The first su(f) allele was isolated on the basis of its ability to suppress the phenotype of the forked-one (f') mutation (Whittinghill 1937) in which a gypsy element inserted into an intron of forked results in gnarled bristles (Parkhurst and Corces 1985). In f' mutants, forked mRNAs are truncated within the 5' long terminal repeat (LTR) of gypsy (Hoover et al. 1993). However, viable mutations at su(f) result in increased amounts of apparently wild-type RNAs made from f'. The simplest explanation of this is that reduced activity of su(f) leads to less efficient mRNA 3' end formation within the 5' LTR of gypsy, so that more transcripts are made that contain the complete gypsy element and reach the 3' end of forked. The gypsy element can then be spliced out to give a structurally wild-type forked mRNA. Null alleles of su(f) are lethal; this is consistent with a vital role for the gene.

The su(f) locus is the most proximal complementation group on the X chromosome and lies at the euchromatin/heterochromatin boundary (Schaeft and Leefurve 1973). The gene is flanked by two copies of a 1.5-kb direct repeat that may play a role in insulating expression of the gene from the effects of neighboring heterochromatin (M. Tudor and K. O'Hare, unpublished data). Alternative polyadenylation results in production of three su(f) transcripts (Mitchelson et al. 1993; see Figure 1). The 2.6 and 2.9-kb mRNAs encode the same 84-kD protein that has homology to two proteins involved in mRNA processing: the product of the yeast RNA14 gene and the 77-kD subunit of human cleavage stimulation factor. Three su(f) mRNAs are produced by alternative polyadenylation. The 2.6- and 2.9-kb mRNAs encode the same 84-kD protein while a 1.3-kb RNA, which terminates within the fourth intron, is unusual in having no stop codon. Using P-element-mediated gene replacement we have copied sequences from a transformation construct into the su(f) gene creating a su(f) allele at the normal genomic location that lacks the first five introns. This allele is viable and appears wild type for su(f) function, demonstrating that the 1.3-kb RNA and the sequences contained within the deleted introns are dispensable for su(f) function. Compared with studies on gene replacement at the white locus, chromosomal breaks at su(f) appear to be less efficiently repaired from ectopic sites, perhaps because of the location of su(f) at the euchromatin/heterochromatin boundary on the X chromosome.

ABSTRACT

The su(f) suppressor of forked (su(f)) locus affects the phenotype of mutations caused by transposable element insertions at unlinked loci. It encodes a putative 84-kD protein with homology to two proteins involved in mRNA 3' end processing: the product of the yeast RNA14 gene and the 77-kD subunit of human cleavage stimulation factor. Three su(f) mRNAs are produced by alternative polyadenylation. The 2.6- and 2.9-kb mRNAs encode the same 84-kD protein while a 1.3-kb RNA, which terminates within the fourth intron, is unusual in having no stop codon. Using P-element-mediated gene replacement we have copied sequences from a transformation construct into the su(f) gene creating a su(f) allele at the normal genomic location that lacks the first five introns. This allele is viable and appears wild type for su(f) function, demonstrating that the 1.3-kb RNA and the sequences contained within the deleted introns are dispensable for su(f) function. Compared with studies on gene replacement at the white locus, chromosomal breaks at su(f) appear to be less efficiently repaired from ectopic sites, perhaps because of the location of su(f) at the euchromatin/heterochromatin boundary on the X chromosome.

The first su(f) allele was isolated on the basis of its ability to suppress the phenotype of the forked-one (f') mutation (Whittinghill 1937) in which a gypsy element inserted into an intron of forked results in gnarled bristles (Parkhurst and Corces 1985). In f' mutants, forked mRNAs are truncated within the 5' long terminal repeat (LTR) of gypsy (Hoover et al. 1993). However, viable mutations at su(f) result in increased amounts of apparently wild-type RNAs made from f'. The simplest explanation of this is that reduced activity of su(f) leads to less efficient mRNA 3' end formation within the 5' LTR of gypsy, so that more transcripts are made that contain the complete gypsy element and reach the 3' end of forked. The gypsy element can then be spliced out to give a structurally wild-type forked mRNA. Null alleles of su(f) are lethal; this is consistent with a vital role for the gene.

The su(f) locus is the most proximal complementation group on the X chromosome and lies at the euchromatin/heterochromatin boundary (Schaeft and Leefurve 1973). The gene is flanked by two copies of a 1.5-kb direct repeat that may play a role in insulating expression of the gene from the effects of neighboring heterochromatin (M. Tudor and K. O'Hare, unpublished data). Alternative polyadenylation results in production of three su(f) transcripts (Mitchelson et al. 1993; see Figure 1). The 2.6 and 2.9-kb mRNAs encode the same 84-kD protein that has homology to two proteins involved in mRNA processing: the product of the yeast RNA14 gene (Mitchelson et al. 1993) and the 77-kD subunit of human cleavage stimulation factor (CstF) (Takagaki and Manley 1994). RNA14 encodes a subunit of the Saccharomyces cerevisiae cleavage factor 1 (CF1), which is required for both 3' end cleavage and polyadenylation of mRNA (Minvielle-Sebastia et al. 1994). CstF enhances 3' end cleavage of mRNAs by stabilizing the interaction between the polyadenylation machinery and the mRNA (Takagaki et al. 1990).

The 1.3-kb su(f) transcript terminates within the fourth intron and is unusual in that it has no stop codon (Mitchelson et al. 1993). If translated, this RNA would encode a protein of ~39 kD sharing its amino-terminal 313 amino acids with the 84-kD protein but having a C terminus encoded by the fourth intron and a poly-lysine tail encoded by the poly(A) tail. Several lines of evidence suggested that this RNA might have a role in the function of su(f). First, temperature-sensitive (ts) alleles of su(f) do not have this RNA at the restrictive temperature (M. Simonelig, personal communication; K. Elliott, personal communication). The lack of this RNA could be the cause or simply a symptom of the ts phenotype. Second, this RNA is conserved as a transcript with no stop codon between Drosophila species (C. J. Williams and K. O'Hare, unpublished results). Third, RNAs lacking stop codons are also derived from two yeast genes with roles in RNA processing: RNA14 (Min-
C. J. Williams and K. O'Hare

![Gene Replacement Diagram]

**Figure 1.** Gene replacement. Boxes are exons. Open boxes are untranslated regions. Restriction sites are as follows: E, EcoRI; H, HindIII; S, SalI; B, BglII; X, XhoI. Probes used for DNA blots are shown. Oligonucleotide primers used for PCR are denoted by numbered arrows. The 1.5-kb direct repeats flanking the gene are shown as hatched boxes. Wild-type *su(f)* transcripts are shown above the map. BX64 is a genomic construct able to make all three *su(f)* mRNAs. WG8 is a cDNA/genomic chimera able to make only the 2.6- and 2.9-kb mRNAs and not the 1.3-kb RNA.


A transformation construct, WG8, containing the *su(f)* gene, was made by substitution of a cDNA fragment for the genomic fragment from exons 1 to 6, thus deleting the first five introns of the gene. This construct cannot make the 1.3-kb *su(f)* RNA and, in some insertion sites, rescues null alleles of *su(f)*. This shows that the 1.3-kb RNA is not always required for normal *su(f)* function (SIMONELIG et al. 1996). However, this construct shows a position sensitivity not observed with a genomic construct, BX64, which can make all three *su(f)* mRNAs. The structures of these transformation constructs are depicted in Figure 1. The position sensitivity of WG8 could be due to deletion of cis-acting regulatory sequences in one or more of the missing introns. To eliminate these position effects and verify that the 1.3-kb RNA is indeed dispensable, we have used P-element-mediated gene replacement (GLOOR et al. 1991; reviewed by LANKENAU 1995) to copy the WG8 construct from an ectopic site into *su(f)*. P-element excision appears to generate a double stranded gap, which may then be widened by exonucleases. The broken ends are able to invade homologous sequences and copy them to repair the break. Normally, the homologous chromosome is used as a template for repair, but in its absence an ectopic template can be employed at rates of ≤10%. This technique has previously been used to copy insertions (NASSIF et al. 1994) and deletions (W. R. ENGELS, personal communication) into the white locus from ectopic sites.

**MATERIALS AND METHODS**

**Drosophila stocks:** Drosophila stocks were reared on cornmeal/treacle/yeast medium. Crosses were made at 25°. Genetic symbols not otherwise explained are in LINDSLEY and ZIMM (1992). *su(f)*<sup>AW97</sup> and *su(f)*<sup>HO2792</sup> are lethal P-element insertions as described by MITCHELSON et al. (1993). Transformation constructs were in the vector pW8, which bears white as a marker (KLEMEMZ et al. 1987). *P[w<sup>+<sup>BX64)* contains a 6.4-kb XbaI-BamHI genomic fragment that rescues *su(f)* null alleles (MITCHELSON et al. 1993). *P[w<sup>WG8]<sup>+<sup> is a cDNA/genomic chimera where the genomic fragment comprising exons 1–6 has been replaced with the corresponding cDNA fragment (SIMONELIG et al. 1996). The 5′ and 3′ flanking sequences are identical in the two constructs (Figure 1). The X-linked insertions of these constructs used as templates for conversion were mapped by recombination and found to lie between yellow and white (*P[w<sup>+<sup>BX64]<sup>]<sub>1</sub>* and close to forked (*P[w<sup>WG8]<sup>]<sub>1</sub>*). Excision of *P* elements was induced using *P[<sup>r<sup>]<sub>T+<sub>](99B)* (ROBERTSON et al. 1988).

In experiments where a homologous *su(f)*<sup>+</sup> chromosome was used as a template for conversion in females, the possibility that viable F<sub>2</sub> males were merely recombinitants was effectively excluded by using a chromosome bearing a lethal muta-
Gene conversion at \( su(f) \)

\[
P: \quad w^{1118} \text{WG8.1 } su(f)^{MS97} \quad \times \quad w^{1118}/Y; \text{Sb e } \Delta 2-3/TM6 \\
\downarrow \\
F1: \quad C(1)DX y f/Y \quad \times \quad w^{1118} \text{WG8.1 } su(f)^{MS97}/Y; \text{Sb e } \Delta 2-3/+ \\
\downarrow \\
F2: \quad w^{1118} \text{WG8.1 } su(f)^*/Y; + \text{ or Sb e } \Delta 2-3/+ \\
\]
mated to \( C(1)DX y f / Y \) then analysed by PCR.

**Figure 2.**—Crossing scheme. WG8.1 is a viable \( X \) chromosome insertion of the \( P[w^*,\text{WG8}] \) construct. \( su(f)^* \) denotes the putative replacement allele.

Molecular biology: Drosophila genomic DNA was prepared from frozen adult flies as described by LEVIS et al. (1982). Subcloning and DNA blotting were performed by standard techniques (SAMBRUG et al. 1989). Preparation of RNA and analysis by RNA blotting using single stranded DNA probes was as described by MITCHELSON et al. (1993). To control for loading, RNA blots were probed with sequences from the actin 5C gene (FISBerg et al. 1983). Polymerase chain reactions (PCR) were carried out on a Biometra or Perkin Elmer thermocycler as recommended by the manufacturer. Oligonucleotides used for PCR were: \( su(f) \) primers: 1, 5'-AGTTCCTGA-AACCTGAAGCTATAAAC-3'; 2, 5'CGATGACACTATCGCAGT-ACGTGACTGTGCGTT; 3'P, 5'-CATATCGATGTCTCACT-G3'; 4, 5'-ATTCGTCCGCACACAACCTT-3'; 5'-CATATCGACTGTCCGT; 3'P, 5'-CATATCGACTGTCCGT-TCAG-3'. The positions of these primers are shown in Figure 1. PCR products were cloned into M13 vectors for sequencing. DNA sequencing was performed by the dideoxy chain termination method (BANNER and BARRELL 1983) using T7 DNA polymerase (Pharmacia). Sequence reactions were fractionated on 6% polyacrylamide gels (19:1 mono/bisacrylamide). Sequences were analyzed using the MacVector package (Kodak IBI) on a Macintosh LCIII computer.

**RESULTS**

Gene conversion strategy: \( P \)-element-mediated gene conversion from ectopic sites occurs at rates of \( \approx 10\% \) at the \( white \) locus and exhibits a strong preference for templates on the same chromosome (ENGELS et al. 1994). With this in mind, a small scale experiment using the \( X \)-linked template, \( P[w^*, \text{WG8}] \) (referred to as WG8.1), was designed. The target for conversion was the allele \( su(f)^{MS97} \) (Figure 1). This lethal allele has an insertion of a 1.15-kb \( P \) element into the second exon of the gene and is rescued by WG8.1. As a result, no phenotypic selection for conversion events at \( su(f) \) could be used. However, as the ultimate aim of the experiment was to determine whether a \( su(f) \) allele lacking intron 4 was viable, selection for viability of convertants would have been prejudgmental. Crosses were set up to induce excision of the MS97 \( P \) element, as depicted in Figure 2, and individual \( F_2 \) males were mated, then screened for excision of the \( P \) element using PCR. Any male that gave no PCR product with primers 2 (3' of the insertion) and PP2 (within the \( P \) element) was analyzed further by PCR and DNA blotting. Out of a total of 370 individual \( F_2 \) males, 92 gave no PCR product with primers 2 and PP2. These were found to be the products of six independent events, which are described below.

Gene conversion event: One male was isolated that gave no mutant specific PCR products. The only PCR products observed were those for WG8.1. DNA blotting (Figure 3) showed that in this allele, \( su(f)^{333} \), the HindIII-SalI interval containing the insertion site of the \( P \) element in \( su(f)^{MS97} \), is the same size as in WG8, while the distal HindIII-SalI interval is the same size as in \( su(f)^* \). Subsequent DNA sequencing of PCR products, after recombination of \( su(f)^{333} \) away from WG8.1 (see below), showed that the introns had been precisely deleted confirming that a gene conversion between WG8.1 and \( su(f)^{MS97} \) had occurred. It is impossible to determine the length of the conversion tract as 5' to intron 1 and 3' to intron 5 the WG8 construct is identical to the genome. At least 203 bp 5' from the \( P \) element and at least 1549 bp 3' from the \( P \) element have been replaced.

Recombination of \( su(f)^{333} \) away from WG8.1 demonstrated that males of the genotype \( y^2 w^o ct^6 le_{16} v^f su(f)^{333}/Y \) are viable and fertile and do not show suppression of the \( forked \) phenotype. RNA blotting (Figure 4) confirmed that these males produce only the 2.6- and 2.9-kb \( su(f) \) mRNAs and no 1.3 kb RNA is observed. The 2.6- and 2.9-kb mRNAs appear to be expressed at a slightly higher level than wild type. Females of the genotype \( w m f su(f)^{333} \) are also viable and fertile and have forked bristles as expected for wild-type expression of \( su(f) \). This indicates that \( su(f)^{333} \) acts as a wild-type \( su(f) \) gene.

Imprecise excision events: Three clusters of males were identified where the \( P \) element had excised imprecisely. In \( su(f)^{138} \), both \( P \) element termini are still present but the bulk of the \( P \) sequences have been deleted. In the alleles \( su(f)^{107} \) and \( su(f)^{27} \), the 3' terminus of
DNA blotting analysis of su(f)^+.

Genomic DNA was digested with restriction enzymes as indicated above the blot. The lanes contain DNA from the following strains: 1, WG8.1 su(f)^+; 2, WG8.1 su(f)^+/Y; 3, su(f)^+/Y. Blot (a) was hybridized with the pXS probe (Figure 1). The HindIII fragment from su(f)^+ is 600 bp smaller than the fragment from wild type. This is consistent with the cumulative size of introns 1-5. The su(f)^+/Y HindIII-SalI fragment is the same size as wild type confirming that there is no disruption to the 5' end of the gene. Blot (b) was hybridized with the pB1.4 probe (Figure 1). The HindIII-SalI su(f)^+ fragment is the same size as that derived from WG8.1.

Deletion of su(f) sequences: The allele su(f)^+ gave no mutant specific band on DNA blots probed with su(f) sequences. PCR with primers 1 (5' to the distal 1.5-kb repeat) and 3 (3' to the proximal 1.5-kb repeat) gave a fragment of 1.5 kb in size. These data indicate that this allele has a deletion of all su(f) coding sequences between the 1.5-kb direct repeats flanking the gene (Figure 5). Sequencing of this fragment revealed the presence of polymorphisms from both the proximal and distal repeats. This structure is essentially identical to that of the allele su(f)^-26, which was recovered after X-ray mutagenesis. Recombination between direct repeats has been observed in yeast after generation of a double stranded break by HO endonuclease (RUDIN and HABER 1988) and these alleles could have been produced by a similar mechanism.

Pelement recruitment: One allele, su(f)^+27, proved to be the result of secondary insertions into the original P element. DNA blots (not shown) indicate that the insertion has increased in size from 1.15 to 3.6 kb. PCR using combinations of P-element and su(f) primers and partial DNA sequencing of PCR products suggests that the insertion has the structure shown in Figure 5, where a second element has inserted at the 3' end of the original and a third element has inserted into the MS97 element. Both of the new insertions are in the opposite orientation to the original element. Secondary insertion events have been reported at the vestigial locus (WILLIAMS et al. 1988) and are thought to occur by a gene conversion-like mechanism.

Conversion from an autosomal WG8 template: The rate of gene conversion using WG8.1 as a template appears to be much lower (1/370 or \( \approx 0.27\% \)) than has been observed using X-linked templates at white (3.6-10%). An experiment using a second chromosome insertion of WG8 and selecting for viable derivatives gave no convertants out of 1842 (<0.05%). Poor utilization of WG8 could be due to the absence of introns, so that

**Figure 3.** DNA blotting analysis of su(f)^+33. The lanes contain approximately 5 µg poly(A)^+ RNA. Lane 1 su(f)^+ females; lane 2 su(f)^+ males; lane 3 su(f)^- males. The probe used was a single-stranded probe complementary to sequences from the 5' end of the gene.

**Figure 4.** RNA blotting analysis of su(f)^+33. The lanes contain approximately 5 µg poly(A)^+ RNA. Lane 1 su(f)^+ females; lane 2 su(f)^+ males; lane 3 su(f)^- males. The probe used was a single-stranded probe complementary to sequences from the 5' end of the gene.
Using a second chromosome insertion of BX64 it was possible to select for viable convertants. The lethal allele \(sf(\text{M}8252)\) was used in this experiment. This is an insertion of a \(P\) element identical to the \(MS97\) element but in the opposite orientation and 65 bp upstream (Mitelson et al. 1993). The crosses shown in Figure 6 were made. In a screen of 16,259 chromosomes, no conversion events were observed. Analysis of 50 individuals by PCR suggested that excision of the \(P\) element occurred at a similar frequency to that observed in the above experiments using \(sf(\text{M}8252)\). These results suggest that gene conversion at \(sf(f)\) is inherently less frequent than at \(white\).

**Conversion from a homologous chromosome:** \(P\) element excision was induced in females where the homologous chromosome could be used as a template for repair. Precise excision events were identified by their viability (Table 1). The rate of reversion at \(sf(f)\) using a wild-type chromosome as template was 1.7% and using a rearranged chromosome, the reversion rate was 0.9%. Conversion occurred at a rate of 1.8–2% in males using an \(XY\) translocation bearing \(sf(f)^+\) (\(B^E\)\(Y^y\) or \(B^Y\)).

**DISCUSSION**

The 1.3-kb RNA is dispensable: The viability of \(sf(f)^{331}\) flies that lack introns 1–5 shows that the 1.3-kb RNA and any protein encoded by it are not necessary for \(sf(f)\) function. When 3' end formation occurs within the fourth intron in \(sf(f)^+\), the 1.3-kb RNA is generated at the expense of the larger mRNAs. As a consequence, levels of the SU(F) 84-kD protein might be reduced. It is possible that the 1.3-kb RNA has a posttranscriptional regulatory function and is produced simply to “fine-tune” expression of the 84-kD protein. The yeast gene \(CBP1\) encodes two transcripts (one lacking a stop codon) that are reciprocally regulated in a similar manner (Mayer and Dieckmann 1989).

As mRNAs that would normally terminate within intron 4 now presumably have their 3' ends generated at the end of the gene (see Figures 1 and 4), the 84-kD protein may be slightly overexpressed in \(sf(f)^{331}\) flies. If so, this slight overexpression of the 84-kD SU(F) protein does not appear to be deleterious. One might expect overexpression of SU(F) protein to result in polyadenylation at inappropriate sites. Reduction of either \(sf(f)\) activity in viable \(sf(f)\) mutants (Parkhurst and Corces 1985) or human CstF activity in adenovirus infection (Mann et al. 1993) does lead to changes in utilization of polyadenylation sites. However, it is not clear if \(sf(f)\) is limiting in production of the Drosophila equivalent of CstF.

The viability of \(sf(f)^{331}\) shows that the failure of WG8 in some genomic locations to rescue null alleles of \(sf(f)\) cannot be due to inadequate expression because of the absence of enhancers in the introns. Some nonrescuing
WG8 insertions have been mobilized to new sites where they do rescue, indicating that they have not undergone some rearrangement that rendered them inactive (St-MONELIG et al. 1996). Although the level of expression of the white marker is not obviously lower in nonrescuing WG8 insertions compared with those that do rescue, the level of expression of su(f) must be insufficient for viability. This may reflect the presence and/or tissue specificity of silencer elements in the genomic DNA flanking the insertion.

**Gene conversion at su(f):** The rate of gene conversion from an ectopic template is much lower at su(f) than has been observed, even with the poorest ectopic templates, at white. Table 1 compares the rate of precise P-element excision in the presence of various templates at w and at su(f). In these experiments, excision of P elements from su(f), as assayed by PCR, occurred at a high frequency, typically 8–10%. The frequency at w<sup>md</sup> may be several-fold higher (JOHNSON-SCHLITZ and ENGELS 1993), but this is not sufficient to explain the difference in reversion frequency in the presence of ectopic templates. Reversion of su(f)<sup>Bs2</sup> in the presence of a homolog in females occurred at a rate of 1.7% compared with 13.6% at w<sup>md</sup>. The combination of less frequent excision with less frequent repair (even from homologues) suggests that if the search for autosomal templates occurred at a similar efficiency for su(f) as it does for w, then the overall expected reversion rate using an ectopic template would perhaps be 25-fold lower. We have only examined two autosomal templates (one WG8 and one BX64) and have recovered no revertants, so we estimate that the reversion frequency in the presence of autosomal templates is <0.005%. A range of frequencies has been observed for w<sup>md</sup>, but only the least effective are comparable (after a 25-fold reduction) with the frequency at su(f). We suspect that the location of su(f) at the euchromatin/heterochromatin border interferes with the search for homologous sequences. It has been proposed that during interphase, euchromatin and heterochromatin are segregated into different nuclear compartments (TALBERT et al. 1994). This segregation might hinder pairing between su(f) and euchromatic templates but would not affect pairing with the homologous chromosome. To formally exclude the possibility that less frequent excision combined with less frequent repair is sufficient to explain our lower frequency of gene conversion, either significantly larger experiments or a much wider range of templates would have to be tested.

The technique of P-element-mediated gene conversion has been used here to answer a specific question. We have been able to show that an unusual RNA lacking a stop codon has no significant role in su(f) function. This demonstrates that the technique can be applied to genes other than white where phenotypic selection and viability of insertion mutations has facilitated model studies.
We thank Bill Engeis for advice and sharing unpublished data, Alan Cheshire for technical assistance, and Kate Elliott and Jyrki Eloranta for comments on the manuscript. C.J.W. was the recipient of an Medical Research Council studentship. This work was funded by grant 040633 from the Wellcome Trust.

LITERATURE CITED


Mitchelson, A., M. Simonelig, C. Williams and K. O'Hare, 1995 Homology with Saccharomyces cerevisiae RNA14 suggests that phe-