The EMBO Journal vol.6 no.13 pp.4105–4111, 1987

Evidence that a regulatory gene autoregulates splicing of its transcript

Zuzana Zachar, Tze-Bin Chou and Paul M.Bingham

Department of Biochemistry, State University of New York, Stony Brook, NY 11794, USA

Communicated by V.Pirrotta

Expression of the presumptive regulatory gene, *suppressor-of-white-apricot* [*su(w^a)*], is controlled at the level of splicing. Results reported here indicate that this control represents autorepression of *su(w^a)* expression. Specifically, reverse genetic studies demonstrate that the 3.5 kb mature *su(w^a)* RNA (produced by removal of seven introns) is a message essential for *su(w^a)$^+$* function and indicate that the abundant 4.4 kb and 5.2 kb mature *su(w^a)* RNAs (resulting when the first or first and second of the seven introns are not removed) are, unexpectedly, byproducts of repression of production of the functional 3.5 kb RNA. Moreover, several experiments indicate that this repression of splices necessary to produce the 3.5 kb RNA is dependent on the translation product of the 3.5 kb RNA itself. We propose that this regulatory gene autoregulates its expression by controlling splicing of its primary transcript.

Key words: regulation/splicing/*Drosophila/suppressor-of-white-apricot*

Introduction

Retrotransposon insertion allele-specific suppressor loci are powerful and convenient experimental systems for analysis of metazoan regulatory genes (see Chou *et al.*, 1987, for details). We report analysis of one of these loci, *suppressor-of-white-apricot* [*su(w^a)*].

Results described in the accompanying paper (Chou *et al.*, 1987) demonstrate that the *su(w^a)* transcript pattern undergoes a developmentally programmed change resulting from post-transcriptional regulation. During the first several hours of development, a 3.5 kb mature RNA is produced by the removal of seven introns from the primary transcript. The structure of this 3.5 kb RNA is that of a conventional message. After the first several hours of development, removal of the first or the first and second of these seven introns largely ceases, resulting in the nearly exclusive production of 4.4 kb and 5.2 kb mature RNA species whose structures are unlike conventional messenger RNAs. (The 3.5 kb RNA continues to be made as a minority species.)

Results reported here indicate that the 3.5 kb RNA is the messenger RNA responsible for *su(w^a)* genetic function and that the 4.4 kb and 5.2 kb RNAs are byproducts of post-transcriptional repression of production of the 3.5 kb RNA. Further, our results indicate that this post-transcriptional repression requires the *su(w^a)* protein itself (3.5 kb RNA translation product). Thus, *su(w^a)*—which apparently regulates a *w^a* RNA processing event in *trans* (see Chou *et al.*, 1987, for a review)—appears to regulate its own expression at the level of RNA processing.

Results

Reverse genetic analysis of functional peptide coding sequences at su(w^a)

Three major *su(w^a)* transcripts result from a developmentally programmed change from splicing to not splicing the first two of seven introns (Figure 1). The smallest mature transcript (the 3.5 kb RNA) has a structure consistent with its being a conventional messenger RNA, while the two larger *su(w^a)* RNAs (4.4 kb and 5.2 kb) have structures unexpected of messenger RNAs (see Figure 6 in Chou *et al.*, 1987, for additional details).

To establish which of these three RNAs is the functional *su(w^a)* messenger RNA we carried out an extensive reverse genetic analysis. Frameshift mutations were introduced at selected positions in the *su(w^a)* locus, these frameshift alleles were reintroduced into the fly and their capacity to engender a *su(w^a)$^+$* eye color phenotype in otherwise *su(w^a)$^-$w^a* individuals was assessed (see Materials and methods for technical details). The results of these experiments are summarized in Figure 1. We conclude that the 3.5 kb RNA is a messenger RNA necessary for *su(w^a)$^+$* function based on the following.

(i)The first exon of the 3.5 kb RNA extends from coordinates 333 to 823 and the first AUG in frame with the open reading frame (ORF) extending through the 3.5 kb RNA is at coordinate 528 (Figure 1). The following frameshift mutations in exon 1 inactivate *su(w^a)$^+$* function as assessed by eye color phenotype (Figure 1): a +10 frameshift at coordinate 593, a +13 frameshift at coordinate 619 and a +11 frameshift at coordinate 641 (Figure 1). The following observation strongly corroborates the presumption that these three derivatives mutationally inactive *su(w^a)* by reading frame shift rather than for some other reason. When the number of bases inserted at coordinate 641 is increased from 11 ($3n-1$) to 15 ($3n$) (Materials and methods) the mutant phenotype of the +11 frameshift is reverted to wild type.

(ii) If the 3.5 kb RNA is the sole functional *su(w^a)* message, frameshift mutations in exons of the 3.5 kb RNA subsequent to exon 1 should inactivate the gene. In contrast, if the 4.4 kb or 5.2 kb RNA is the sole genetically functional message, such frameshifts will not be mutant as they are in coding sequences not part of the same uninterrupted ORF as that containing the three mutant frameshifts in exon 1 (Figure 1). Exon 2 of the 3.5 kb RNA extends from coordinates 1680 to 1853 and a −4 frameshift at coordinate 1808 inactivates *su(w^a)$^+$* function. Exon 4 extends from coordinates 2911 to 3631 and a +10 frameshift at coordinate 3055 inactivates *su(w^a)$^+$* function. Exon 8 extends from coordinates 4572 to 5460 and a +14 frameshift at coordinate 4633 inactivates *su(w^a)$^+$* function.

(iii) All frameshift mutations inactivating *su(w^a)$^+$* function fail to complement both *su(w^a)^1* and *su(w^a)^{EMSD5}* (see Materials and methods). Moreover, the following pairwise combinations of frameshifts (identified as in Figure 1) fail to complement one another (Materials and methods): 593/1808; 619/1808; 641/1808 and 1808/3505. Collectively, these results indicate that all tested mutant frameshifts (including those in exons 1, 2 and 4) inac-

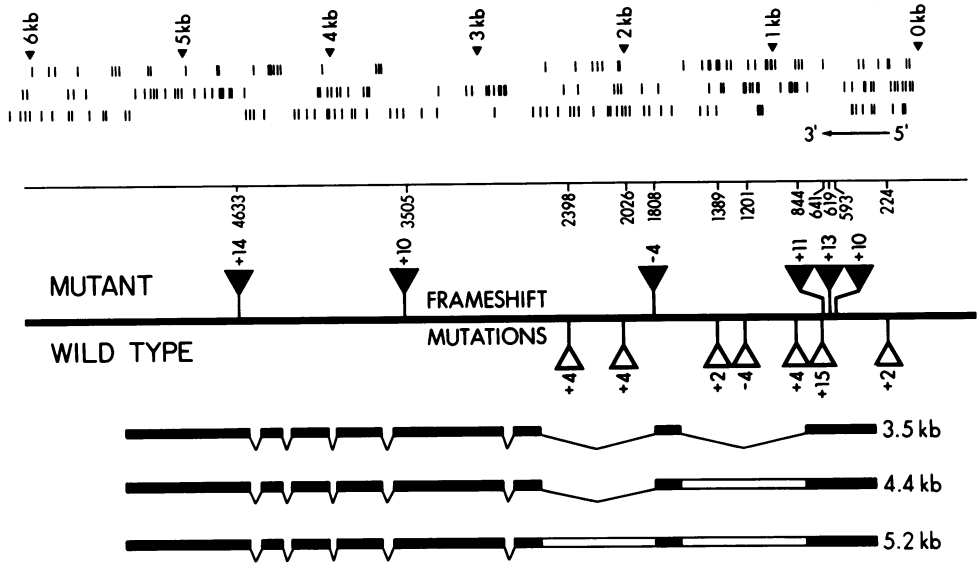

Fig. 1. Reverse genetic analysis of sequences required for $su(w^a)^+$ function. Frameshift mutations were introduced at the coordinates indicated on the thin horizontal line. The triangles indicate the position and eye color phenotype of each frameshift (solid triangles designate a mutant eye color phenotype, open triangles a wild type eye color phenotype; see Materials and method). The number of bases added or deleted in each frameshift (Materials and methods) is indicated by the number associated with each triangle. At the top is the ORF diagram of the $su(w^a)$ sequence and at the bottom are the structures of the three major $su(w^a)$ transcripts (see Figure 6 of accompanying paper for details). The 3.5 kb RNA contains an uninterrupted open reading frame extending from an AUG at coordinate 528 through a UGA at coordinate 5364. The various diagrams are to the same scale and can be projected onto the standard coordinate grid at top.

tivate the same block of peptide coding sequence. The only case in which exons 1, 2 and 4 are in the same continuous segment of peptide coding sequence is the 3.5 kb RNA (Figure 1; Chou *et al.*, 1987).

(iv) The second exon (coordinates 1680–1853) is part of the long ORF extending through the 3.5 kb RNA. In contrast, the two long ORFs in the 4.4 kb and 5.2 kb RNAs extend from the AUG at 528 through 1059 and from the AUG at 2811 through 5363 (Figure 1). Thus, the mutant frameshift at 1808 inactivates a potential coding sequence only in the case of the 3.5 kb RNA.

These various observations demonstrate that the 3.5 kb RNA is necessary for $su(w^a)^+$ genetic function. The question arises as to whether the 4.4 kb or 5.2 kb RNA might also be necessary. The following considerations suggest that neither the 4.4 kb nor the 5.2 kb RNA is required as a messenger RNA to allow $su(w^a)^+$ genetic function.

(i) Translation of the 3.5 kb RNA must begin 5′ to the frameshift at coordinate 593 and presumably at the AUG at coordinate 528 (Figure 1). If the 4.4 kb and/or 5.2 kb RNAs are translated, translation is likely to also begin at coordinate 528. The ORF in the 4.4 kb and 5.2 kb RNAs inactivated by the three exon 1 frameshifts extends from the AUG at coordinate 528 to a UGA at coordinate 1059, whereas exon 1 of the 3.5 kb RNA ends at coordinate 823 (Figure 1). If the 3.5 kb RNA is the sole functional $su(w^a)$ mRNA, frameshift mutations in the portion of this ORF extending beyond exon 1 (interval 823–1059) will not disrupt $su(w^a)^+$ function. In contrast, if the 4.4 kb and/or 5.2 kb RNA are also necessary for $su(w^a)^+$ genetic function, frameshifts in the interval 823–1059 would probably disrupt $su(w^a)^+$ function (Figure 1). A +4 frameshift at coordinate 844 is $su(w^a)^+$ suggesting that the 3.5 kb RNA is the sole genetically functional $su(w^a)$ message. [In addition to being wild type in the presence of the $su(w^a)^1$ and $su(w^a)^{EMSD5}$ endogenous alleles (Materials and methods), the 844 frameshift is wild type in individuals carrying the $su(w^a)^{g165}$ deletion mutant allele at the endogenous $su(w^a)$ locus. $su(w^a)^{g165}$ lacks a segment beginning between coordinates 2100 and 2600 and extending rightward (5′) off the end of the sequenced $su(w^a)$ interval and thus cannot provide either wild type 3.5 kb RNA translation product or a wild type translation product of the 528 through 1059 ORF (our unpublished results).]

(ii) Additional frameshifts in sequences unique to the 4.4 kb and 5.2 kb RNAs do not inactivate $su(w^a)^+$ function. Frameshifts at coordinates 1201, 1389, 2026 and 2398 fail to disrupt $su(w^a)^+$ function (Figure 1).

(iii) While it is difficult to exclude the possibility that the extended ORF in the 3′ portion of the 4.4 kb or 5.2 kb RNA (presumptive initiator AUG at coordinate 2811; Figure 1) is necessary for $su(w^a)^+$ function, this possibility seems remote. Specifically, translation of the 3′ ORF of the 5.2 kb RNA requires initiation at the twenty-fifth AUG and, for the 4.4 kb RNA, at the twentieth AUG in the RNA (Chou *et al.*, 1987). Moreover, it is not apparent that addition of the first and second introns to the RNA would prevent use of the AUG at 528 where translation of the 3.5 kb RNA is apparently initiated.

Timing of effective $su(w^a)$ expression

The genetically functional 3.5 kb $su(w^a)$ RNA is the major precellular blastoderm transcript and it is rapidly replaced (within 6–8 h) by the 4.4 kb and 5.2 kb RNAs; however, the 3.5 kb RNA persists as a trace species through postcellular blastoderm development (Chou *et al.*, 1987). The question arises as to whether $su(w^a)^+$ genetic function is attributable to translation of the precellular blastoderm 3.5 kb RNA or to translation of the trace amounts of the 3.5 kb RNA present throughout postcellular blastoderm development.

The following experiment demonstrates that most or all of $su(w^a)$ expression necessary for a $su(w^a)^+$ adult eye color phenotype occurs after early second larval instar and, thus, must involve the trace amounts of postcellular blastoderm 3.5 kb RNA. Organisms were irradiated between two and three days of age (25°C) with 1000 R from a Cesium-137 gamma ray source (dose

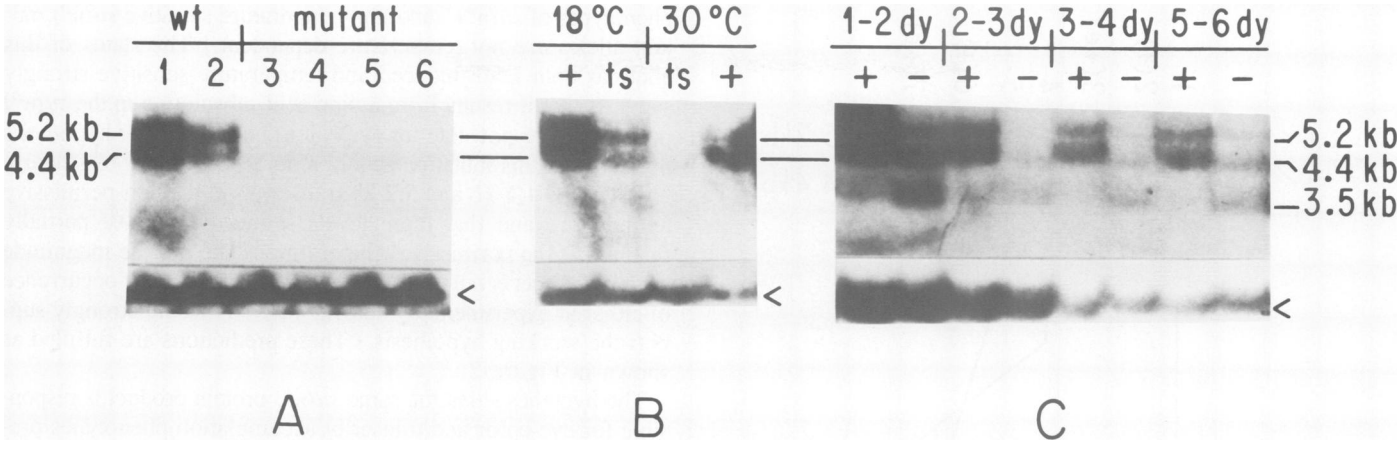

Fig. 2. Northern analysis of effects of mutational inactivation of *su(w^a)* on postcellular blastoderm *su(w^a)* transcript pattern. The genotypes of individuals donating RNAs is indicated above each channel. In **panel A, channels 1** and **2** are two different wild type strains, **channel 3** is *su(w^a)^{EMSD5}*, **channel 4** is *su(w^a)^{EMSA12}*, **channel 5** is *su(w^a)^{EMSDM17}* and **channel 6** is *su(w^a)^1*. **Channels 3–5** are EMS-induced point mutations and **channel 6** is an X-ray induced point mutation. The *ts* allele in **panel B** is *su(w^a)^{SD10}* and the mutant ('−') allele in **panel C** is *su(w^a)^{EMSD5}*. Pupal RNAs were used in **panels A** and **B**, larval RNAs (age of larvae indicated) in **panel C.** Culture temperatures are indicated below **panel B** and are 25°C in all other cases. The *su(w^a)* probe segment extends from coordinates 2022 through 341. Positions of the three major transcripts are indicated. Results of a matching control experiment (Materials and methods) assessing relative RNA inputs are superposed at bottom of each panel (indicated by caret).

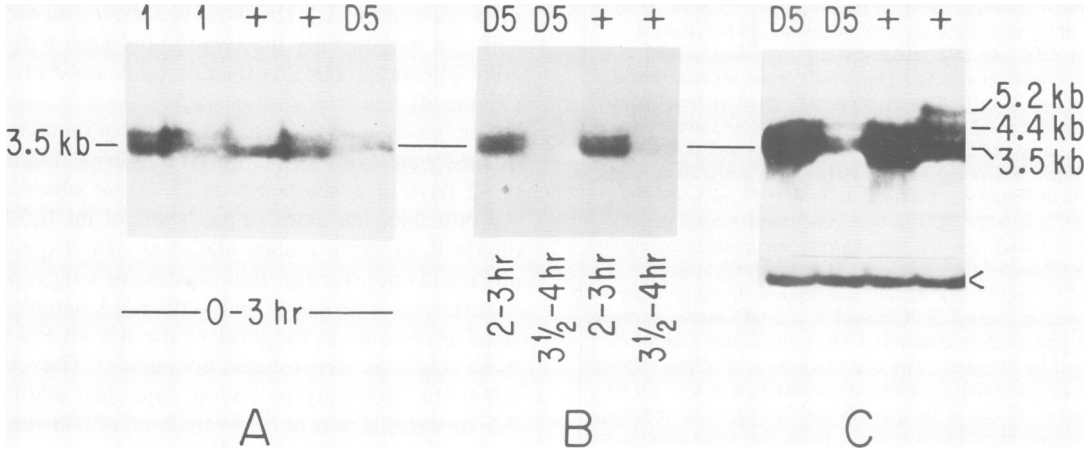

Fig. 3. Northern analysis of the effects of mutational inactivation of *su(w^a)* on *su(w^a)* transcript pattern before and during cellular blastoderm. The genotype of individuals donating RNAs is indicated above each channel as the allele superscript [+ indicates wild type, D5 *su(w^a)^{EMSD5}*, and 1 *su(w^a)^1*]. Age of embryos (collected at 25°C) donating RNAs is indicated below **panels A** and **B. Panel C** is as long exposure of **panel B**. Results of a matching control experiment (Materials and methods) assessing the relative RNA inputs is superposed below **panel C** (indicated by caret). The *su(w^a)* probe segment extends from coordinates 2022 to 341. Positions of the three major transcripts are indicated.

rate 500 R/minute). The irradiated females were of genotype *su(w^a)^{1}w^{a}sn^3/su(w^a)$^+$w^{a}sn$^+$*. Irradiation produces mitotic recombination leading to somatic clones of cells that are homozygous for the *sn^3* mutant allele and/or for the *su(w^a)^1* mutant allele. [Such homozygous *sn^3* mutant clones have been shown previously to produce mutant bristle morphology. *sn^3* thus represents a positive control in this experiment (see Postlethwaite, 1978, and references therein).] If *su(w^a)* genetic function results from translation of precellular blastoderm 3.5 kb RNA, homozygous *su(w^a)^1* somatic clones induced several days after cellular blastoderm should be phenotypically wild type as they were heterozygous for the dominant *su(w^a)$^+$* allele during the relevant developmental period. However, if adult *su(w^a)$^+$* function results from translation of the trace amounts of postcellular blastoderm 3.5 kb RNA, homozygous *su(w^a)^1* somatic clones should be mutant. The latter prediction is fulfilled. Among ~600 females examined, nine independent sectors of *su(w^a)$^-$* eye color phenotype were observed. (All sectors affected one of the two eyes in an individual and ranged in size from ~5% to ~15% of the surface area of the eye. In the same sample of females, four thoracic sectors of *sn* mutant bristles were observed ranging in size from ~10% to ~20% of the dorsal thoracic surface.)

Effects of allelic state of su(w^a) on production of the various mature su(w^a) transcripts

During a conventional genetic anaysis of *su(w^a)*, we collected several EMS-induced mutant alleles (Zachar *et al.*, 1987). None was associated with abberation in the *su(w^a)* region detectable by detailed restriction mapping. Moreover, these mutations arose at relatively high frequencies suggesting a large mutational target size. Collectively, these observations suggested that these mutations result from small lesions (e.g. base change mutations) in a large target such as an extended block of peptide coding sequence.

We were thus surprised to observe that all of these mutations produce drastic reductions in the levels of the 4.4 kb and 5.2 kb

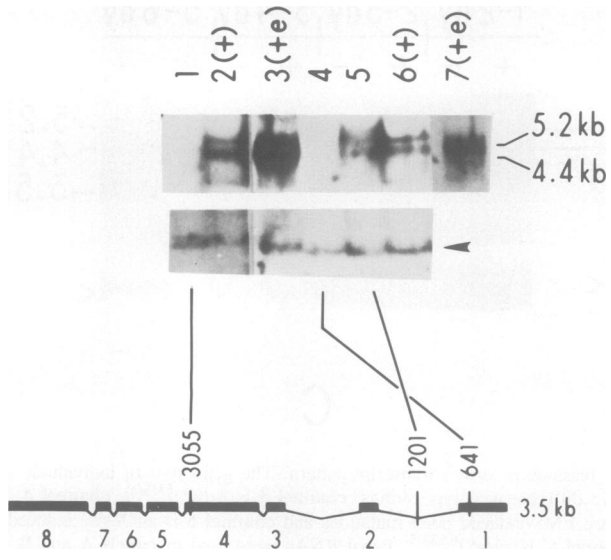

Fig. 4. Northern analysis of the effects of synthetic *su(w^a)* frameshift mutations on *su(w^a)* transcript pattern. RNAs were extracted from adult males carrying a single, autosomal copy of a wild type or frameshift *su(w^a)* allele and the *su(w^a)^{g165}* deletion allele on the X. *su(w^a)^{g165}* is an X-ray induced deficiency allele deleting a segment beginning on the interval between coordinates 2100 and 2600 and extending rightward beyond coordinate 1 (our unpublished results). The *su(w^a)* probe segment extends from coordinate 2022 to 341. Thus, only transcripts from the reintroduced allele are detected (positions of the 4.4 kb and 5.2 kb RNAs are indicated). The diagram shows the structure of the 3.5 kb transcript and positions of tested frameshift alleles. RNAs are as follows (*fs* allele designation refers to synthetic frameshifts in Figure 1). **Channel 1,** *su(w^a)^{g165}; su(w^a)^{fs3505}*. **Channel 2,** *su(w^a)^{g165}; su(w^a)$^+$* (a wild type gene transfer construction). **Channel 3,** *su(w^a)$^+$* (a wild type fly carrying only the endogenous allele). **Channel 4,** *su(w^a)^{g165}; su(w^a)^{fs641}*. **Channel 5,** *su(w^a)g165; su(w^a)^{fs1201}*. **Channel 6,** *su(w^a)^{g165}; su(w^a)$^+$* (a second, independent insertion of the wild type construction used in channel 2). **Channel 7,** a 2-fold shorter exposure of channel 3. Superposed below **channels 1–6** is a matching control for RNA input (indicated by the caret); this control consists of a long exposure of a portion of the experimental gel allowing visualization of a ~1.0 kb transcript originating elsewhere in the genome and having weak homology to this *su(w^a)* probe (our unpublished results). [Note that these experiments suggest that transplantation to an autosome of the 8.2 kb *su(w^a)* interval used removes the gene from dosage compensation resulting in ~2-fold lower levels of expression from wild-type autosomal constructions than from the endogenous wild type (X-linked) allele. We are currently investigating this in detail.]

RNAs beginning at the time of first production of these two forms (~3.5 h) and extending throughout the rest of postcellular blastoderm development (Figures 2 and 3 and results not shown). In addition to the five mutations in Figures 2 and 3, we have examined two other EMS-induced mutations with similar results, bringing to seven the number of independent point mutations analyzed producing drastic reductions in postcellular blastoderm levels of the 4.4 kb and 5.2 kb RNAs (results not shown).

These various observations suggest that the large effects on 4.4 kb and 5.2 kb transcript levels result from mutational inactivation of a protein product of *su(w^a)*. Moreover, the correlation between eye color phenotype and postcellular blastoderm 4.4 kb and 5.2 kb RNA levels suggests that the same protein product is responsible for both phenotypes.

We initially tested this hypothesis by isolating and characterizing an EMS-induced temperature-sensitive *su(w^a)* mutant allele, *su(w^a)^{SD10}*. This allele produces a *su(w^a)$^+$* eye color phenotype at 18°C, a *su(w^a)$^-$* eye color phenotype at 29–30°C and intermediate eye color phenotypes at intermediate temperatures (increasingly mutant with increasing temperature). [In contrast, phenotypes of *su(w^a)$^+$* and non-temperature-sensitive *su(w^a)* mutant alleles are not temperature-dependent.] The status of this allele as both EMS-induced and temperature-sensitive strongly suggests that it results from amino acid substitution in the *su(w^a)* polypeptide responsible for eye color phenotype. Our hypothesis predicts that this mutation should reduce postcellular blastoderm levels of the 4.4 kb and 5.2 kb transcripts at the non-permissive temperature and that these levels should be at least partially restored at the permissive temperature. (The precise magnitude of these effects is not predictable, a *priori*; however, occurrence of either at experimentally detectable levels would strongly support the working hypothesis.) These predictions are fulfilled as shown in Figure 2.

The hypothesis that the same *su(w^a)* protein product is responsible for eye color and transcript accumulation phenotypes predicts that the synthetic frameshift mutations described above (Figure 1) should have effects on postcellular blastoderm *su(w^a)* transcript pattern similar to those of EMS-induced mutations. As predicted, each of three tested mutant frameshift alleles (producing mutant eye color phenotypes) substantially reduce adult levels of the 4.4 kb and 5.2 kb RNAs (Figures 4 and 5). Note especially that the 1808 frameshift reduces 4.4 kb and 5.2 kb RNA levels even though it inactivates potential peptide coding sequences only in the case of the 3.5 kb RNA (see above). In contrast, a frameshift not mutationally inactivating the gene with respect to eye color phenotype has no effect on adult *su(w^a)* transcript pattern (Figure 4).

The question arises as to the effects of mutational inactivation of *su(w^a)* on levels of the 3.5 kb transcript. During larval stages (1–5 days of development at 25°C) we observed modest, but reproducible, increases in the levels of the 3.5 kb RNA in individuals carrying mutant alleles at *su(w^a)* (panel C, Figure 2 and results not shown). Examination of long exposures of some experiments with pupal and adult RNA samples indicates that small elevations in the level of the 3.5 kb RNA occur during these stages as well (results not shown). (Levels of the 3.5 kb RNA are frequently so low at pupal and adult stages that the 3.5 kb signal is near or below the level of non-specific filter labeling preventing definition of effects.)

In contrast to postcellular blastoderm stages, levels of the 3.5 kb RNA during precellular blastoderm stages are not affected by the allelic state of *su(w^a)* (Figure 3).

Direct test of capacity of the su(w^a) protein to regulate postcellular blastoderm transcript pattern in trans

The hypothesis that the *su(w^a)* protein is necessary for efficient postcellular blastoderm accumulation of the 4.4 kb and 5.2 kb RNAs predicts that a wild type *su(w^a)* allele should control the transcript pattern of a mutationally inactivated copy of *su(w^a)* in *trans*. This prediction is demonstrated to be fulfilled by the following experiment. Strains were constructed homozygous for either of two independent, autosomal insertions (generated by gene transfer) of the −4 frameshift at coordinate 1808 [Figure 1; referred to as *su(w^a)^{fs1808}*]. These strains carry either a wild type allele or the EMSD5 point mutant allele at the endogenous, X-linked *su(w^a)* locus. The presence of the frameshift in *su(w^a)^{fs1808}* has two effects. First, it inactivates the allele with respect to accumulation of the 4.4 kb and 5.2 kb RNAs as demonstrated by Northern analysis (Figure 5 and results not shown). Thus, if the pupal transcript pattern produced by *su(w^a)^{fs1808}* is to be wild type, the *su(w^a)$^+$* protein must be provided in *trans*. Second, this frameshift allows transcripts originating from *su(w^a)^{fs1808}* to be distinguished from those of the endogenous copy of *su(w^a)*. This discrimination is achieved

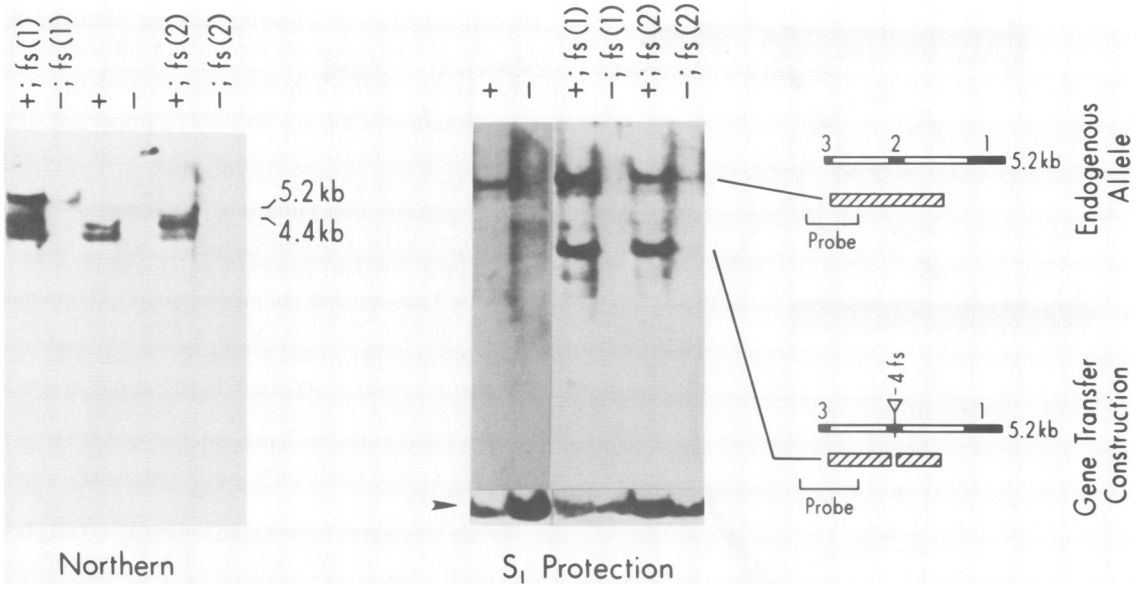

Fig. 5. Control in *trans* of transcript pattern from a frameshift mutant $su(w^a)$ allele by a wild type $su(w^a)$ allele. Northern and S_1 protection experiments using pupal RNAs are shown. The allelic state of the endogenous $su(w^a)$ locus is indicated as follows: '+' is a wild type allele and '−' is the $su(w^a)^{EMSD5}$ mutant allele. 'fs(1)' is one autosomal insertion of the $su(w^a)^{fs1808}$ allele (a mutant allele resulting from a −4 frameshift at coordinate 1808; Figure 1) and 'fs(2)' is a second independent, autosomal $su(w^a)^{fs1808}$ insertion.

Production of the 5.2 kb RNA by $su(w^a)^{fs1808}$ is assayed by S_1 protection with a protector extending from coordinates 2627 to 1198. The Southern transfer of S_1 protection products was probed with a sequence extending from coordinates 2395 to 2826 as diagrammed. The protection products expected are indicated by hatched boxes. [Note that only one of the two products form the 5.2 kb $su(w^a)^{fs1808}$ transcript is homologous to the probe segment.] $su(w^a)^+$ produces a 1430 base protection product and $su(w^a)^{fs1808}$ produces a 820 base protection product as a result of S_1 cleavage of the DNA strand in the RNA:DNA heteroduplex at the point of the frameshift. A matching control experiment (Materials and methods) is superimposed at the bottom of the S_1 protection panel and shows that all samples are well matched except for the $su(w^a)^{EMSD5}$ control channel (second from left) which is overloaded by about 2-fold. Pupal levels of the 5.2 kb RNA from mutant strains show slight sample-to-sample variation when assayed in this way (compare '−' channels) but are always substantially lower than the corresponding wild-type case.

fs(1) produces a new, large transcript class in addition to the 4.4 kb and 5.2 kb classes (Northern panel). Such anomalous transcript classes are seen with some integrants of this and related gene transfer constructions. We believe these result from transcription promoted upstream from and extending into the inserted construction (our unpublished results). Note that this putative fusion transcript is also apparently partially responsive to *trans* regulation by the wild type $su(w^a)$ allele.

by assaying transcript levels by S_1 protection using a DNA segment from the wild type allele (Figure 5). In this circumstance, transcripts from each of the alleles generates a differently sized protection product due to S_1 cleavage of the DNA strand at the point of the −4 frameshift in the heteroduplex with the transcript of $su(w^a)^{fs1808}$ but not in the heteroduplex with the transcript of the endogenous allele. For reasons of technical convenience we assay the transcript pattern produced by the various alleles in question by assaying production of the 5.2 kb RNA (Figure 5).

When the endogenous allele is the EMSD5 mutant allele, both the endogenous and $su(w^a)^{fs1808}$ alleles produce the low levels of the 5.2 kb RNA expected in the absence of functional $su(w^a)$ protein (Figure 5). In contrast, when the endogenous allele is $su(w^a)^+$, both the endogenous and frameshift mutant alleles produce the high levels of the 5.2 kb RNA predicted in the presence of the wild type $su(w^a)$ protein (Figure 5).

Discussion

In this and the accompanying paper we have characterized the developmental regulation of $su(w^a)$ expression by control of splicing. The observations to be accounted for are as follows. (i) $su(w^a)$ produces three major mature transcripts. The smallest of these (the 3.5 kb RNA) is the majority species during precellular blastoderm development after which its levels drop rapidly. The 3.5 kb RNA persists as a minority species throughout the rest of the life of the organism. The larger two mature transcipts (the 4.4 kb and 5.2kb RNAs) first appear around cellular blastoderm and their levels increase substantially during the next few hours. The 4.4 kb and 5.2 kb RNAs are the preponderant RNA species throughout the remainder of the life of the organism. (ii) The 3.5 kb mature RNA is produced by the removal of seven introns. The 4.4 kb RNA results when the first of these is not removed and the 5.2 kb RNA when the first and second are not removed. Thus, the removal of the first intron is developmentally controlled. [The second intron may or may not be controlled independently of the first (Chou *et al.*, 1987).] (iii) The 3.5 kb RNA is clearly a message essential for $su(w^a)^+$ genetic function. In contrast, we find no evidence for mRNA function attributable to the 4.4 kb and 5.2 kb RNAs in spite of their status as the preponderant postcellular blastoderm RNAs. Our somatic recombination experiments indicate that the trace amount of the 3.5 kb RNA present during postcellular blastoderm development is responsible for postembryonic $su(w^a)^+$ function. (iv) The postcellular blastoderm transition to production of the 4.4 kb and 5.2 kb RNAs requires the protein resulting from translation of the 3.5 kb RNA. This requirement for the $su(w^a)$ protein can be detected from the first two hours of the transition through the remainder of the life of the organism. When the protein product of the 3.5 kb RNA is not present (mutationally inactivated) in postcellular blastoderm individuals—resulting in failure to produce high levels of the 4.4 kb and 5.2 kb RNAs—the levels of the 3.5 kb RNA are somewhat elevated. Moreover, during precellular blastoderm development—when the 3.5 kb RNA is the exclusive mature transcript and no 4.4 kb or 5.2 kb mature RNAs are produced—the $su(w^a)$ transcript pat-

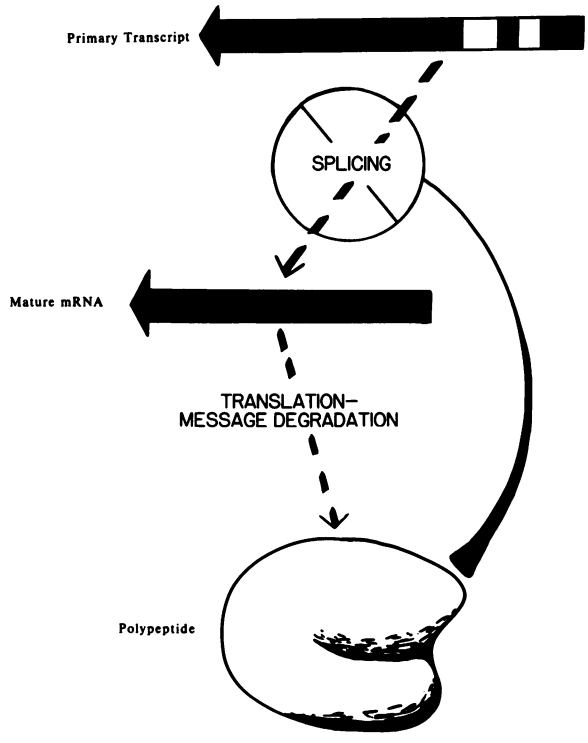

Fig. 6. Model for autogenous regulation of *su(w^a)*. The *su(w^a)* primary transcript is processed to produce a mature, functional messenger RNA (the 3.5 kb RNA; solid arrow at center) in the absence of effective levels of the *su(w^a)* protein. This functional message is translated to produce the *su(w^a)* protein and is also rapidly degraded. The *su(w^a)* protein represses further production of the *su(w^a)* mRNA by blocking the removal of one or both of the first two introns from the continuously produced *su(w^a)* primary transcript (regulated introns symbolized by open portions of the primary transcript diagram) leading to the accumulation of the incompletely processed 4.4 kb and 5.2 kb RNAs.

tern is indifferent to mutational inactivation of the *su(w^a)* protein.

The autoregulation hypothesis

A variety of models are capable of accounting for isolated aspects of these results. However, one model seems uniquely capable of economically unifying all details of our results. This model consists of the following elements and is diagrammed in Figure 6. (i) The *su(w^a)* promoter fires either continuously or sporadically independently of the *su(w^a)* protein. (ii) In the absence of effective levels of the *su(w^a)* protein, the resulting primary transcript is processed by the removal of seven introns to produce the functional *su(w^a)* messenger RNA, the 3.5 kb RNA. [This condition exists when *su(w^a)* is mutationally inactivated and, by hypothesis, during precellular blastoderm development.] (iii) As the *su(w^a)* protein (the translation product of the 3.5 kb RNA) accumulates, it blocks removal of the first or the first and second introns from the *su(w^a)* primary transcript leading to the production of the 4.4 kb and 5.2 kb RNAs at the expense of the 3.5 kb RNA. [By hypothesis, the *su(w^a)* protein begins accumulating during or before precellular blastoderm and reaches effective levels around cellular blastoderm. These levels are presumed to be maintained by continued titration throughout the remainder of the life of the organism.] This constitutes autorepression at the level of splicing. (iv) The 3.5 kb RNA is quite unstable and is more unstable than the 4.4 kb and 5.2 kb RNAs. (The results in panels B and C of Figure 3 demonstrate a half-life for the 3.5 kb RNA of substantially less than an hour during the period immediately following cellular blastoderm. Presumably this low stability persists throughout subsequent development.)

Implications of the autoregulation hypothesis

The instability of the 3.5 kb RNA is important. If the functional *su(w^a)* messenger RNA were stable, the autoregulatory feedback circuit we propose (Figure 6) would allow large oscillations in *su(w^a)* protein concentration. In contrast, if the functional messenger RNA is shortlived, repression of a splice necessary to produce it would result in rapid cessation of production of the *su(w^a)* protein and the autoregulatory loop would be capable of very precise, delicate control of protein levels. Further, this relative instability of the 3.5 kb RNA accounts for the observation that removal (mutational inactivation) of the *su(w^a)* protein from postcellular blastoderm individuals produces a larger decrease in the steady state levels of the 4.4 kb and 5.2 kb RNA than the concomitant increase in the steady state levels of the 3.5 kb RNA.

While it remains to be unambiguously established whether the 4.4 kb and 5.2 kb RNAs have mRNA functions we are currently unable to recognize, our results suggest that these two larger RNAs are merely the byproducts of post-transcriptional repression of production of the 3.5 kb functional mRNA. In spite of this, these two RNAs are the preponderant species throughout most of the life of the organism. Our model economically accounts for this superficially paradoxical result; the *relatively* high steady state levels of the these two RNAs is an indirect result of the requirement for unusually low stability on the part of the functional messenger (3.5 kb) RNA. For the reasons discussed above, this may be a common or universal property of RNAs that are byproducts of repression at the level of splicing. Preponderant transcripts have been identified in the cases of several genes whose structures are strikingly similar to the unusual structures of the 4.4 kb and 5.2 kb *su(w^a)* RNAs (see, for example, Hall and Spierer, 1986; Theissen *et al.*, 1986). We speculate that some or all of these RNAs will prove to be the byproducts of repression at the level of splicing and that the authentic messages in such cases will prove to be shortlived species present at relatively low steady state levels and as yet unidentified.

It is noteworthy that this type of regulatory circuit does not require the regulated splices to be in peptide coding sequences; addition or deletion of sequences from an RNA segment 5′ to peptide coding sequences is demonstrably capable of influencing translation of a message (Hinnebush, 1984).

Previous studies indicate that *su(w^a)* acts in *trans* to regulate or influence an RNA processing event involving transcripts of the *w^a* allele (Zachar *et al.*, 1985; reviewed in Chou *et al.*, 1987). Our results indicating that *su(w^a)* regulates its own expression by controlling an RNA processing event are thus resonant with these earlier observations. We note especially that one of several allowable interpretations of these earlier results is that the *su(w^a)* protein represses splicing of the second *w^a* intron.

Lastly, we note that autogeneous regulation of *su(w^a)* splicing should be unusually accessible to detailed analysis of the interaction of a primary transcript with a protein regulating its splicing. This is in contrast to all other currently available regulated metazoan splices in which no simple route exists to the clear identification of the *trans*-acting factors presumably involved. We are currently pursuing this line of investigation.

Materials and methods

Fly strains

Fly strains not described in the text are described in Zachar *et al.* (1985) and Zachar *et al.* (1987).

Northern and Southern analysis

Formaldehyde-agarose gels were used for Northern and Southern analyses (Maniatis *et al.*, 1982). Nitrocellulose transfers were probed with single-stranded M13 probes made according to Hu and Messing (1982) and used according to Bingham and Zachar (1985).

Molecular weight standards for RNA measurements were the following *Drosophila* transcripts (visualized by reprobing Northern filters with the appropriate sequences): *gypsy* (Marlor *et al.*, 1986), *copia* (Emori *et al.*, 1985), *white* (Pirrotta and Brockl, 1984; Levis *et al.*, 1984; Davison *et al.*, 1985), *Adh* (Benyajati *et al.*, 1983) and *rp49* (O'Connell and Rosbash, 1984). We estimate that our size measurements for RNAs are accurate to within less than about 10%.

rp49 is a ribosomal protein gene whose message is used as a control for the amount of RNA in Northern gel channels. This control measurement consisted of running a small aliquot of each experimental sample on a second gel and probing a transfer with the *Eco*RI − *Hin*dIII fragment containing most of the *rp49* transcription unit (O'Connell and Rosbash, 1984).

Molecular weight standards for Southern analysis were a commerically available (BRL) 123 base ladder. These fragments were end-labelled by conventional Klenow filling which adds two bases to each fragment class. Thus, the size of the *n*-mer in this ladder is (*n*) (123)+2. Size estimates using this standard ladder are reliable within less than about 15 bases in the mol. wt. range of 125−900 bases.

Gene transfer

All gene transfer constructions carried the 6191 base *Sal*I−*Nru*I *su(w^a)* interval (see Chou *et al.*, 1987 for sequence) contained within an 8.2 kb *Sal*I−*Xho*I fragment beginning at the *Sal*I site at coordinate 1 and extending to the *Xho*I site ~2 kb beyond the *Nru*I site at the 3′ end of the sequenced interval. [See Zachar *et al.*, 1987 for a restriction map of the *su(w^a)* region.] In all cases this segment was inserted into the Carnegie 20 P gene transfer vector (Rubin and Spradling, 1983) such that the *Sal*I site was oriented toward the *Eco*RI site in the polylinker and the *Xho*I site was toward the polylinker *Hin*dIII site.

Frameshift mutations diagrammed in Figure 1 were made as follows (frameshifts are designated here and in Figure 1 according to the first base of the restriction cleavage site used to make them): 224 is a +2 frameshift made by Klenow polishing of the 224 *Cla*I site; 593 is a +10 frameshift made by inserting a 10 base *Bgl*II linker (linker sequence GAAGATCTTC) into the 593 *Hae*III site; 619 is a +13 frameshift made by inserting this same 10 base linker into the 619 *Hin*fI site after the ends of the cleaved site had been rendered flush by Klenow polishing. Two variants exist at coordinate 641. The first is a +11 frameshift made by inserting the 10 base *Bgl*II linker into the Klenow polished 641 *Nci*I site. The second was made by opening this inserted linker with *Bgl*II, Klenow polishing and religating; this duplicates the central four bases of the linker (GATC) and increases the total number of bases added to the wild type *su(w^a)* sequence at this site to 15. 844 is a +4 frameshift made by Klenow polishing of the 844 *Ava*I site. 1201 is a −4 frameshift made by Klenow polishing of the 1201 *Pst*I site. 1389 is a +2 frameshift made by Klenow polishing of the 1389 *Acc*I site. 1808 is a −4 frameshift made by Klenow polishing of the 1808 *Kpn*I site. 2026 is a +4 frameshift made by Klenow polishing of the 2026 *Hin*dIII site. 2398 is a +4 frameshift made by Klenow polishing of the 2398 *Bam*HI site. 3505 is a + 10 frameshift made by inserting the 10 base *Bgl*II linker in the 3505 *Hin*cII site. 4633 is a +14 frameshift made by inserting the 10 base *Bgl*II linker into the Klenow polished 4633 *Xma*III site.

P mediated germline gene transfer was carried out using conventional procedures (Rubin and Spradling, 1983; Karess and Rubin, 1984; Zachar *et al.*, 1987).

Analysis of su(w^a) genetic function of gene transfer constructions

The capacity of a construction to confer *su(w^a)$^+$* function was assessed by crossing *su(w^a)$^+$w$^+$: cn$^-$:ry$^-$* males homozygous or heterozygous for an autosomal insertion of the construction to *su(w^a)$^-$w^a: cn$^+$:ry$^+$* females [*su(w^a)$^-$* indicates that both the *su(w^a)^1* and *su(w^a)^{EMSD5}* mutant alleles were tested]. The eye color phenotypes of male progeny of such crosses define whether the construction is *su(w^a)$^+$* or *su(w^a)$^-$* as follows. (i) Crosses in which the construction is *su(w^a)$^+$* and the parental male is heterozygous for the inserted construction produce males with a '*white-apricot*' eye color and males with the '*suppressed white-apricot*' eye color in a 1:1 ratio. (ii) *su(w^a)$^+$* constructions present in homozygous condition produce only '*white-apricot*' eye color in progeny males. (iii) All crosses in which the construction is mutant produce only '*suppressed white-apricot*' males. For all *su(w^a)$^+$* constructions, at least three independent transformants were analyzed and, for all *su(w^a)$^-$* constructions, at least five independent transformants were analyzed.

The capacity of two different frameshifts to complement was assayed by crossing a *su(w^a)^{EMSD5}w^a* female homozygous for an autosomal insertion of one of the two frameshifts to a male carrying an autosomal insertion of the second frameshift heterozygous with a balancer. Eye color phenotypes of male progeny of this cross carrying the balancer (not carrying one of the two frameshifts) and those not carrying the balancer (heterozygous for both frameshift constructions) were compared.

S_1 protection

S_1 protection was carried out by the modified method described in Chou *et al.* (1987). In addition, an aliquot (1/20th) of each sample was withdrawn immediately before S_1 digestion, run on a parallel gel and probed with the *Hin*dIII − *Eco*RI fragment from the *rp49* ribosomal protein gene (O'Connell and Rosbash, 1984) to assess the relative amounts of RNA in each otherwise identical S_1 protection reaction. The result of this matching control experiment is superimposed at the bottom of the S_1 protection panel in Figure 5.

Acknowledgements

We are grateful to Carolyn Chapman, Inka Mims and Michael Belanich for criticism of early drafts of this manuscript. This work was supported by NIH grant GM32003 to P.M.B.

References

Benyajati,C., Spoerel,N., Haymerle,H. and Ashburner,M. (1983) *Cell*, **33**, 125−133.

Bingham,P.M. and Zachar,Z. (1985) *Cell*, **40**, 819−825.

Chou,T.-B., Zachar,Z. and Bingham,P.M. (1987) *EMBO J.*, **6**, 4095−4104.

Davison,D., Chapman,C., Wedeen,C. and Bingham,P.M. (1985) *Genetics*, **110**, 479−494.

Emori,Y., Shiba,T., Kanaya,S., Inouye,S., Yuki,S. and Saigo,D. (1985) *Nature*, **315**, 773−776.

Hall,L.M.C. and Spierer,P. (1986) *EMBO J.*, **5**, 2949−2954.

Hinnebush,A.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6442−6446.

Hu,N. and Messing,J. (1982) *Gene*, **17**, 271−277.

Karess,R.E. and Rubin,G.M. (1984) *Cell*, **38**, 135−146.

Levis,R., O'Hare,K. and Rubin,G.M. (1984) *Cell*, **38**, 471−480.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 202−203.

Marlor,R.L., Parkhurst,S.M. and Corces,V.G. (1986) *Mol. Cell Biol.*, **6**, 1129−1134.

O'Connell,P. and Rosbash,M. (1984) *Nucleic Acids Res.*, **12**, 5495−5513.

Pirrotta,V. and Brockl,Ch. (1984) *EMBO,J.*, **3**, 563−568.

Postlethwaite,J.H. (1978) In Ashburner,M. and Wright,T.R.F., (eds), *The Genetics and Biology of Drosophila*. Academic Press, New York, pp. 359−442.

Rubin,G.M. and Spradling,A.C. (1983) *Nucleic Acids Res.*, **11**, 6341−6351.

Theissen,H., Etzerodt,M., Reuter,R., Schneider,C., Lottspeich,F., Argos,P., Lührmann,R. and Philipson,L. (1986) *EMBO J.*, **5**, 3209−3217.

Zachar,Z., Davison,D., Garza,D. and Bingham,P.M. (1985) *Genetics*, **111**, 495−515.

Zachar,Z., Garza,D., Chou,T.-B., Goland,J. and Bingham,P.M. (1987) *Mol. Cell Biol.*, **7**, 2498−2505.

Received on September 1, 1987; revised on October 12, 1987