. Cloning of *su(f)* from *Drosophila virilis* and analysis of its RNA profile suggest that *su(f)* autoregulation is conserved in this species. Sequence comparison between *su(f)* from both species allows us to point out three conserved regions in intron 4 downstream of the truncated RNA poly(A) site. These conserved regions include the GU-rich downstream sequence involved in poly(A) site definition. Using transgenes truncated within intron 4, we show that sequence up to the conserved GU-rich domain is sufficient for production of the truncated RNA and for regulation of this production by *su(f)*. Our results indicate a role of *su(f)* in the regulation of poly(A) site utilization and an important role of the GU-rich sequence for this regulation to occur.

Polyadenylation is an important step in the processing of most eukaryotic mRNA. Poly(A) site choice or efficiency can be regulated, and such regulations can lead to changes in gene expression.

To gain insight into the mechanism of polyadenylation and its regulation *in vivo*, we are studying the *suppressor of forked* [*su(f)*] gene of *Drosophila melanogaster*. Genetic and molecular data are consistent with a role of *su(f)* in mRNA 3′ end formation (1, 2). Such a role for *su(f)* also is suggested by homology between the Su(f) protein and proteins involved in mRNA 3′ end formation in yeast and in humans. The Su(f) protein shows 26% identity and 47% similarity with the yeast RNA14 protein (3), which is part of cleavage factor I (CFI) (4). In yeast as in mammalian cells, the 3′ end processing reaction requires several complexes and proceeds in two main steps: the cleavage of pre-mRNA and the addition of poly(A) to the newly generated 3′ end (5–7). Yeast CFI has a role in both steps of the reaction. The Su(f) protein is also 56% identical and 69% similar to the 77-kDa human protein, a subunit of cleavage stimulation factor (CstF) (8). CstF is required for cleavage of pre-mRNA but not for polyadenylation of the cleaved molecule. In HeLa cells, CstF consists of three subunits of 77 kDa, 64 kDa, and 50 kDa (9). CstF binds GU-rich sequences located downstream of the cleavage site in the pre-mRNA and it does so through the 64-kDa protein (10). The 77-kDa protein makes a bridge between the two other subunits of CstF (8). This protein also interacts with a 160-kDa subunit of another complex, cleavage and polyadenylation specificity factor (CPSF) (11), which binds to the polyadenylation signal (AAUAAA), located upstream of the cleavage site. CPSF is required for both cleavage of pre-mRNA and polyadenylation. Cooperative binding of CPSF and of CstF to the pre-mRNA results from interactions between CstF and CPSF and allows the cleavage site to be defined.

The high level of similarity that extends over the entire length of Su(f) and human 77-kDa proteins suggests that the Su(f) protein is part of a CstF complex in *Drosophila*. This hypothesis is reinforced by the presence in *Drosophila* of a 64-kDa homologous protein and by interaction *in vitro* between this *Drosophila* protein and human 77-kDa protein (8).

The *su(f)* gene produces three polyadenylated transcripts resulting from the utilization of alternative poly(A) sites (3) (Fig. 1*A*). Two of these transcripts, which are 2.6 and 2.9 kb long, are polyadenylated at the end of the transcription unit and encode the 84-kDa Su(f) protein. The third transcript is 1.3 kb long and is polyadenylated within intron 4. This transcript is dispensable for *su(f)* function since the construct *WP10*, able to produce only the 1.3-kb RNA, is unable to rescue any *su(f)* phenotype, whereas the construct *WG8.4*, able to produce only the 2.6-kb and 2.9-kb mRNAs, rescues the lethality of null *su(f)* mutants (12). Using different temperature-sensitive (ts) *su(f)* mutants in Northern blots, we found that, at a restrictive temperature, the lack of *su(f)* function is correlated with the disappearance of the 1.3-kb *su(f)* RNA (ref. 3; Fig. 6). The accumulation of this 1.3-kb RNA is rescued by the construct *WG8.4*, able to encode the wild-type 84-kDa Su(f) protein (K. Elliott, C. Williams, K. O'Hare, and M.S., unpublished data). This indicates that the Su(f) protein is required for accumulation of the 1.3-kb *su(f)* RNA. Given the probable role of *su(f)* in mRNA 3′ end formation, these data suggest that the Su(f) protein is involved in 3′ end processing of the 1.3-kb *su(f)* RNA and that 3′ end formation of this RNA is particularly sensitive to *su(f)* activity. This negative autoregulatory loop would serve to control the amount of the 84-kDa protein by stimulating the formation of a truncated RNA and thus reducing the amount of the coding 2.6-kb and 2.9-kb mRNAs (Fig. 1*A*).

Genetics: Audibert and Simonelig

Proc. Natl. Acad. Sci. USA 95 (1998) 14303

We have used an evolutionary approach to examine the importance of *su(f)* autoregulation and to identify sequences potentially involved in this autoregulation. We have cloned the *su(f)* gene of *D. virilis* and analyzed its RNA profile. Our results show that the *su(f)* coding sequence is highly conserved between *D. melanogaster* and *D. virilis*. In addition, *D. virilis su(f)* also produces a polyadenylated transcript truncated in intron 4. Sequence comparison of this intron from both species allows us to point out three blocks of conserved sequences, one of which corresponds to the GU-rich sequence downstream of the poly(A) site. Moreover, production of the 1.3-kb RNA in *D. melanogaster*, from a construct including *su(f)* 5′ sequence up to this GU-rich sequence in intron 4, depends on Su(f) wild-type protein.

Taken together, our results suggest a negative autoregulation of *su(f)* at the level of 3′ end formation and a conservation of this autoregulation in *D. virilis*. Therefore, this negative autoregulatory loop seems to be of functional importance, possibly to regulate the amount of Su(f) protein. Our study also reveals an important role of the GU-rich sequence in the regulation of poly(A) site utilization by *su(f)* activity.

MATERIALS AND METHODS

***Drosophila* Stocks and Transformation.** The *su(f)^{L26}* allele is described by Tudor *et al.* (13). *su(f)^{ts67g}* and *su(f)^{ts726}* are ts alleles described in Lindsley and Zimm (14). The *WG8.4* and *WP10* constructs and their genetic properties are described by Simonelig *et al.* (12). *P*-element transformation was carried out as described by Rubin and Spradling (15). Construct DNA (500 μg/ml) with 250 μg/ml of the helper plasmid, pUChsΠΔ2–3, were injected into *w^{1118}* embryos.

Constructs. The *WP24* and *WP13* constructs were generated from the *WP10* construct. *WP10* has an *Xba*I-*Xho*I genomic fragment from the 5′ region of *su(f)* up to exon 6, cloned into the pW8 vector (12). Deletions of *WP10* from exon 6 to intron 4 were carried out by using exonuclease III. The precise location of the truncation in intron 4 was determined by sequencing.

Isolation and Characterization of *D. virilis* Recombinant Phages. An *Mbo*I genomic *D. virilis* library in λEMBL3 (gift of R. Blackman, Cambridge, MA) was screened at low stringency with a probe corresponding to a fragment of the *D. melanogaster su(f)* locus, which spans exon 2 to exon 6. Hybridization was at 42°, in 2× SSC/35% of formamide/5× Denhardt's solution/100 μg/ml sonicated salmon sperm DNA. Washes were at 50°, in 1× SSC/0.1% SDS. Four positive phages were recovered and characterized by restriction mapping and Southern blotting. Three of these phages were identical and contain the complete *D. virilis su(f)* locus. One of them, λV9, was selected for further analysis. Four fragments of this phage were subcloned into pBluescript and sequenced. Sequences of junctions between the different subclones were determined by direct sequencing of PCR products amplified from *D. virilis* genomic DNA and/or from λV9. PCR amplifications were carried out by using the following oligonucleotides: [4000–4021], [4514–4494], [2586–2606], [2965–2984], [3449–3430]; coordinates are from the *D. virilis* sequence in GenBank (accession no. AF097830). Sequence data were analyzed with the BESTFIT and COMPARE programs of the GCG package. Intron scores were calculated with the SIGNALX program of the SQX package, using the SENAPATHY matrix.

RNA Blots, RNase Protection Assays, and Reverse Transcription–PCR (RT-PCR). RNA blots were as reported by Simonelig *et al.* (12). The RNase protection assays were performed as described by Neel *et al.* (16). Total RNA (8 μg) was used in each assay. The antisense probe transcribed with T7 RNA polymerase was complementary to a part of exon 4 and a part of intron 4, from nucleotides 4707 to 4970. After hybridization and RNase digestion, protected fragments with

sizes of 207 nt (exon 4 and intron 4) for the 1.3-kb RNA and 93 nt for the 2.6-kb and 2.9-kb RNAs were produced. Loading was controlled by independent RNase protection assays by using an *rp49* probe and 1 μg of each RNA. RT-PCR was carried out on total RNA extracted from 100 mg of *D. virilis* adults. To determine the boundaries of intron 4, 1% of the RNA was used in a 50-μl PCR in 1× buffer II (Perkin–Elmer/Cetus) containing 1.5 mM MgCl$_2$, 10 mM dNTPs, 50 pmol of oligonucleotides [2658–2633] and [1429–1514 without intron 2]. After 3 min at 94°, 1 unit of AMV-RT (avian myoblastoma virus–reverse transcriptase) was added for 20 min at 55°, followed by 3 min at 94° and the addition of 2 units of *Taq* polymerase. Two microliters of this PCR was used as a template in a second 50-μl PCR using oligonucleotides [2658–2633] and [2122–2146]. The PCR product was directly sequenced. For 3′ rapid amplification of cDNA ends–PCR, 1/40 of the RNA was denatured at 65° for 3 min and put in a 20-μl reaction with 1× RT buffer (Boehringer Mannheim), 1 mM dNTPs, 50 pmol of oligonucleotide 5′-CGTGTCGGAAT-TCACTTA(T)$_{18}$ [oligo(T)-adapter], and 10 units of AMV-RT for 2 hr at 41°. Two successive 50-μl PCRs were performed by using specific primers listed below (50 pmol) and oligonucleotide 5′-CGTGTCGGAATTCACTTATT (adapter) (50 pmol) in 1× buffer II (Perkin–Elmer/Cetus) with 1.5 mM MgCl$_2$/10 mM dNTPs/2 units of *Taq* polymerase. For the 3′ end of *D. virilis* 1.3-kb RNA, oligonucleotides were [1429–1514 without intron 2] and [2122–2146]. For the 3′ end of *D. virilis* 2.6-kb mRNA, oligonucleotides were [4000–4021], and [4661–4683] or [4948–4967]. To determine the 3′ ends of RNAs produced by *WP13* and *WP24*, total RNA from *w^{1118} su(f)^{L26}/Y;WG8.4/+;WP24/+* or *w^{1118} su(f)^{L26}/Y;WG8.4/+;WP13/+* males were used to generate the cDNA pool. Primers specific to *D. melanogaster su(f)* sequence were [4426–4509 without intron 3] and [4707–4728 with the sequence 5′-CGGGATC to create a *Bam*HI site]. Coordinates are from the *D. melanogaster* sequence in GenBank (accession no. X62679). The PCR products were cloned into pBluescript and sequenced.

RESULTS

Molecular Organization of the *su(f)* Locus of *D. virilis*. To clone the *D. virilis su(f)* gene, a genomic library of *D. virilis* was screened at low stringency. One clone, λV9, covering the whole *su(f)* locus, was used to sequence a total of 5.5 kb. The overall sequence identity between *D. virilis* and *D. melanogaster su(f)* genes is 76.2%. The coding regions show a very high degree of similarity; this allowed the determination of intron–exon boundaries of the *D. virilis* gene by comparison with the *D. melanogaster* sequence. Boundaries of intron 4 were confirmed by sequencing PCR-amplified DNA using cDNA synthesized from total *D. virilis* RNA as a template. The number of introns is conserved, but a comparison of intron sequences shows that they have diverged almost completely. One exception is within intron 4, where several islands of striking conservation are detected over 137 bp (see below). No significant homology was detected upstream of the initiation codon and downstream of the coding sequence between *D. melanogaster* and *D. virilis*.

The Su(f) Proteins from *D. virilis* and *D. melanogaster* Are Highly Conserved. The protein sequences are 96.3% identical and 97.7% similar over their entire length (Fig. 1*B*). The predicted *D. virilis* Su(f) protein contains 737 aa, which is 4 aa longer than the *D. melanogaster* protein. Apart from the 4 aa lacking in the *D. melanogaster* protein, only 25 aa are different between both species and only 2 changes are nonconservative. Much of the difference between the proteins lies within a single region of 20 aa (positions 561–581). This region contains the 4 aa found exclusively in the *D. virilis* protein and also shows 6 aa differences between the two proteins. The N-terminal two-thirds of the *D. melanogaster* Su(f) protein contains eight repeats similar to tetratricopeptide repeat motifs (7), which are
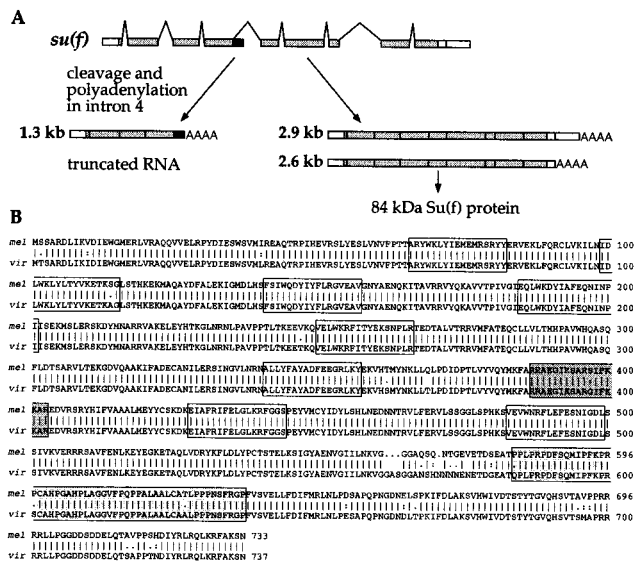
A

su(f)

cleavage and
polyadenylation
in intron 4

1.3 kb ◻▭▭▭AAAA   2.9 kb ◻▭▭▭▭▭▭AAAA

truncated RNA   2.6 kb ◻▭▭▭▭▭AAAA

84 kDa Su(f) protein

B

```
mel  MSSARDLIKVDIEWGMERLVRAQQVVELRFYDIESWSVMIREAQCTRPIHEVRSLYESLVNVFPTTARYWKLYIIEMEMRSRYYERVEKLFQRCLVKILNID  100
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
vir  MTSARDLIKIDIEWGMERLVRAQQVVELRFYDIESWSVMLREAQTRPIHEVRSLYESLVNVFPTTARYWKLYIIEMEMRSRYYERVEKLFQRCLVKILNID  100

mel  LWKLYLTVVKETKSGLSTHKEKMAQAYDFALEKIGMDLHSFSIWQDYIYFLRGVEAVGNYAENQKITAVRRVVQKAVVTPIVGIEQLWKDYIAFEQNINF  200
     ||||||||||||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
vir  LWKLYLTVVKETKAGLSTHKEKMAQAYDFALEKIGMDLHSFSIWQDYIYFLRGVEAVGNYAENQKITAVRRVVQKAVVTPIVGIEQLWKDYIAFEQNINF  200

mel  ISEKMSLERSKDYMNARRVAKELEYHTKGLNRNLPAVPPTLTKEEVKQVELWKRFITYSKSNPLRTEDTALVTRRVMFATEQCLLVLTKHPAVMHQASQ  300
     |||||||||||||||||||||||||||||||||||||||||| ||||||||||| ||||||||| |||||||||||||||||||||||||||||||||||
vir  ISEKMSLERSKDYMNARRVAKELEYHTKGLNRNLPAVPPTLTKEETKQVELWKRFITYEKSNPLRTEDTALVTRRVMFATEQCLLVLTKHPAVMHQASQ  300

mel  FLDTSARVLTEKGDVQAAKIFADECANILERSINGVLNRNALLYFAYADFEEGRLKVEKVHTMYNKLLQLPDIDPTLVYVVGYMKFARSASOIKSARSIFK  400
     ||||||||||||||||||||||||||||||| |||||||| ||||||||||| ||||||||||||||||||||||||||||||
vir  FLDTSARVLTEKGDVQAAKIFADECANILERSINGVLNRNALLYFAYADFEEGRLKYEKVHSMYNKLLQLPDIDPTLVYVVGYMKFARHAAGIRSARAGIFK  400

mel  KARSEDVRSRYHIFVAAALKEYYCSKDNEIAFRIPLGLRKFGGSPEYVMCYIDYLSHLNEDNNTRVLFERVLSSGGLSPHKSVEVWNRFLEFESNIGDLE  500
     |||                 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
vir  KARSEDVRSRYHIFVAAALKEYYCSKDNEIAFRIPLGLRKFGGSPEYVMCYIDYLSHLNEDNNTRVLFERVLSGGGLSPHKSVEVWNRFLEFESNIGDLE  500

mel  SIVKVERRRSAVFENLKEYEGKETAQLVDRYKFLDLYPCTSTELKSIGYAENVGIILNKVG...GGAQSQ.NTGEVETDSEATPPLRPDFSQMIPFKPR  596
     .|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||     |||||  |||||||||||||||||||||||||||
vir  SIVKVERRRSAVFENLKEYEGKETAQLVDRYKFLDLYPCTSTELKSIGYAENVGIILNKVGGASGGANSHDNMNENETDGEATQPLRPDFSQMIFFKPR  600

mel  PCANFPGAHPLAGGVFPQPPALAALCATLPPPNSFRGPFVSVELLFDIFMRLNLPDSAPQPNGDNELSPRKIFDLAKSVHWIVDTSTYTGVQHSVTAVPPRR  696
     .|||||||||||||||||||||||||||||||||||| |||||||||||||||| | ||||||||||||||||||||||||||||||| ||||| ||||
vir  SCAHPGAHPLAGGVFPQPPALAALCAALPPPNSFRGPFVSVELLFDIFMRLNLPESAPQPNGDNDLTPKIFDLAKSVHWIVDTSTYTGVQHSVTSMAPRR  700

mel  RRLLPGGDDSDDELQTAVPPSHDIYRLRQLKRFAKSN  733
     |||||||||||||||||||||||||||||||||||||
vir  RRLLPGGDDSDDELQTSAPPTNDIYRLRQLKRFAKSN  737
```

FIG. 1. Schematic representation of the *su(f)* gene and comparison of Su(f) proteins from *D. melanogaster* (*mel*) and *D. virilis* (*vir*). (*A*) The *su(f)* locus is depicted. Shaded boxes are coding sequences, open boxes are noncoding sequences, and the solid box is the part of intron 4 that is incorporated in the 1.3-kb RNA. The model for *su(f)* autoregulation proposes that the 84-kDa Su(f) protein, encoded by the full-length *su(f)* transcripts, stimulates utilization of the poly(A) site in intron 4, leading to formation of a truncated transcript. (*B*) Comparison of Su(f) proteins from *D. melanogaster* and *D. virilis*. Vertical lines indicate identical residues, colons represent similarities, and single dots represent less similar amino acids. Open boxes are tetratricopeptide-like repeats, the darkly shaded box indicates the nuclear localization signal, and the lightly shaded box is the proline-rich domain.

known to be capable of mediating protein–protein interactions. These repeats are perfectly conserved in *D. virilis* (Fig. 1*B*), except for a single conservative change. The protein also contains a bipartite nuclear localization signal (positions 387–403), which shows only one conservative change in *D. virilis*. A proline-rich domain (positions 580–633) is present toward the C terminus of the protein, in which 16 residues among 54 are prolines (3). This domain is conserved in the *D. virilis* protein, where 14 prolines are found at the same positions. Such an extensive conservation over the entire coding region (except for a 20-aa domain) probably reflects an important contribution of most parts of the protein in *su(f)* function.

mRNA Profile of *su(f)* in *D. virilis* and *D. melanogaster*. To determine whether the *su(f)* autoregulation is conserved in *D. virilis*, we analyzed its mRNA profile in this species. In *D. melanogaster*, *su(f)* produces three polyadenylated RNAs that are 2.9, 2.6, and 1.3 kb in length (Fig. 1*A*). RNA blots were carried out with RNA from *D. virilis* and *D. melanogaster* adults, using a specific probe for each species (Fig. 2). Two RNAs were detected in *D. virilis*, a major 2.6-kb RNA and a minor 1.3-kb RNA.

The 3′ ends of the *D. virilis* 1.3-kb and 2.6-kb RNAs were mapped by using 3′ rapid amplification of cDNA ends–PCR. cDNAs were synthesized by using an oligo(T) primer, and
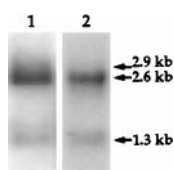
FIG. 2. mRNA profile of *su(f)* in *D. melanogaster* and in *D. virilis*. Northern blot of poly(A)⁺ RNA from adult *D. melanogaster* (lane 1) and *D. virilis* (lane 2). RNA probes of *su(f)* from each species were hybridized to blotted RNA from the corresponding species.

these cDNAs were used as a template for two rounds of PCR using two sets of nested primers. PCR products corresponding to the 1.3-kb RNA were cloned, and 13 independent clones were sequenced. Among these 13 clones, we identified three different 3′ ends: at positions 2280, 2283, and 2289 within intron 4 (Fig. 3*A*). In each case, the cleavage site has the same sequence, CA. One of the clones is polyadenylated at position 2280, and six are polyadenylated at each of the other sites. In *D. melanogaster*, the 1.3-kb RNA is polyadenylated at a single site at position 4913 within intron 4 (3). The length of intron 4 sequence that is incorporated in the 1.3-kb RNA is similar in both species (112 bp in *D. melanogaster* versus 93, 96, and 102 bp in *D. virilis*). These data indicate that, as in *D. melanogaster*, *su(f)* produces a 1.3-kb RNA polyadenylated in intron 4 in *D. virilis*. Therefore, alternative polyadenylation also is used in this species to produce different *su(f)* mRNAs. PCR products corresponding to the 2.6-kb mRNA 3′ end were generated in two independent experiments, using two different primers for the second round of PCR, located either in the coding sequence or in the 3′ untranslated region. PCR products from both experiments were cloned, and six and three independent clones, respectively, were sequenced. We found two different sets of cleavage sites for the 2.6-kb mRNA (Fig. 3*B*). Among the six clones from the first experiment, we identified four different 3′ ends at positions 4932, 4935, 4942, and 4947. Two clones are polyadenylated at each of sites 4932 and 4947, and one clone is polyadenylated at each of sites 4935 and 4942. Among the three clones from the second experiment, we identified two different 3′ ends at positions 5026 and 5031. In *D. melanogaster*, the 3′ end of the 2.6-kb mRNA was determined by sequencing of a single cDNA and is located at position 7076 (Fig. 3*B*) (3).

Conserved Features of Intron 4 and of the 1.3-kb RNA. Production of the 1.3-kb RNA results from a competition between splicing of intron 4 and 3′ end formation within intron 4. Therefore, we examined splice junctions of intron 4 in both species (Fig. 4*B*). Intron 4 boundaries in *D. virilis* were determined by sequencing the products of PCR amplification of cDNAs. The 5′ splice site of intron 4 is identical in *D. melanogaster* and *D. virilis* and it has the lowest score among the other splice sites of *su(f)*. This splice site is the only one in *su(f)* to be 100% conserved between both species. This suggests that this poor 5′ splice site is important for the production of *su(f)* mRNAs, possibly in controlling the kinetics of splicing of intron 4. In *D. melanogaster*, the 1.3-kb RNA is unusual in that it lacks an encoded stop codon. Therefore, the ORF continues to be open for translation into the 3′ poly(A). This RNA could encode a 39-kDa protein that would have the same first 313 aa as the 84-kDa protein, then 38 aa encoded by the beginning of intron 4, and a polylysine tract encoded by the poly(A). This putative protein is dispensable for *su(f)* function (12). Nevertheless, the 1.3-kb RNA is produced in the two species. We have compared the amino acid sequences encoded by intron 4 that would correspond to the C terminus of the putative 39-kDa protein. Fig. 4*A* shows that this region encoded by intron 4 is not conserved between the two species. This result suggests that this C-terminal domain is not required for *su(f)* function and corroborates the fact that the putative 39-kDa protein is dispensable for *su(f)* function in *D. melanogaster*. The only conserved feature between the 1.3-kb RNA of both species is that they both lack an encoded stop codon (Fig. 4*A*). It therefore is possible, given the unusual structure of this RNA, that it would not be translated. These results show that it is not the coding capacity of the 1.3-kb RNA that is conserved, but its production. This suggests that synthesis of this truncated RNA is important, possibly to regulate the amount of the larger mRNAs encoding the Su(f) protein.

Conserved Sequences in Intron 4. A consensus poly(A) signal AAUAAA is not present upstream of the cleavage sites in intron 4 of *su(f)* from *D. melanogaster* and *D. virilis*. Thus,
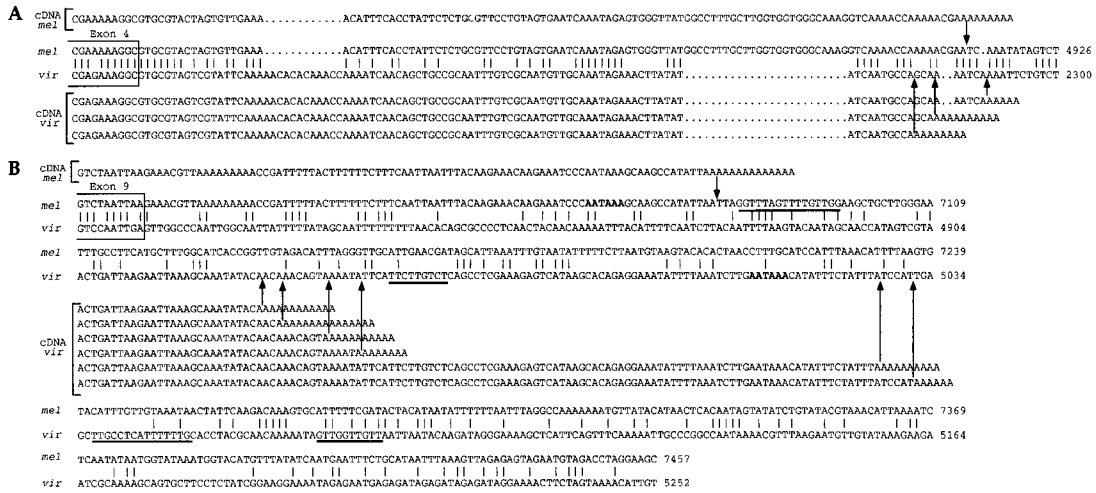
A

```
cDNA  CGAAAAAGGCGTGCGTACTAGTGTTGAAA...........ACATTTCACCTATTCTCTGCGTTCCTGTAGTGAATCAAATAGAGTGGGTTATGGCCTTTGCTTGGTGGTGGGCAAAGGTCAAAACCAAAAACGAAAAAAAAA
mel
mel  CGAAAAAGGCGTGCGTACTAGTGTTGAAA...........ACATTTCACCTATTCTCTGCGTTCCTGTAGTGAATCAAATAGAGTGGGTTATGGCCTTTGCTTGGTGGTGGGCAAAGGTCAAAACCAAAAACGAATC.AAATATAGTCT 4926
     ||| |||||| |||||||||||| | || ||| ||           ||  ||  ||||| || ||| |  ||| || | | | | ||||||| ||||        |||| ||| || |||| ||||  |||||
vir  CGAGAAAGGCGTGCGTAGTCGTATTCAAAAACACACAAACCAAAATCAACAGCTGCCGCAATTTGTCGCAATGTTGCAAATAGAAACTTATAT...............ATCAATGCCAGCAA..AATCAAAATTCTGTCT 2300
cDNA CGAGAAAGGCGTGCGTAGTCGTATTCAAAAACACACAAACCAAAATCAACAGCTGCCGCAATTTGTCGCAATGTTGCAAATAGAAACTTATAT...............ATCAATGCCAGCAAAAAAAAAAA
vir  CGAGAAAGGCGTGCGTAGTCGTATTCAAAAACACACAAACCAAAATCAACAGCTGCCGCAATTTGTCGCAATGTTGCAAATAGAAACTTATAT...............ATCAATGCCAAAAAAAAAA
```

B

```
cDNA  GTCTAATTAAGAAACGTTAAAAAAAAAACCGATTTTTACTTTTTTCTTTCAATTAATTTACAAGAAACAAGAAATCCCAATAAAGCAAGCCATATTAAAAAAAAAAAAAAA
mel
mel  GTCTAATTAAGAAACGTTAAAAAAAAAACCGATTTTTACTTTTTTTCTTTCAATTAATTTACAAGAAACAAGAAATCCCAATAAAGCAAGCCATATTAATTAGGTTTAGTTTTGTTGGAAGCTGCTTGGGAA 7109
     ||| ||||| |||| ||            : .| |: ::.: : .... ...|||
vir  GTCCAATTGAGTTGGCCCAATTGGCAATTATTTTTATAGCAATTTTTTTTTAACACAGCGCCCTCAACTAACAACAAAAATTTACATTTTCAATCTTACAATTTAAGTACAATAGCAACCATAGTCGTA 4904

mel  TTTGCCTTCATGCTTGGCATCACCGGTTGTAGACATTTAGGGTTGCATTGAACGATAGCATTAAATTTGTAATATTTTCTTAATGTAAGTACACACTAACCTTTGCATCCATTTAAACATTTTAAGTG 7239
     || |       |||   ||| | |           |   || |||  ||| | | |  |  |||      || |
vir  ACTGATTAAGAATTAAAGCAAATATACAACAAACAGTTAAATATTCATTCTTTGTCTCAGCCTCGAAAGAGTCATAAGCACAGAGGAAATATTTTAAATCTTGAATAAACATATTTCTATTTATTCCATTGA 5034

cDNA  ACTGATTAAGAATTAAAGCAAATATACAAAAAAAAAA
vir   ACTGATTAAGAATTAAAGCAAATATACAACAAAAAAAAAAAAAA
      ACTGATTAAGAATTAAAGCAAATATACAACAAACAGTAAAAAAAAAAA
      ACTGATTAAGAATTAAAGCAAATATACAACAAACAGTAAAATAAAAAAA
      ACTGATTAAGAATTAAAGCAAATATACAACAAACAGTAAAATATTCATTCTTTGTCTCAGCCTCGAAAGAGTCATAAGCACAGAGGAAATATTTTAAATCTTGAATAAACATATTTCTATTTAAAAAAAAA
      ACTGATTAAGAATTAAAGCAAATATACAACAAACAGTAAAATATTCATTCTTTGTCTCAGCCTCGAAAGAGTCATAAGCACAGAGGAAATATTTTAAATCTTGAATAAACATATTTCTATTTATTCCATTAAAAAA

mel  TACATTTGTTGTAAATAACTATTCAAGACAAAGTGCATTTTCGATACTACATAATATTTTTTAATTTAGGCCAAAAAAATGTTATACATAACTCACAATAGTATATCTGTATACGTAAACATTAAAATC 7369
                          |  || |||||   || |   ||| | | | |        |   |    || ||| ||       |       | |           |   | | |||
vir  GCTTGCCTCATTTTTTGCACCTACGCAACAAAATAGTTGGTTGTTAATTAATACAACGATAGGGAAAAGCTCATTCAGTTTCAAAAATTGCCCGGCCAATAAAACGTTTAAGAATGTTGTATAAAGAAGA 5164

mel  TCAATATAATGGTATAAATGGTACATGTTTATATCAATGCAATTTCTGCATAATTTAAAGTTAGAGAGTAGAAGTGTAGACCTAGGAAGC 7457
                                   || ||       |       |  |||
vir  ATCGCAAAAGCAGTGCTTCCTCTATCGGAAGGAAAATAGAGAATGAGAGATAGAGATAGAGATAGGAAAACTTCTAGTAAAACATTGT 5252
```

FIG. 3.   3′ ends of *su(f)* 1.3-kb and 2.6-kb mRNAs. Sequences of cDNAs corresponding to the 1.3-kb (*A*) and to the 2.6-kb (*B*) RNAs from *D. melanogaster* (*mel*) and from *D. virilis* (*vir*). In *B*, AAUAAA signals are in bold and GU-rich sequences are underlined.

comparison of intron 4 from the two species could point out sequences required for utilization of this poly(A) site. Sequence comparison of this intron reveals that a region of 137 bp shows a high degree of conservation (75.8% identity) between the two species (Fig. 4*B*). This conserved region is located in both species just downstream of the 3′ end of the 1.3-kb RNA (Fig. 4*B*). Two distinct domains can be identified in this conserved region. The first domain is GU-rich; it corresponds to the GU-rich sequence located downstream of poly(A) sites, which is required for poly(A) site definition. In *D. virilis,* this domain is 17 bp long, which is 2 bp longer than the homologous domain in *D. melanogaster*. The remaining 15 bp are identical (Fig. 4*B*). The second domain (called the downstream domain) is about 100 bp long and contains two highly conserved regions (90% identity) of 30 bp and 40 bp, respectively, separated by a 30-bp nonconserved region (Fig. 4*B*). Comparison with sequences in the databases did not identify other sequences homologous to this domain. This region could form a stem–loop structure, in which part of the conserved regions would form the stem and the internal nonconserved region would form the loop. To determine whether the downstream domain has a role in the 1.3-kb RNA 3′ end formation, we made transgenes containing the *su(f)* locus of *D. melanogaster* up to intron 4 and truncated at two different positions within intron 4 (Fig. 4*B*). *WP24* is truncated just downstream of the GU-rich sequence and does not contain the downstream domain. In contrast, *WP13* contains both the GU-rich and the downstream domains. The constructs were introduced into *D. melanogaster* by *P*-mediated transformation. Two *WP24* and two *WP13* transformants were used in Northern blots to determine whether the transgenes can produce the 1.3-kb RNA. Transformants were crossed with the

stock *su(f)*^L26^*; WG8.4*. The *su(f)*^L26^ allele is a deletion of the *su(f)* locus, and *WG8.4* is a transgene containing the whole *su(f)* locus, but lacking the first five introns of the gene. Thus, in the stock *su(f)*^L26^*; WG8.4*, only the 2.6-kb and 2.9-kb mRNAs of *su(f)* are produced (ref. 12, Fig. 5). Fig. 5 shows that both *WP24* and *WP13* transgenes do produce the 1.3-kb RNA. We determined 3′ ends of this 1.3-kb RNA produced by the transgenes using 3′ rapid amplification of cDNA ends–PCR. For each transgene, PCR products were cloned and six independent clones were sequenced. All six clones generated from the *WP13* construct, which contains the GU-rich and the downstream domains, are polyadenylated at the wild-type poly(A) site at position 4913. Among the six clones generated from the *WP24* construct, which has the GU-rich domain only, five are polyadenylated at the wild-type site at position 4913 and one is polyadenylated at position 4917. Therefore, the 1.3-kb RNA is produced by both constructs and the downstream domain does not appear to be absolutely required for the production of the 1.3-kb RNA.

The conservation of the GU-rich sequence in intron 4 between *D. melanogaster* and *D. virilis* suggests that this sequence could have an important role in the regulation of this poly(A) site utilization by *su(f)* activity. Therefore, we tested whether sequence up to this poly(A) site, including the GU-rich sequence, is sufficient for regulation by *su(f)*. *su(f)* autoregulation is shown in Fig. 6*A*. In the *su(f)*^ts67g^ allele at restrictive temperature (25°), accumulation of the 1.3-kb RNA is low (Fig. 6*A*, compare lanes 1 and 4). In addition, a transgene, *WP10*, containing *su(f)* 5′ sequence up to exon 6 (12), is subject to *su(f)* regulation. In wild type, the presence of the *WP10* transgene induces an increase in the accumulation of the 1.3-kb RNA (Fig. 6*A*, compare lanes 1 and 2). In
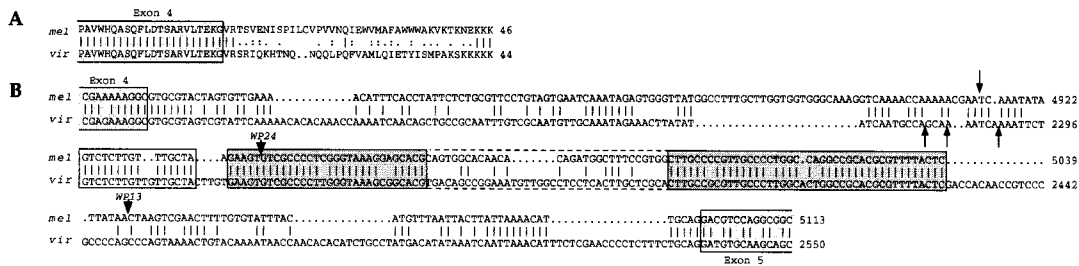
A

```
        Exon 4
mel  PAVWHQASQFLDTSARVLTEKGVRTSVENISPILCVPVVNQIEWVMAFAWWWAKVKTKNEKKK 46
     ||||||||||||||||||||||  ...: .    .: |:  ::.:  : .....||||
vir  PAVWHQASQFLDTSARVLTEKGVRSRIQKHTNQ..NQQLPQFVAMLQIETYISMPAKSKKKKK 44
```

B

```
        Exon 4
mel  CGAAAAAGGCGTGCGTACTAGTGTTGAAA...........ACATTTCACCTATTCTCTGCGTTCCTGTAGTGAATCAAATAGAGTGGGTTATGGCCTTTGCTTGGTGGTGGGCAAAGGTCAAAACCAAAAACGAATC.AAATATA 4922
     ||| |||||| |||||||||| | || ||| ||           ||  ||  ||||| || ||| |  ||| || | | | | ||||||| |||         |||| ||| || |||| |||||
vir  CGAGAAAGGCGTGCGTAGTCGTATTCAAAAACACACAAACCAAAATCAACAGCTGCCGCAATTTGTCGCAATGTTGCAAATAGAAACTTATAT...............ATCAATGCCAGCAA..AATCAAAATTCT 2296
                                                                                                          WP24

mel  GTCTCTTGT..TTGCTA............GGAAGTGTCGCCCTCGGGTAAAGGAGCACGCAGTGCACAACA......CAGATGGCTTTCCGTGCTTTGCCCGTTGCCCCTGGC.CAGGCCGCACGCGTTTTACTC............. 5039
     |||||||| ||||||             ||||||||||||| |||||||||||||||        || || |||||| || | |  ||||||||||||||||||||||||||||||
vir  GTCTCTTGTTGTTGCTAGTTGTGAAGTGTCGCCCCTTGGGTAAAGCGGCACGTGACAGCCGGAAATGTTGGCCTCCTCACTTGCTCGCACTTTGCCCGTTGCCCTTGGGCACTGGCCGCACGCGTTTTACTGCACCACAACCGTCCC 2442
                    WP13

mel  .TTATAACTAAGTCGAACTTTGTGTATTTAC............ATGTTTAATTACTTATTAAAACAT.................TGCAGGACGTCCAGGCGGC 5113
      || ||| ||| ||     ||   |       || ||||  |                      ||||||| ||||
vir  GCCCCAGCCCAGTAAAACTGTACAAAATAACCAACACACATCTGCCTATGACATATAAATCAATTAAACATTTCTCGAACCCCTCTTTCTGCAGGATGTGCAAGCAGC 2550
                                                                                                    Exon 5
```

FIG. 4.   Comparison of *su(f)* intron 4 from *D. melanogaster* (*mel*) and *D. virilis* (*vir*). (*A*) Comparison of amino acid sequences encoded by the part of intron 4 that is included in the 1.3-kb RNA. (*B*) Sequence comparison of *su(f)* intron 4 from *D. melanogaster* and *D. virilis*. Lightly shaded boxes are part of exons, the open box indicates the GU-rich domain, darkly shaded boxes and dashed box show the downstream domain, the conserved regions being in the darkly shaded boxes. Arrows show the 3′ ends of the 1.3-kb RNA. Arrowheads indicate the localization of the ends of the *WP13* and *WP24* constructs.
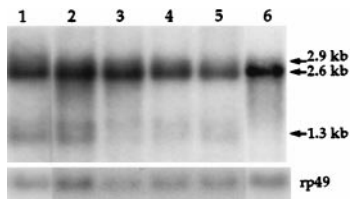
FIG. 5. Production of the 1.3-kb RNA from the WP13 and *WP24* transgenes. Northern blot of poly(A)$^+$ RNA from *D. melanogaster* adults containing the *WP13* or *WP24* constructs. Hybridization was with a *su(f)* RNA probe. The blot was reprobed with the *rp49* clone as a loading control. Lanes: 1, w^{1118}/Y; 2, w^{1118} $su(f)^{L26}$/Y;WG8.4/ +;WP13.1/+; 3, w^{1118} $su(f)^{L26}$/Y;WG8.4/WP13.2; 4, w^{1118} $su(f)^{L26}$/ Y;WG8.4/+;WP24.1/+; 5, w^{1118} $su(f)^{L26}$/Y;WG8.4/WP24.2; 6, w^{1118} $su(f)^{L26}$/Y;WG8.4. The *WG8.4* construct does not produce the 1.3-kb RNA (lane 6); in contrast, both *WP13* and *WP24* constructs do. An additional band is present on this Northern blot, including in the control (lane 1); it is not reproducible and corresponds to nonspecific hybridization of the probe.

contrast, in the $su(f)^{ts67g}$ mutant, raised at 25°, the amount of 1.3-kb RNA produced by *WP10* and the endogenous *su(f)* locus is very reduced (Fig. 6*A*, lane 3). We tested whether production of the 1.3-kb RNA from the *WP13* and *WP24* transgenes is also regulated by *su(f)* activity. We used another ts allele, $su(f)^{ts726}$, that shows a stronger phenotype than $su(f)^{ts67g}$. In wild type, the presence of *WP13* or *WP24* transgenes also induces an increase in the 1.3-kb RNA accumulation (Fig. 6 *B* and *C*, compare lane 1 with lanes 2 and 3). In the $su(f)^{ts726}$ mutant, the 1.3-kb RNA disappears at both restrictive temperatures we have used (25°, Fig. 6*B*, lane 4, and 29°, Fig. 6*C*, lane 4). In the $su(f)^{ts726}$ mutant containing *WP13* or *WP24* transgenes, the 1.3-kb RNA produced by the transgene accumulates at very low levels at 25° (Fig. 6*B*, lanes 5 and 6) and is barely detectable at 29° (Fig. 6*C*, lanes 5 and 6). Results of Northern blots were confirmed by RNase protection assays (Fig. 6*D*), which show that *WP13* and *WP24* transgenes produce a large amount of the 1.3-kb RNA in wild type (lanes 2 and 3), whereas accumulation of this RNA is very low in the $su(f)^{ts726}$ mutant at 29° (lanes 5 and 6). These results show that accumulation of the 1.3-kb RNA from the *WP13* and *WP24* constructs, which are truncated downstream of the poly(A) site within intron 4, requires wild-type Su(f) protein.

DISCUSSION

Conservation of the Su(f) Protein Between *D. melanogaster* and *D. virilis*. The *D. melanogaster* and *D. virilis* Su(f) proteins

show a very high level of identity, 96.3%, which is comparable to that of most conserved proteins between these two species [e.g., 97% identity for the Hsp82 protein (17)]. The human protein homologous to Su(f), 77 kDa, interacts with at least three other proteins, the two other subunits of CstF (8), and the 160-kDa subunit of CPSF (11). Similar interactions probably occur in *Drosophila* between Su(f) and other proteins from CstF and CPSF. These multiple protein–protein interactions could present a strong constraint on the Su(f) protein. Different domains of the protein are likely to be involved in interactions with different proteins, and, indeed, the high level of conservation between *D. melanogaster* and *D. virilis* covers the whole length of the Su(f) protein, except for a short 20-aa region. This less-conserved region might correspond to a hinge between different domains of the protein. It is located between two domains very well conserved, the C-terminal proline-rich domain (3) and the N-terminal two-thirds of the protein, which contains tetratricopeptide-like repeats (7). This short, less-conserved region also is not conserved between the Su(f) and 77-kDa proteins (8).

$su(f)$ Autoregulation in *D. melanogaster* and *D. virilis*. Analysis of the *su(f)* mRNA profile in *D. virilis* indicates that, as in *D. melanogaster*, a 1.3-kb RNA, polyadenylated in intron 4, accumulates in this species. Mapping of the 3′ end of this RNA indicates that polyadenylation can occur at three different positions in intron 4 within 9 bp. That there are multiple cleavage sites is not particular to the poly(A) site within intron 4, since we also have found multiple cleavage sites associated with both poly(A) sites of the 2.6-kb mRNA in *D. virilis*.

In both species, the 1.3-kb RNA does not encode a stop codon; in addition, the sequence encoded by intron 4 shows no homology between the two species. This corroborates the fact that a hypothetical protein encoded by the 1.3-kb RNA is dispensable for *su(f)* function, and we have already proposed that this RNA might remain untranslated (12). These data indicate that the production of the 1.3-kb RNA, but not its coding capacity, is conserved in *D. virilis*. Moreover, we have shown by Northern blots in *D. melanogaster*, using two ts *su(f)* mutants, that accumulation of the 1.3-kb RNA requires wild-type Su(f) protein, since this accumulation strongly decreases in the mutants. This suggests that the Su(f) protein stimulates 3′ end formation of the truncated 1.3-kb RNA. This feedback loop would serve to control accumulation of complete *su(f)* mRNAs and, consequently, the amount of the Su(f) protein. The production of this truncated 1.3-kb RNA is conserved in *D. virilis*, indicating that autoregulation of *su(f)* at the level of 3′ end formation of the 1.3-kb RNA probably also is conserved in *D. virilis*. That this regulatory process occurs in another
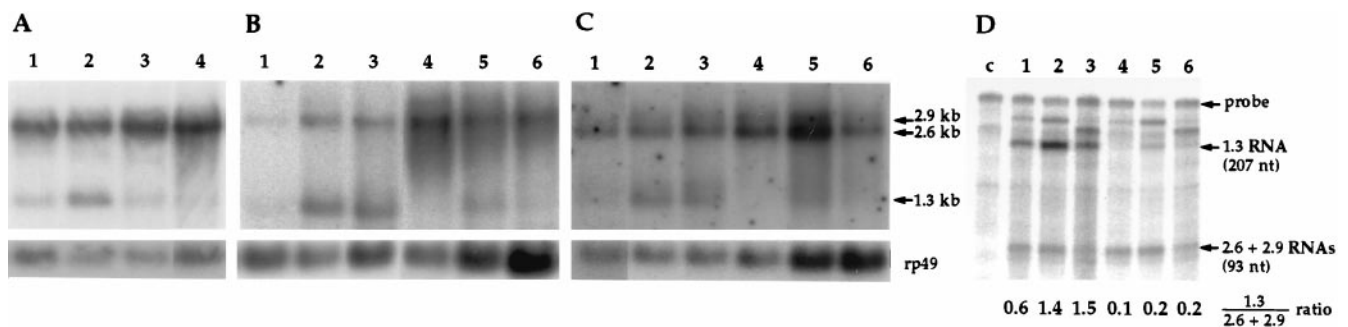

FIG. 6. Autoregulation of *su(f)* and regulation by *su(f)* of poly(A) site utilization from the *WP13* and *WP24* transgenes. (*A–C*) Northern blots of poly(A)$^+$ RNA from wild type and *su(f)* ts mutant adults containing *WP10*, *WP13*, or *WP24* constructs. Hybridization was with a *su(f)* RNA probe. The blot was reprobed with the *rp49* clone as a loading control. (*A*) Flies were raised at 25°. Lanes: 1, w^{1118}; 2, w^{1118};WP10; 3, w^a *f* $su(f)^{ts67g}$;WP10; 4, w^a *f* $su(f)^{ts67g}$. (*B*) Flies were raised at 25°. Lanes: 1, w^a; 2, w^a;WP13/+, 3: w^a;WP24/+; 4, y^2 w^a *f* $su(f)^{ts726}$; 5, y^2 w^a *f* $su(f)^{ts726}$;WP13/+; 6, y^2 w^a *f* $su(f)^{ts726}$;WP24/+. (*C*) Flies were raised at 25° and shifted for 4 days to 29°. Lanes: 1, w^a/Y; 2, w^a/Y;WP13/+; 3, w^a/Y;WP24/+; 4, y^2 w^a *f* $su(f)^{ts726}$/Y; 5, y^2 w^a *f* $su(f)^{ts726}$/Y;WP13/+; 6, y^2 w^a *f* $su(f)^{ts726}$/Y;WP24/+. (*D*) RNase protection assays with total RNA from wild type and $su(f)^{ts726}$ adults raised at 25° and shifted for 4 days to 29°. Positions of the undigested probe and of the protected fragments are indicated. Ratios of the 1.3-kb RNA to the 2.6-kb and 2.9-kb RNAs were obtained by using PhosphorImager scanning (Molecular Dynamics), taking into account the number of labeled residues in each protected fragment. Lane c, control with tRNA (10 μg); lanes 1–6: as in *C*.

Genetics: Audibert and Simonelig

Proc. Natl. Acad. Sci. USA 95 (1998)   14307

Drosophila species, distant from *D. melanogaster*, suggests that the control of the levels of Su(f) protein is important for *Drosophila* development.

Role of *su(f)* and of GU-Rich Downstream Domains in the Regulation of Poly(A) Site Utilization. Several data indicate a role of CstF in the regulation of poly(A) site choice. Poly(A) site definition results from cooperative binding of CPSF and of CstF on both sides of the poly(A) site (5, 7). However, whereas CPSF binds to the conserved upstream polyadenylation signal AAUAAA (18), CstF binds to downstream GU-rich sequences that are highly variable between poly(A) sites (19). Moreover, the stability of the ternary complex pre-mRNA-CPSF-CstF depends on the affinity of CstF for this variable GU-rich sequence. This affinity thus would determine poly(A) site efficiency (20). Until now, shifts in poly(A) site selection have been associated with variations in activity of the 64-kDa subunit of CstF (21–23), which is responsible for the binding of CstF to GU-rich sequences (10). We have shown that utilization of the poly(A) site in intron 4 of *su(f)* depends on the activity of the Su(f) protein. This indicates that variations of another subunit of CstF also can lead to regulation of poly(A) site utilization. Such a role for *su(f)* in the regulation of poly(A) site utilization also is consistent with strong variations in the Su(f) protein level in different *Drosophila* tissues (24).

Comparison of sequences downstream of the poly(A) site in intron 4 between *D. melanogaster* and *D. virilis* reveals that a 137-bp region is conserved. This region consists of a GU-rich sequence just downstream of the intronic poly(A) site and in a more downstream domain composed of two sequences (30 bp and 40 bp) that show 90% identity between *D. melanogaster* and *D. virilis*. We have shown, using the construct *WP24* truncated upstream of this domain, that it is not essential for recognition of the poly(A) site in the absence of a AAUAAA signal. However, this domain could stimulate poly(A) site utilization in intron 4 in the endogenous *su(f)* gene, where 3′ end formation competes with the splicing of intron 4. This downstream domain can form a stem loop and could influence the binding of CstF to the GU-rich sequence; it thus could be involved in modulation of competition between the weak splicing and poly(A) sites in intron 4.

The GU-rich sequence downstream of the poly(A) site in intron 4 is 17 bp in *D. virilis*, and, except for a 2-bp gap, the remaining 15 bp are identical to those in *D. melanogaster*. This could indicate that this particular GU-rich sequence is important for the definition of this poly(A) site, which has no consensus poly(A) signal. We have compared the GU-rich sequences downstream of the poly(A) sites of the 2.6-kb *su(f)* mRNA in *D. melanogaster* and *D. virilis*. One poly(A) site is used for 3′ end formation of the 2.6-kb mRNA in *D. melanogaster*; it has a AAUAAA signal and a GU-rich downstream region (Fig. 3*B*). In *D. virilis*, two sets of cleavage sites, separated by 79 bp, appear to be used for 3′ end formation of the 2.6-kb mRNA. The upstream set of cleavage sites shows no AAUAAA signal and has a short downstream GU-rich sequence; the downstream set of cleavage sites is surrounded by an AAUAAA and two downstream GU-rich sequences (Fig. 3*B*). The sequences of these GU-rich domains are not conserved between *D. melanogaster* and *D. virilis*. This reinforces the significance of sequence conservation between both species of the GU-rich domain downstream of the 1.3-kb RNA poly(A) site. *In vitro* selection experiments (SELEX) were carried out to determine sequence requirements for CstF-RNA interaction. SELEX with the RNA-binding domain of the 64-kDa subunit selected short GU-rich sequence elements without sequence consensus (25), whereas SELEX with purified CstF allowed the selection of sequences that are GU-rich but that also contain A and/or C residues, and that fall into three consensus sequences (26). The GU-rich domain downstream of the poly(A) site in intron 4 does not match any of these three consensus sequences, but it contains A and C residues that are conserved between *D. melanogaster* and *D. virilis*. We have shown that the sequence up to the GU-rich domain is sufficient for utilization of this poly(A) site, and that this site, either including the GU-rich domain only or the GU-rich and the more downstream domains, is very sensitive to a decrease in *su(f)* activity. Regulation of this poly(A) site utilization by *su(f)* could result from the fact that it is a weak poly(A) site that would not allow the formation of a stable polyadenylation complex. In *su(f)* ts mutants, cooperativity of RNA binding by CPSF-CstF [ensured in part by Su(f) since its human homolog contacts proteins within CstF and in CPSF] would be weakened and utilization of this poly(A) site would be reduced. The conservation of the GU-rich sequence, including A and C residues, suggests that a particular sequence for the GU-rich domain is required for this regulation to occur. Our *in vivo* study corroborates previous findings that indicate the role of GU-rich sequences in poly(A) site efficiency.

1. Ishimaru, S. & Saigo, K. (1993) *Mol. Gen. Genet.* **241,** 647–656.
2. O'Hare, K. (1995) *Trends Genet.* **11,** 255–257.
3. Mitchelson, A., Simonelig, M., Williams, C. & O'Hare, K. (1993) *Genes Dev.* **7,** 241–249.
4. Minvielle-Sebastia, L., Preker, P. J. & Keller, W. (1994) *Science* **266,** 1702–1705.
5. Keller, W. (1995) *Cell* **81,** 829–832.
6. Keller, W. & Minvielle-Sebastia, L. (1997) *Curr. Opin. Cell Biol.* **9,** 329–336.
7. Colgan, D. F. & Manley, J. L. (1997) *Genes Dev.* **11,** 2755–2766.
8. Takagaki, Y. & Manley, J. L. (1994) *Nature (London)* **372,** 471–474.
9. Takagaki, Y., Manley, J. L., MacDonald, C. C., Wilusz, J. & Shenk, T. (1990) *Genes Dev.* **4,** 2112–2120.
10. MacDonald, C. C., Wilusz, J. & Shenk, T. (1994) *Mol. Cell. Biol.* **14,** 6647–6654.
11. Murthy, K. G. K. & Manley, J. L. (1995) *Genes Dev.* **9,** 2672–2683.
12. Simonelig, M., Elliott, K., Mitchelson, A. & O'Hare, K. (1996) *Genetics* **142,** 1225–1235.
13. Tudor, M., Mitchelson, A. & O'Hare, K. (1996) *Genet. Res.* **68,** 191–202.
14. Lindsley, D. L. & Zimm, G. G. (1992) in *The Genome of Drosophila Melanogaster* (Academic, San Diego).
15. Rubin, G. M. & Spradling, A. C. (1982) *Science* **218,** 348–353.
16. Neel, H., Weil, D., Giansante, C. & Dautry, F. (1993) *Genes Dev.* **7,** 2194–2205.
17. Blackman, R. K. & Meselson, M. (1986) *J. Mol. Biol.* **188,** 499–515.
18. Keller, W., Bienroth, S., Lang, K. M. & Christofori, G. (1991) *EMBO J.* **10,** 4241–4249.
19. McLauchlan, J., Gaffney, D., Whitton, J. L. & Clements, J. B. (1985) *Nucleic Acids Res.* **13,** 1347–1368.
20. Weiss, E. A., Gilmartin, G. M. & Nevins, J. R. (1991) *EMBO J.* **10,** 215–219.
21. Mann, K. P., Weiss, E. A. & Nevins, J. R. (1993) *Mol. Cell. Biol.* **13,** 2411–2419.
22. Edwalds-Gilbert, G. & Milcarek, C. (1995) *Mol. Cell. Biol.* **15,** 6420–6429.
23. Takagaki, Y., Seipelt, R. L., Peterson, M. L. & Manley, J. L. (1996) *Cell* **87,** 941–952.
24. Audibert, A., Juge, F. & Simonelig, M. (1998) *Mech. Dev.* **72,** 53–63.
25. Takagaki, Y. & Manley, J. L. (1997) *Mol. Cell. Biol.* **17,** 3907–3914.
26. Beyer, K., Dandekar, T. & Keller, W. (1997) *J. Biol. Chem.* **272,** 26769–26779.