Effects of Transposable Elements on the Expression of the *forked* Gene of *Drosophila melanogaster*

Kenley K. Hoover, Andy J. Chien and Victor G. Corces

Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

Manuscript received April 13, 1993
Accepted for publication June 23, 1993

**ABSTRACT**

The products of the *forked* gene are involved in the formation and/or maintenance of a temporary fibrillar structure within the developing bristle rudiment of *Drosophila melanogaster*. Mutations in the *forked* locus alter this structure and result in aberrant development of macrochaetae, microchaetae and trichomes. The locus has been characterized at the molecular level by walking, mutant characterization and transcript analysis. Expression of the six *forked* transcripts is temporally restricted to mid-late pupal development. At this time, RNAs of 6.4, 5.6, 5.4, 2.5, 1.9 and 1.1 kilobases (kb) are detected by Northern analysis. The coding region of these RNAs has been found to be within a 21-kb stretch of genomic DNA. The amino terminus of the proteins encoded by the 5.4- and 5.6-kb *forked* transcripts contain tandem copies of ankyrin-like repeats that may play an important role in the function of *forked*-encoded products. The profile of *forked* RNA expression is altered in seven spontaneous mutations characterized during this study. Three *forked* mutations induced by the insertion of the *gypsy* retrotransposon contain a copy of this element inserted into an intron of the gene. In these mutants, the 5.6-, 5.4- and 2.5-kb *forked* mRNAs are truncated via recognition of the polyadenylation site in the 5' long terminal repeat of the *gypsy* retrotransposon. These results help explain the role of the *forked* gene in fly development and further our understanding of the role of transposable elements in mutagenesis.

**DEVELOPMENT** of macrochaetae and microchaetae in *Drosophila melanogaster* involves choreographed interactions between four sister cells. These cells are derived from two successive mitotic divisions of a sensory mother cell. At approximately 15 hr after puparium formation, the two predecessor cells generated from the first of these divisions and located within the epidermal layer divide to give rise to two sets of cells. One pair will differentiate to form the external sensory neuron and its protective sheath cell, the thecogen. The other set is composed of two precursor cells, the socket-forming tormogen cell and the bristle-secreting trichogen cell. Between 18 and 30 hr of pupal development, the tormogen and trichogen precursor cells grow larger than surrounding epidermal cells in volume and become polyploid through a series of endomitotic divisions (POODRY 1980). Bristle secretion commences at approximately 30 hr after puparium formation and proceeds rapidly during the next 25 hr of development. Socket formation begins after bristle secretion has started. The socket cell has completely circumscribed the bristle by 48 hr after puparium formation.

Growth of the developing bristle occurs predominantly at its tip and proceeds perpendicular to the body axis (LEES and WADDINGTON 1942; LEES and PICKEN 1945). Histological studies indicate that bristle morphogenesis depends upon the structural grouping and orientation of microtubules and microfibrils within these cells. In particular, the formation and arrangement of fiber bundles associated with the cell surface of the developing bristle rudiment plays a critical role in bristle development. These intracytoplasmic fiber bundles are closely applied to the cell surface and are oriented parallel to the long axis of the bristle rudiment. Cross sectional analysis of the bristle rudiments of *Stubble* and *singed* mutants shows an alteration in the amount, shape, and distribution of these fiber bundles (OVERTON 1967). In contrast, these mutants do not show abnormalities in the more centrally located microtubule structures when compared to wild-type bristle sections. Cross-sectional analysis of the bristle rudiments in the *forked* (*f*) mutant *f*^36a^ 40 hr after puparium formation fails to detect these longitudinally arrayed fiber bundles (N. PETERSON, personal communication). The *forked* gene (1-56.7) is located in chromosomal subdivision 15F. The product of this gene is essential during a 24-hr period just prior to bristle secretion (DUDICK, WRIGHT and BROTHERS 1974). Failure of the *forked* product to be made in wild-type quantities results in bristles that are shortened and aberrantly bent or forked with ends that may be sharply twisted or split. The severity of the bristle phenotype ranges from the relatively mild *f*^173^ and *f*^7^ alleles to the severe *f*^bd^, *f*^36a^ and *f*^146^ alleles.

The primary source of spontaneous mutations in
yeast and Drosophila is the insertion of transposable elements. The mechanism of transposable element mutagenesis upon the affected locus depends on the location of the insertions site with respect to the different structural and functional domains of the mutated gene. Possible modes of mutagenesis include interference with enhancer and/or promoter activity, disruption of the coding region, alteration of message processing, and perturbation through more complicated mechanisms inherent to the nature of the element causing the mutation. For example, insertion of the gypsy retrotransposon in the 5' region of the yellow (y) or cut (ct) loci effects the rate of transcription of these genes in a tissue-specific fashion (CORCES and GEYER 1991; JACO et al. 1991), whereas insertion of the copia element in an intron of the white (w) gene results in premature truncation of transcripts initiating in the white promoter (ZACHAR et al. 1985; MOUNT, GREEN and RUBIN 1988; PENG and MOUNT 1990). Characterization of the forked locus has been undertaken in order to analyze the mode of gypsy mutagenesis in alleles where the element has inserted into the intron of the gene. The forked locus was cloned by utilizing the gypsy retrotransposon as a tag to screen a genomic library prepared from the DNA of a f' stock (PARK-HURST and CORCES 1985; MCLACHLAN 1986). The f', f' and f alleles of this gene are gypsy-induced and bring the expression of the forked gene under the control of the su(Hw) gene as well as the suppressor of forked [su(f)], suppressor of sable [su(s)], suppressor of white apricot [su(w)], suppressor of purple [su(pr)], suppressor of lozenge [su(lz)], and enhancer of white eosin [en(w)] genes (RUTLEDGE et al. 1988; SMITH and CORCES 1991). The number of modifier loci that alter forked expression in these gypsy-induced alleles suggests that the gypsy element may alter the expression of the forked gene by multiple mechanisms. The su(Hw) protein has been implicated to act at the tran-
The Drosophila foriked Gene

RESULTS

Analysis of transcripts encoded by the foriked gene: A chromosomal walk of contiguous genomic clones was performed in both directions flanking the insertion of the gypsy retrotransposon in the f allele (Parkhurst and Corces 1985) in order to derive a DNA map of the foriked region. This walk spans a distance of 50 kb (Figure 1). The 6.2-kb EcoRI-SalI genomic fragment within which the gypsy is inserted was used as a probe to hybridize to a Northern blot of poly(A)+ RNA prepared from various stages during Drosophila development. The results (Figure 2) show that the foriked locus is temporally expressed during mid-late pupal development, during which time six transcripts are detected. The amount of these mRNAs was quantitated with a Molecular Dynamics PhosphorImager. A prominent 2.5-kb transcript accounts for 42% of the total foriked RNA. Two RNAs of 1.9 and 1.1 kb make up 17 and 18% of the total foriked RNA, respectively. Transcripts of 5.6 and 5.4 kb in length each constitute 10% of the total foriked RNA. A rare 6.4-kb RNA constitutes 3% of the total foriked RNA.

Characterization of foriked transcripts by North-
ern and cDNA analysis: An oligo(dT)-primed cDNA library was generated from poly(A*) mRNA of mid-late Canton S pupae. Several cDNA clones were isolated using the 6.2-kb EcoRI-SalI genomic forked fragment as a probe. Genome and cDNA isolates were sequenced and the organization of exons and introns was determined by comparing the sequence of the cDNA isolates with that of the corresponding genomic DNA. The splice site donor and acceptor sequences were identified as consensus splice sites (MOUNT 1982).

As depicted in Figure 3, all clones that were obtained from this screen recognized the first of the two closely polyadenylation signals located at base 18,423 of the genomic sequence (PROUDFOOT and BROWNLEE 1976). Sites of polyadenylation were observed to begin 16–25 bases downstream of this polyadenylation signal. Analysis of these cDNA clones allowed the determination of the structure of forked encoded transcripts. The 2.5-kb transcript: One of the clones isolated during the screen, named 22G, defines cDNA A (Figure 1). This cDNA includes six exons. To determine which forked transcript is encoded by cDNA A, exon-specific probes were made by the polymerase chain reaction (PCR), with primers flanking each exon, and used for Northern analysis (data not shown). Exon A2 also hybridizes to the 5.4-, 5.6- and 6.4-kb RNAs. Northern analysis (data not shown). Exon A2 also hybridizes to the 5.4-, 5.6- and 6.4-kb RNAs.
The Drosophila forcked Gene

Exons A3, A4, A5 and A6 hybridize to all forked transcripts with the exception of the 1.1-kb RNA, which only hybridizes to A3 and A6 (see below). These results suggest that cDNA A, defined by clone 22G, corresponds to the 2.5-kb transcript.

Sequence analysis of clone 22G indicates that the 5` end has a guanosine nucleotide that is not present in the genomic sequence. This difference is presumably the result of the 7-methyl guanosine cap of the mRNA (BROWN et al. 1989). The transcription start site of this transcript is AGGCTGC, located at base 12,982 in the genomic sequence (Figure 3) and has weak homology to the ATCAG/TTC/T sequence that is conserved in the transcription initiation site of other Drosophila genes (HULTMARK, KLEMENZ and GEHRING 1986). A sequence resembling a TATA box is flanked by a G-C rich sequence and is located 32 bp upstream of this transcription site. There is a possible CAAT box located 61 bp upstream of this transcription initiation site (Figure 3). The low extent of homology of the transcription start site, TATA box, and CAAT box with that of the consensus for these promoter signals in other eukaryotes may explain the relatively low abundance of this transcript or may be due to a pupal or tissue specific promoter sequence. Similar low homologies have been described for the pupal specific transcript of the singed locus (PATERSON and O`HARE 1991). Three in-frame ATG codons are

Figure 3.—Continued.
and 6.4-kb transcripts. In one instance, the alternative in translation initiation sites best fits the nucleotide bias of acceptor splice sites are acknowledged at two locations in exons common to the 1.9-, 2.5-, 5.4-, 5.6- and 7.0-kb cDNAs. The open reading frame of this mRNA is 1.9 kb upstream of this ATG site. The 2.5-kb RNA contains an in-frame termination codon 2 nucleotides upstream of this ATG site. The 2.5-kb RNA contains an in-frame non-sense codon. Sequence analysis of 75% of the clones. The alternative acceptor splice sites correspond to the fourth exon of the 2.5-kb transcript. These alternative sites are located 24 bp apart at nucleotides 16,892 and 16,916 in the genomic sequence. The downstream site is acknowledged in 75% of the clones. The alternative acceptor splice sites recognized during sequence analysis of the isolated clones. Twelve nucleotides separate these sites. In this case, the upstream site located at nucleotide 17,206, is acknowledged in 75% of the clones. 

Sequence analysis of cDNA clones shows that alternate acceptor splice sites are acknowledged at two locations in exons common to the 1.9-, 2.5-, 5.4-, 5.6- and 6.4-kb transcripts. In one instance, the alternative acceptor-splice sites correspond to the fourth exon of the 2.5-kb transcript. These alternative sites are located 24 bp apart at nucleotides 16,892 and 16,916 in the genomic sequence. The downstream site is acknowledged in 75% of the clones. The alternative acceptor splice sites recognized during sequence analysis of the isolated clones. Twelve nucleotides separate these sites. In this case, the upstream site located at nucleotide 17,206, is acknowledged in 75% of the clones. The alternative acceptor splice sites recognized during sequence analysis also agree with the Drosophila splice site consensus (MOUNT 1982). Alternatives are of lengths such that the reading frame of this mRNA, preceded by an in-frame non-sense codon. Sequence flanking the first of these upstream sites best fits the nucleotide bias in Drosophila melanogaster (CAVENER and RAY 1991). There is an inframe termination codon 21 nucleotides upstream of this ATG site. The 2.5-kb RNA contains six exons and spans 5.5 kb of genomic DNA (Figure 3). The open reading frame of this mRNA is 1.9 kb and encodes a polypeptide of 71 KD.

Sequence analysis of cDNA clones shows that alternate acceptor splice sites are acknowledged at two locations in exons common to the 1.9-, 2.5-, 5.4-, 5.6- and 6.4-kb transcripts. In one instance, the alternative acceptor-splice sites correspond to the fourth exon of the 2.5-kb transcript. These alternative sites are located 24 bp apart at nucleotides 16,892 and 16,916 in the genomic sequence. The downstream site is acknowledged in 75% of the clones. The alternative acceptor splice sites recognized during sequence analysis of the isolated clones. Twelve nucleotides separate these sites. In this case, the upstream site located at nucleotide 17,206, is acknowledged in 75% of the clones. The alternative acceptor splice sites recognized during sequence analysis also agree with the Drosophila splice site consensus (MOUNT 1982). Alternatives are of lengths such that the reading frame
will not be altered depending upon the selection of the site. Acceptor splice site selection does not appear to be specific to each size class forked RNA, as two independent clones of the 2.5-kb transcript display differences in both alternate acceptor-splice site selection and B6, hybridize to the 5.6-kb transcript, suggesting that cDNA B encodes the 5.6-kb transcript which is assembled with clones 86E, 11E, and B6, hybridize to the 5.6-kb transcript, whereas clones 86E and 11E suggest that the exons are common to these transcribed, untranslated leader of the transcript (Figure 3). Probes corresponding to exons B2, B3, B4, B5 on Northern blots whereas clones 86E and 11E hybridize to the 5.4-, 5.6- and 6.4-kb RNAs. To determine which of these two transcripts is encoded by the cDNAs, exon-specific probes were made by PCR of primers flanking each exon and used for Northern analysis. Results from these experiments (data not shown) indicate that exon B1 hybridizes only to the 5.6-kb transcript, suggesting that cDNA B which is assembled with clones 86E, 11E and 15T, encodes the 5.6-kb forked transcript. Analysis of clones 86E and 11E indicates that they extend into the 5′ transcribed, untranslated leader of the transcript (Figure 3). Probes corresponding to exons B2, B3, B4, B5 and B6, hybridize to the 5.4-, 5.6- and 6.4-kb RNAs, suggesting that the exons are common to these tran-
The Drosophila forked Gene

scripts. The 5.6-kb RNA is encoded within 12 kb of genomic DNA from coordinate +6.4 to +18.4 of the chromosomal walk and has eleven exons. A 4.3-kb open reading frame in this mRNA could encode a protein of 157 kD.

The 5.4-kb transcript: A cDNA clone, designated 77X, composes cDNA C and was isolated from the screening of the random hexanucleotide primed library. This clone hybridizes to 6.4-, 5.6- and 5.4-kb RNAs when used as a probe. Four exons are contained within the clone. The most 5' of these exons, named C1, hybridizes only to the 6.4- and 5.4-kb RNA. The relative abundance of these RNAs suggest that cDNA C, which is composed of clones 77X and 15T, corresponds to the 5.4-kb transcript. The three terminal exons of clone 77X are shown by sequence analysis to be the same as exons B2, B3 and B4 in the 5.6-kb RNA. The 5' end of the C1 exon does not contain an open reading frame and represents the transcribed, untranslated leader of the transcript. Sequence flanking the translation initiation site of the 5.4-kb RNA are in agreement with that found in other Drosophila genes (CAVENGER and RAY 1991). This transcript spans 18 kb of genomic DNA from interval +0.4 to +18.4 of the chromosomal walk and has a 4.3-kb open reading frame that encodes a polypeptide of 155 kD.

The 1.9-kb transcript: Clone 2W was isolated during the screen of the oligo-(dT) primed cDNA library. This clone, composing cDNA D, is 1.6 kb in length and contains five exons. Northern analysis had established that probes of exons A3, A4, A5 and A6 hybridize to the 1.9-kb transcript. In accordance with these findings, sequence analysis indicates that the four terminal exons of clone 2W are the same as those

Figure 5.—Bristle phenotypes of forked mutants. Scanning electron micrographs showing dorsal thorax of Canton S, f, f
d, f, f, f, f and f flies. The forked phenotype of the flies ranges from the mild f mutant to the severe f and f mutants.
in transcript A. A fragment that represents exon D1 (Figure 1), hybridizes on Northern blots to both the 1.9- and 1.1-kb RNAs, suggesting that cDNA D, which contains clone 2W, represents the 1.9 kb transcript. Genomic sequence analysis indicates that a 90-bp open reading frame extends upstream of the 5' end of clone 2W. This open reading frame contains three translation start sites. Sequences flanking the first of these sites is most similar to sequences that are adjacent to the translation start sites of other Drosophila genes (CAVENER and RAY 1991). Assuming that this is the 5' most exon of the 1.9-kb transcript and that the ATG site within this open reading frame is the translation start site of this transcript, the open reading frame of the 1.9-kb RNA extends 1.4 kb and encodes a 53-kD protein.

The 1.1-kb transcript: The structure of the 1.1-kb RNA was determined by Northern analysis using exon specific probes that were generated by PCR. No cDNAs were isolated for this RNA, and therefore the structure of this transcript is only tentative. The 1.1-kb transcript hybridizes to probes representing exons D1, A3 and A6. Exon A6 was further divided into two equal 550-bp fragments and each were used as probes. The 1.1-kb RNA hybridizes only to the fragment corresponding to the terminal half of exon A6. These results show that the 1.1- and 1.9-kb RNAs probably share the same promoter and that alternative splice selection occurs in the 1.1-kb RNA. Analysis of consensus splice sequences suggests that the splice acceptor site for the last exon of the 1.1-kb RNA is the translation start site of this transcript, the open reading frame of the 1.9-kb RNA extends 1.4 kb and encodes a 53-kD protein.

The 6.4-kb transcript: Northern analysis indicates that the 6.4- and 5.4-kb RNAs hybridize to the same genomic and exon specific probes from interval −0.5 to +19.4 of the chromosomal walk. A genomic fragment that extends 500 bp upstream of the 5' end of cDNA clone 77X of the 5.4-kb RNA, from coordinates −0.5 to 0, encodes multiple stop signals in all reading frames and selectively hybridizes to both of these RNAs. Probing Northern with a 1.0-kb fragment that extends distally from −0.5 to −1.5, detects only the 6.4-kb RNA and reveals that the transcription start site of this RNA is upstream of that for the 5.4-kb RNA. Northern hybridizations using genomic fragments that extend 10 kb upstream of interval −1.5 as probes do not detect forked RNAs. Therefore, the transcription start site of the 6.4-kb RNA should be located between the XhoI site at position 0 and the BamHI site at −2.0 kb. Sequence of the genomic DNA in this region reveals no significant open reading frames. Based on these results, a tentative structure for the 6.4-kb transcript is represented in Figure 1, assuming that the only difference between the 5.4- and 6.4-kb RNAs is a longer transcribed untranslated region in the latter.

**Structural analysis of the forked proteins:** To determine if a function could be predicted from the structural domains of the forked products, the amino acid structure of the proteins encoded by the 2.5-, 5.4- and 5.6-kb mRNAs was analyzed. A feature common to all three of these proteins is the presence of stretches of at least 16 hydrophobic residues that may represent regions of the protein spanning the cell membrane. Two such hydrophobic stretches occur in exon A6 (Figure 1) which is common to all three RNAs encoded by nucleotides 17,714–17,777 and 17,909–17,972 of the genomic sequence (Figure 3). The coding region shared by the 5.6- and 5.4-kb transcripts has as many as three of these hydrophobic stretches in exon B4 (Figure 1), spanning nucleotides 9,285–9,348, 9,450–9,512 and 9,822–9,870. The 5' exon of the 5.4-kb RNA also contains a significant stretch of residues that commence at nucleotide 458 and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.

The characterized transcripts of the forked locus contain several regions in which there are concentrations of specific residues. Thirteen of the 24 amino acids in exon A3, between nucleotides 16,558 and 16,630 in the genomic sequence, encode proline residues. The 5.4- and 5.6-kb RNAs have glycine rich stretches encoded in exon B4, between nucleotides 9,267–9,341 and 9,444–9,482 in which 13 of 25 amino acids and 11 of 13 amino acids are glycine, respectively. Polyglycine tracts are devoid of structure and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.

The characterized transcripts of the forked locus contain several regions in which there are concentrations of specific residues. Thirteen of the 24 amino acids in exon A3, between nucleotides 16,558 and 16,630 in the genomic sequence, encode proline residues. The 5.4- and 5.6-kb RNAs have glycine rich stretches encoded in exon B4, between nucleotides 9,267–9,341 and 9,444–9,482 in which 13 of 25 amino acids and 11 of 13 amino acids are glycine, respectively. Polyglycine tracts are devoid of structure and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.

The characterized transcripts of the forked locus contain several regions in which there are concentrations of specific residues. Thirteen of the 24 amino acids in exon A3, between nucleotides 16,558 and 16,630 in the genomic sequence, encode proline residues. The 5.4- and 5.6-kb RNAs have glycine rich stretches encoded in exon B4, between nucleotides 9,267–9,341 and 9,444–9,482 in which 13 of 25 amino acids and 11 of 13 amino acids are glycine, respectively. Polyglycine tracts are devoid of structure and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.

The characterized transcripts of the forked locus contain several regions in which there are concentrations of specific residues. Thirteen of the 24 amino acids in exon A3, between nucleotides 16,558 and 16,630 in the genomic sequence, encode proline residues. The 5.4- and 5.6-kb RNAs have glycine rich stretches encoded in exon B4, between nucleotides 9,267–9,341 and 9,444–9,482 in which 13 of 25 amino acids and 11 of 13 amino acids are glycine, respectively. Polyglycine tracts are devoid of structure and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.

The characterized transcripts of the forked locus contain several regions in which there are concentrations of specific residues. Thirteen of the 24 amino acids in exon A3, between nucleotides 16,558 and 16,630 in the genomic sequence, encode proline residues. The 5.4- and 5.6-kb RNAs have glycine rich stretches encoded in exon B4, between nucleotides 9,267–9,341 and 9,444–9,482 in which 13 of 25 amino acids and 11 of 13 amino acids are glycine, respectively. Polyglycine tracts are devoid of structure and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.

The characterized transcripts of the forked locus contain several regions in which there are concentrations of specific residues. Thirteen of the 24 amino acids in exon A3, between nucleotides 16,558 and 16,630 in the genomic sequence, encode proline residues. The 5.4- and 5.6-kb RNAs have glycine rich stretches encoded in exon B4, between nucleotides 9,267–9,341 and 9,444–9,482 in which 13 of 25 amino acids and 11 of 13 amino acids are glycine, respectively. Polyglycine tracts are devoid of structure and includes three residues of the next exon. Hydrophobic stretches of this size are characteristic of transmembrane spanning peptides, α-helices, and membrane associated proteins. It is appealing that these hydrophobic stretches are present in proteins that are involved in the formation of microfibrillar structures of bristles that are apposed to the cell membrane.
has been used to probe developmentally staged Northern blots. This screen demonstrated that many other transcripts contain the PRD motif, but its role has not been determined (FRIGERIO et al. 1986).

The forked gene has two copies of a sequence defined by the presence of a repeated (CAX), motif, where X is the nucleotide G, A or C, and n usually has a value of less than 30. These repeats are located within exon B4 (Figure 1) which is the fourth exon of the 5.4- and 5.6-kb RNAs. The (CAX), motif is present eighteen times from nucleotides 9,078 to 9,135 in the genomic sequence (Figure 3). The second repeat is located near the 3' end of the exon and contains thirty-two of these codons from nucleotides 9,987-10,094. This repeat has been found in organisms ranging from yeast to humans, in genes that are developmentally or tissue specifically regulated (WHARTON et al. 1985a; SCHULTZ and CARLSON 1987; SUZUKI et al. 1988; CHANG et al. 1987; DUBOULE et al. 1987). The (CAX), motif encodes a polyglutamine stretch that is often interrupted by histidine residues and has been recognized under the names of opa (WHARTON et al. 1985a,b), strep (KIDD, LOCKETT and YOUNG 1983), and the M repeat (SCHNEWLY et al. 1986). In D. melanogaster, the opa repeat occurs at least 500 times in the haploid genome and has been found to be expressed differentially in a developmental profile (WHARTON et al. 1985b). The opa motif is associated with homebox containing genes such as Antennapedia (LAUGHON et al. 1986), Deformed (REGULSKI et al. 1985), cut (BLOCHINGER et al. 1988), engrailed (KUNER et al. 1985) and fushi tarazu (LAUGHON and SCOTT 1984). In the transcription factor Sp1, two glutamine rich domains, containing approximately 25% glutamine residues, have been shown to be potent activators of transcription (COUREY et al. 1989). A chimaera containing the Sp1 zinc finger binding domain and the nonhomologous opa repeat of Antennapedia was observed to synergistically stimulate the expression of the SV40 enhancer, suggesting that the repeat may serve an evolutionary conserved structural or functional role (COUREY et al. 1989).

Products of the 5.6- and 5.4-kb RNAs contain ankyrin-like repeats: A search of the Genbank protein database revealed significant homology between regions common to the three largest forked products and polypeptide stretches of proteins known to contain a repeated motif found in the ankyrin protein (LUX, JOHN and BENNETT 1990). The products of the 5.6- and 5.4-kb RNAs contain four and five copies of this ankyrin-like repeat, respectively. The motif has also been recognized as the Cdc10/SW16 repeat (BREEDEN and NASMYTH 1987). The amino terminally situated forked ankyrin-like repeats are arranged in tandem. This repeated motif is encoded in the first four exons of the 5.4-kb transcript and the second, third, and fourth exon of the 5.6-kb RNA (Figure 3). When conservative residue substitution is taken into consideration (Figure 4), the forked ankyrin-like repeats show significant homology to the ankyrin-like consensus (LAMARCO et al. 1991). Additionally, position two is usually an arginine residue within the forked ankyrin-like repeats. As observed in the consensus for the 22 ankyrin repeats (LUX, JOHN and BENNETT 1990) and the six ankyrin-like repeats in the product of the Caenorhabditis elegans gene fem-1 (SPENCE, COULSON and HODGKIN 1990), position 17 of the forked ankyrin-like consensus is typically a glycine residue. Position 22 of the ankyrin-like consensus is a leucine residue as is the case of the consensus for the seven copies of this repeat motif in the product of the human bel-3 gene (OHNO, TAKIMOTO and MCKEITHAN 1990). The length of ankyrin-like repeats has generally been found to be 33 residues, but varies from the half repeats in the human NF-kB binding protein (BOURS et al. 1990), to the presence of noncontiguous repeats in the products of the yeast genes cdc10 and SW16 (AVIS et al. 1985; BREEDEN and NASMYTH 1987). Ankyrin-like repeats of varying numbers and lengths have been found in many classes of proteins in organisms ranging from yeast to humans. The yeast cell cycle control genes cdc10, SW14 and SW16, encode proteins that contain two noncontiguous repeats (BREEDEN and NASMYTH 1987; ANDREWS and HERSKOWITZ 1989). Products of the Notch locus of D. melanogaster and the lin-12 and glp-1 genes of C. elegans are involved in tissue differentiation and contain six contiguous ankyrin-like repeats in their cytoplasmic domains (GREENWALD 1985; YOCHEN and GREENWALD 1989). The fem-1 gene product of C. elegans is involved in sex determination and contains six contiguous ankyrin-like repeats (SPENCE, COULSON and HODGKIN 1990). The ankyrin-like motif has been observed in transcription factors as well. Studies of the heterodimeric DNA binding protein GABP indicate that the amino terminus of the β subunit contains four contiguous ankyrin-like repeats (LAMARCO et al. 1991). These repeats help stabilize GABP β interaction with the α subunit (THOMPSON, BROWN and MCKNIGHT 1991).

Analysis of spontaneous forked mutations: To analyze the regulatory and coding regions of the forked gene further and to determine the mechanism of mutagenesis by the insertion of transposable elements, eight spontaneous forked mutations were characterized. Genomic DNA was isolated from f1, f5, f8, fL6a, fL4b, fL4c, fL4d and f3 mutant stocks. This DNA was then treated with restriction endonucleases. Genomic Southern containing DNA isolated from Canton S and mutant individuals were prepared and analyzed for alterations using fragments from the 50 kb of
contiguous DNA in the chromosomal walk of the forked locus as hybridization probes (Figure 1). In all cases, insertions were found within the analyzed region. The characterized mutations, with the exception of the $f^{17k}$ and $f^3$ alleles, have insertions located within a 6.2-kb EcoRI SalI genomic fragment from interval +9.2 to +19.2 of the chromosomal walk. Insertions in the $f^{17k}$ and $f^3$ alleles are located 2.6 and 9.3 kb upstream of the 6.2-kb region, respectively (Figures 1 and 3).

A minimum of 11.7 kb separates the transposable element insertion site of the $f^3$ mutation from that of the $f^1$, $f^3$ and $f^{36a}$ mutations. This distance between members of the left and right pseudoallelic series explains the intra-allelic recombination that has been observed to occur within the forked locus (GREEN 1955). Furthermore, mutant characterization shows that the forked locus is oriented with its 5' end closest to the telomere. This orientation is supported by the physical map of a phage walk spanning the $Df(1)B^{265-20}$ deletion, which includes the forked locus (HIGASHIJIMA et al. 1992).

The $f^3$ allele is a member of the left pseudoallelic series and has an insertion in the forked locus that is the most distal of those characterized (Figure 1). The bristle phenotype of $f^3$ mutants is the weakest of all forked alleles characterized in this study, with macrochaetae slightly bent (Figure 5). This mutation was cloned by screening a genomic library prepared from a $f^3$ mutant stock with a 1.6-kb EcoRI XhoI genomic fragment located between coordinates +2.6 to +4.1 of the chromosomal walk (Figure 1). Restriction and sequence analysis of clones isolated from the screening of this library indicate that the $f^3$ allele has a full length 7.6-kb 412 retrotransposon (YUKI et al. 1986) inserted at the locus in the same orientation as forked. This insertion is located at nucleotide 3,914, within the first intron of the 5.4-kb transcript and upstream of the other forked transcripts (Figures 1 and 3).

$f^{17k}$ mutants have a weak bristle phenotype, as only macrochaetae are affected (Figure 5). Clones containing the forked locus of this allele were obtained by screening a genomic library made with DNA from $f^{17k}$ flies with a 3.8-kb EcoRI genomic fragment located between coordinates +9.2 to +14.0 of the chromosomal walk (Figure 1). Characterization of isolates obtained from this screen by restriction and sequence analysis reveals that a 2.9-kb hobo element (STRECK, MACGAFFEY and BECKENDORF 1986) is inserted in the same orientation as forked in this allele. The insertion site of the hobo element is within the fifth exon of the 5.6- and 5.4-kb RNAs and upstream of the smaller transcripts. The hobo element is inserted at nucleotide 10,608 (Figures 1 and 3).

The $f^{36a}$ allele was generated by P element-induced hybrid dysgenesis (D. DORER and A. CHRISTENSEN, personal communication). There is no sign of reduced viability or fertility in the $f^{36a}$ stock, but the phenotype of the macrochaetae, microchaetae, and trichomes are severely affected (Figure 5). The bristle phenotype of females that are heterozygous for $f^{36a}$ and the deletion $Df(1)B^{265-20}$ which includes all of the forked coding region is indistinguishable from homozygous $f^{36a}$ females. Screening of a genomic phage library prepared with DNA that was isolated from a $f^{36a}$ stock with the 6.2-kb EcoRI SalI genomic fragment provided clones containing the insertion. Sequence analysis of these clones shows that a 2.9-kb P element (O’HARE and RUBIN 1983) is inserted into the forked locus. The P element insertion is located 14 nucleotides upstream of a donor splice site for an exon that is common to the 6.4-, 5.6-, 5.4- and 2.5-kb transcripts at nucleotide 15,597 of the genomic sequence (Figure 3). This insertion is in the second exon of the 2.5-kb RNA and the seventh for the 5.4- and 5.6-kb transcripts. The P element is inserted in the same orientation as the forked gene (Figure 1).

Macrochaetae, microchaetae and trichomes are severely affected in the $f^{14k}$ and $f^{36a}$ alleles (Figure 5). Both of these mutations are the result of transposable element insertions into an exon that has been determined by screening mid-late pupal poly(A+) RNA of Canton S flies to be present in all forked transcripts. Analysis of clones isolated from a genomic library prepared from the DNA of $f^{14k}$ flies shows that a full length gypsy retrotransposon (MARLOR, PARKHURST and CORCES 1986) has inserted in an opposite orientation to the forked gene. This insertion is at nucleotide 16,517 of the genomic sequence (Figures 1 and 3). The screening of a genomic phage library prepared from the $f^{36a}$ mutant stock with the 6.2-kb EcoRI SalI genomic fragment provided clones of the forked region in this mutant. Analysis of the clones indicates that a full length springer retrotransposon (KARLIK and FYRBERG 1985) is inserted at nucleotide 16,513, four nucleotides upstream of the $f^{14k}$ insertion (Figure 3). The springer insertion is oriented opposite to the forked locus. The insertion sites for both the $f^{14k}$ and $f^{36a}$ alleles are within the coding region containing the PRD repeat.

Unlike the $f^{14k}$ insertion which results from a gypsy insertion into an exon, insertion of the gypsy retrotransposon into an intron of the forked gene in $f^1$, $f^3$ and $f^{36a}$ alleles renders the expression of the bristle phenotype under the control of second site modifier loci. The gypsy insertion of the $f^1$ allele is located within an intron of the characterized transcripts, at nucleotide 15,673 of the genomic sequence (Figures 1 and 3) (HOOVER et al. 1992). This element is inserted in the same orientation as forked. The macrochaetae and microchaetae are mildly affected in the $f^1$ allele (Figure 5).
The \( f^3 \) mutation has a strong bristle phenotype. The macrochaetae and microchaetae are strongly gnarled in \( f^3 \) flies (Figure 5). Southern analysis of genomic DNA cut with the EcoRI restriction endonuclease indicates that two insertions occur within the 23-kb SalI genomic fragment from the -4 to +19.4 interval of the genomic walk (data not shown). A genomic phage library was prepared from a \( f^3 \) stock and screened with the 6.2-kb EcoRI-SalI fragment. Sequence analysis of the clones obtained from this screen indicates that two full-length gypsy retrotransposons are inserted in the same orientation as the forked gene. One of the gypsy insertions is located at the identical site and orientation as that observed for the \( f^1 \) gypsy insertion (Figure 1). The second gypsy insertion of the \( f^3 \) allele is shown by Southern and sequence analysis to be located within the 3' transcribed untranslated region of all forked transcripts, at nucleotide 18,166 of the genomic sequence (Figures 1 and 3). Bridges (1938) isolated the \( f^3 \) allele in a screen 22 years after his initial isolation of the \( f^1 \) mutation. Molecular characterization of the \( f^{m1} \) allele, a third independently derived mutation, indicates that it also contains a gypsy insertion at the identical site and orientation as that observed in the \( f^1 \) and \( f^3 \) alleles. The \( f^{m1} \) allele arose spontaneously (T. Gersgimova, personal communication) and has a weak bristle phenotype. The insertion site common to all of these alleles appears to be a preferred location in the forked locus for gypsy mutagenesis. The TATA target site of these insertions is the same as that recognized by the gypsy insertion of the \( f^K \) allele (see below) and the second \( f^3 \) gypsy retrotransposon insertion. The target site of the gypsy insertion in the \( f^{14} \) allele also consists of alternating pyrimidines and adenines, but is CACA.

The weak bristle phenotype of \( f^K \) mutants is similar to that of the \( f^1 \) allele (Figure 5). Genomic phage libraries were prepared from DNA isolated from a \( f^K \) mutant stock. The library was screened with the 6.2-kb EcoRI-SalI genomic fragment from interval +13.0 to +19.2 of the chromosomal walk. Sequence characterization of the clones isolated from this screen show that a gypsy retrotransposon is inserted at nucleotide 15,821 in the genomic sequence, 148 bp downstream of the insertion site of the gypsy elements in the \( f^1 \) and \( f^3 \) alleles (Figures 1 and 3). This element is oriented in the same manner as the forked gene.

**Effects of transposable element insertions on the expression of the forked gene:** Northern analysis of poly(A*) RNA isolated during pupal stages of development, was carried out for each of the mutant stocks to study the correlation between RNA levels of the forked gene and the phenotypic severity of each mutation. These experiments were also undertaken to further characterize the involvement of transposable elements in eliciting a mutant phenotype. Northern blots were probed with the 6.2-kb EcoRI-SalI genomic fragment containing the coding region of the Drosophila ras2 gene. This gene encodes a 1.6-kb transcript that is expressed at approximately constant levels throughout development and serves as a control for the amount of RNA loaded in each lane (Mozzer et al. 1985).

The mutations have been separated into three groups according to severity of bristle phenotype and identity of the transposable element responsible for the mutation. The weak phenotypes of \( f^3 \) and \( f^{174} \) alleles are the result of similar alterations in the developmental profile of forked RNAs. The 6.2-kb EcoRI-SalI genomic probe does not detect the 6.4-, 5.6- or 5.4-kb transcripts in RNA isolated from a \( f^3 \) mutant stock (Figure 6, Table 1). In this allele, the 412 element insertion is within an intron of the 6.4- and 5.4-kb RNAs that is upstream of the coding regions for the 5.6-, 2.5-, 1.9- and 1.1-kb RNAs (Figure 1). The 412 insertion does not alter abundance of the 2.5-, 1.9- or 1.1-kb RNAs. As is the case for the \( f^3 \) mutation, the 6.4-, 5.6- and 5.4-kb RNAs are also absent from the pupal RNA of \( f^{174} \) flies in which a hobo element has inserted into an exon of these transcripts (Figures 1 and 3). Quantification of transcript levels using ras2 as a standard shows that the accumulation of the 2.5- and 1.9-kb transcripts is

![Figure 6](image_url).
affected in f17k flies. In this mutant, the amount of the 2.5-kb RNA is almost one-third and that of the 1.9 kb RNA is approximately two-thirds of the same size transcript in Canton S flies (Figure 6, Table 1).

The severity of the bristle phenotype in the fhd and f36a mutations is the result of the insertion of a transposable element into an exon of the forked gene. The 6.4-, 5.6-, 5.4- and 2.5-kb RNAs are not detected when a Northern containing poly(A) RNA isolated from pupal stages of development from fhd and f36a stocks was probed with the 6.2-kb EcoRI-SalI fragment (Figure 7). In the fhd mutant, abundance of the 1.9- and 1.1-kb RNA is the same as that observed for wild-type flies (Table 1). The 5′ most exon of these 1.1- and 1.9-kb RNAs is downstream of the site of the P element in the fhd allele and these RNAs are the only forked transcripts within which the coding region has not been disrupted (Figure 1). A signal of approximately 1.1 kb is present in the f36a mutant stock (Figure 7). This signal is amplified eightfold over that observed for the equivalent sized forked transcript in wild-type flies. Longer exposure of the membrane also detects a band corresponding to a transcript that is 4.2 kb. These forked RNAs detected in f36a flies result from premature termination of transcription. The 4.2-kb RNA in this mutant is detected with probes that are specific to the 5.6- and 5.4-kb RNAs. A probe containing the 5′ most exon of the 2.5-kb RNA hybridizes to a 1.1-kb RNA in the f36a mutant stock (Figure 8B). Northern analysis using the long terminal repeat (LTR) of the springer element as a probe shows hybridization to the same RNAs as the 6.2 kb EcoRI-SalI fragment in addition to the expected RNAs of the element (data not shown). These results indicate that all forked RNAs are truncated within the 3′ LTR of the springer element. These findings are supported by the presence of a polyadenylation signal on the nontranscribed strand of the springer LTR (Karlik and Fryberg, 1985).

The insertion of a gypsy retrotransposon into an intron of the forked gene in the f1 mutation lowers the abundance of the five largest RNAs (Figure 8A). The 1.1-kb RNA of this mutant is elevated 1.4-fold over that observed in the Canton S stock (Table 1). The Northern blot in Figure 8 was probed with a fragment specific to the 2.5-kb transcript and found to hybridize to an 1.1-kb RNA in f1 flies (Figure 8B). This fragment hybridizes selectively to the 2.5-kb RNA in wild-type flies (see lane marked CS-8 in Figure 8B), suggesting that the 1.1-kb transcript arises by premature termination of the 2.5-kb RNA. To examine the role of the gypsy insertion in truncation of

### TABLE 1

**Summary of forked phenotypes and quantitation of forked RNA levels in Canton S and mutant stocks**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Phenotype</th>
<th>Nature of mutation</th>
<th>Effects on RNA levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>6.4 kb</td>
</tr>
<tr>
<td>CS</td>
<td>Wild type</td>
<td></td>
<td>ND</td>
</tr>
<tr>
<td>f1</td>
<td>Weak</td>
<td>412</td>
<td>5</td>
</tr>
<tr>
<td>f17k</td>
<td>Weak</td>
<td>hobo</td>
<td>ND</td>
</tr>
<tr>
<td>f36a</td>
<td>Strong</td>
<td>springer</td>
<td>ND</td>
</tr>
<tr>
<td>f412</td>
<td>Strong</td>
<td></td>
<td>ND</td>
</tr>
<tr>
<td>f36a</td>
<td>Intermediate</td>
<td>gypsy</td>
<td>10</td>
</tr>
<tr>
<td>f36a</td>
<td>Intermediate</td>
<td>gypsy</td>
<td>1</td>
</tr>
<tr>
<td>f36a</td>
<td>Strong</td>
<td>gypsy</td>
<td>1</td>
</tr>
</tbody>
</table>

Data summarize the quantitation of forked mRNAs in the strains examined. Values represent the average of at least three different preparations. RNA levels were determined using a Molecular Dynamics PhosphorImager or a densitometer. The value obtained for each recording was divided by that received for the ras2 control in the lane. A value of 100 has been assigned to the amount of 2.5-kb RNA in wild-type flies to which other values have been normalized. ND indicates not detectable.
The Drosophilaforked Gene

FIGURE 8.—Analysis of forked RNA in f' and f36a mutants. A Northern was prepared from poly (A') RNA (10 µg) of pupal stages of development isolated from Canton S, f' and f36a stocks. Numbers at the top of each lane indicate the age of the flies in days after egg laying; day 7 corresponds to young pupae in Figure 1, day 8 to mid pupae, and day 9 includes the mid-late pupae stages. The membrane was hybridized with the 6.2-kb EcoRI-Sall genomic fragment (A); the exon A1 (Figure 1) specific to the 2.5-kb RNA (B); the long terminal repeat of the gypsy element (C); and the ras2 gene as an internal standard (D). The arrowhead in panel C points to the 1.1-kb transcript that initiates in the forked promoter and terminates in the gypsy 5' LTR.

FIGURE 9.—The profile of forked mRNAs in fK flies. A Northern containing poly(A') RNA of Canton S and fK flies was probed with the 6.2-kb EcoRI-Sall genomic fragment (A), the A1 exon specific for the 2.5-kb RNA (B), or the ras2 gene as a control for loading (C). Numbers at the top of each lane indicate the age of the flies in days after egg laying; day 7 corresponds to young pupae in Figure 1, day 8 to mid pupae, and day 9 includes the mid-late pupae stages.

Examination of forked RNAs in the fK mutation indicates that all transcripts are reduced in abundance when compared to Canton S (Table 1). When used as a probe, the 6.2-kb EcoRI-Sall fragment detects RNAs that are 4.0 and 4.2 kb in addition to the other forked RNAs (Figure 9A). Northern analysis using genomic fragments that are proximal to the gypsy insertion as probes do not detect the 4.0- and 4.2-kb RNAs. Further hybridizations of this Northern with probes containing the first exons of the 5.6-, 5.4- and 2.5-kb RNAs indicate that reduced levels of normal sized transcripts are being produced and that some of the RNAs from each of these promoters are truncated to form transcripts that are 1.4 kb smaller than native forked RNAs (Figure 9B). In agreement with the f' data, the three truncated RNAs in fK mutants are detected when the LTR of the gypsy retrotransposon is used as a probe (data not shown). Hybridization of the Northern in Figure 9 with a probe representing exon A3 that is common to all forked RNAs (Figure 1), indicates that levels of the 1.9- and 1.1-kb RNA are reduced when compared to Canton S (data not shown). The coding region of these RNAs is downstream of the gypsy insertion causing the fK mutation (Figure 1).

The f' mutation has a gypsy element insertion into the 3' transcribed, untranslated region of the forked gene in addition to a second gypsy insertion in the exact site and orientation as that for the f' allele (Figure 1). As predicted by the increased severity of the bristle phenotype in this mutant, the abundance
of the *forked* RNAs is further reduced from that observed for the *f* mutant (Figure 10A, Table 1). Like the *f* mutant, some of the transcripts generated from the 5.6-, 5.4- and 2.5-kb promoters in the *f* stock are prematurely terminated and give rise to RNAs that are 1.4 kb smaller than normal. In this mutant, truncation of *forked* RNA in the 5' LTR of the downstream gypsy element would result in RNAs of the same size as wild type. A Northern containing *f* RNA was probed with exon A3 (Figure 1) present in all *forked* RNAs (Figure 10A). Use of this exon, which is located downstream of the gypsy insertion in the intron of *f* (Figure 1), allows for the examination of levels of those transcripts that are not truncated in this element. The blot was then stripped and reprobed with exon A1 (Figure 1), a probe specific to the 2.5-kb RNA (Figure 10B). These results indicate that most of the 1.1-kb RNA that is detected in the *f* mutant when exon A3 is used as a probe is a truncated transcript generated from the 2.5-kb promoter.

**DISCUSSION**

Data presented here show that the *forked* locus encodes at least six different transcripts, and that the coding regions of these RNAs span 21 kb of genomic DNA. Analysis of spontaneous mutations in this study shows that changes in *forked* transcription result in disfiguration of chaetae and trichomes. Correlations exist between the presence or absence of specific *forked*-encoded mRNA and the severity of the mutant phenotype.

The terminally differentiated bristle cell is hollow and elliptical in cross section (Ribbert 1972). The mechanical support necessary to generate this anisometric cell during mid to late stages of pupal development is presumably provided by nine to twelve regularly spaced fiber bundles that run in parallel to the long axis of the bristle shaft close to the cell surface. Electron microscopy of the developing bristle shaft in Oregon R flies shows that these fiber bundles are round, with a diameter of approximately 500 nm, and constitute 20% of the cross sectional area of the bristle (Overton 1967). The fiber bundle structures are temporary and only apparent between forty and fifty-three hours after puparium formation. The disappearance of these structures corresponds to the appearance of a dense cuticle around the periphery of the bristle shaft. Sclerotization and layers of cuticle provide the structural support necessary to maintain the terminally differentiated hollow bristle. Similar fibrillar structures are found during the development of scales in Ephestia (Overton 1966) and the footpads of Sarcophaga (Ribbert 1972).

Electron micrographs of bristle rudiments from the severe *f* mutant, taken at 5-hr intervals of pupal development, fail to detect the temporary fibrillar structures seen in the bristle shafts of wild-type flies (N. Petersen, personal communication). Immunostaining of the bristle rudiments with polyclonal antibodies generated from the five amino terminal exons of the 2.5-kb *forked* mRNA localize to these fibrillar structures, but there is an absence of antibody binding in the bristle rudiments of *f* mutants (Petersen et al. 1993). This implicates *forked* products in the formation and/or maintenance of the fibrillar structures.

The stretches of hydrophobic amino acids and ankyrin-like repeats in some *forked* proteins may help to explain the role of these proteins during assembly of the fibrillar structures. The *forked* protein may be membrane associated, in which case, the presence of stretches of hydrophobic amino acids could represent regions of the protein that span the cell membrane or interact with membrane associated proteins. The ankyrin-like repeats and/or other structural motifs within the *forked* proteins, could interact with proteins of the fibrillar complex and anchor them to the cell membrane. The ankyrin-like repeats present in the proteins encoded by the 6.4-, 5.6- and 5.4-kb mRNAs might serve a function analogous to the amino terminal domain of human erythrocyte ankyrin that contains 22 of these repeats. This domain binds to integral membrane proteins (Davis and Bennett 1984; Weaver, Pasternack and Marchesi 1984) and tubulin (Bennett and Davis 1981). The *forked* proteins that contain these repeats may likewise interact with the central microtubules and/or integral membrane proteins present on the cell membrane that are closely
associated with the fibrillar bundles. This model predicts that mutations in the forked gene result in a failure of the fibrillar assembly to anchor properly to the cell membrane, compromising the structural support. Stress created by cytoplasmic pressure within the bristle rudiment causes buckling of the shaft and results in a forked phenotype.

Alternatively, the forked products may be components of the fibrillar structures. The hydrophobic stretches and the ankyrin-like repeats of the forked proteins may be involved in protein/protein interactions necessary for the formation or packaging of the fibrillar structures. The ankyrin-like repeats of the 6.4-, 5.6- and 5.4-kb RNAs may serve a role in stabilizing the fibrillar structures by interacting with microtubules or integral membrane proteins. The region of the forked proteins most likely to interact to form fibrillar structures is within the carboxy terminal amino acids common to the five largest forked products. This model predicts that mutations in the forked gene would result in insufficient levels or an imbalance of materials required for normal fibrillar assembly.

Characterization of the effects of transposable element insertions on the expression of the forked gene clearly indicates that the 6.4-, 5.6- and 5.4-kb transcripts are necessary for normal bristle development. In the f allele, a 412 element is inserted into an intron of the 5.4- and 6.4-kb RNAs, upstream to the promoters of the other forked RNAs (Figure 1). The 6.4-, 5.6- and 5.4-kb RNAs are absent in f flies, but the other forked RNAs are unaffected. The f mutant has a weak bristle phenotype. Interestingly, elevated levels of the 2.5-, 1.9- and 1.1-kb forked RNAs are capable of rescuing the severe bristle phenotype of f flies (Petersen et al. 1993). Higher levels of these RNAs, which encode proteins that do not have ankyrin-like repeats, are sufficient for normal bristle development. The 12.8-kb BamHI-SalI genomic fragment from the interval +6.4 to +19.2 of the chromosomal walk (Figure 1) was cloned into the CaSpeR cloning vector, and introduced into a forked mutant stock through P element-mediated germline transformation. When the construct is expressed at high levels, the mutant phenotype of f is rescued (Petersen et al. 1993), leading to the conclusion that the ankyrin-like repeats are dispensable, but abundance of the forked protein is critical for normal bristle development.

The insertion of a transposable element into an exon of forked causes the severe phenotypes of the f and f alleles. In the f mutant, a P element is inserted in the same orientation as the forked gene. The insertion site of this element is near the 3' end of exon A2 (Figure 1) that is common to the 6.4-, 5.6-, 5.4- and 2.5-kb RNAs, which are absent in this mutant. This absence may result from effects of the P element on transcription of the promoters of these RNAs or RNA instability resulting from insertion of the element. In contrast, levels of the 1.9- and 1.1-kb RNAs which are initiated downstream of this insertion site are unaffected. Results from Northern analyses of f and f suggest that the severity of the forked phenotype correlates with levels of the 2.5-kb RNA. Southern analysis of DNA isolated from a f stock shows that the springer insertion responsible for this mutation is in exon A3 (Figure 1) common to all forked RNAs. This element is oriented opposite to forked. The RNA in f mutants is truncated. This truncation is caused by recognition of a polyadenylation signal located in the 3' LTR of the springer element.

In the f, f and f alleles, the insertion of a gypsy retrotransposon into an intron of forked is responsible for the mutant phenotype. There is a reduction in the native levels of forked RNAs in these mutants. The mechanism(s) by which the gypsy element acts to reduce the RNA levels in these mutants is unclear. The element may influence the expression of forked by truncating forked RNAs within the element, affecting stability of the forked RNA, or by altering transcription levels of the various forked promoters. Insertion of gypsy in f, f and f also affects the structure of forked RNAs by causing truncation of the transcript at a polyadenylation signal present in the 3' LTR of the gypsy retrotransposon. In the case of the f allele, this truncation leads to the appearance of 4.0 and 4.2-kb transcripts that are absent from the Canton S stock. Northern blots of the mRNA from f, f and f mutants analyzed with a probe specific for the promoter of the 2.5-kb transcript show hybridization to a 2.5-kb RNA and an additional 1.1-kb RNA. In contrast to forked RNAs in the f mutant, premature termination of forked RNAs within the LTR of the gypsy element in the f, f and f mutants is leaky, since low levels of native sized forked RNAs are present in the affected alleles. The fact that wild-type sized RNAs are detected in f, f and f mutants indicates that during RNA processing, the gypsy element is excised with the intron into which it has inserted. The presence of the gypsy in unprocessed RNA may cause transcript instability that would further reduce abundance of the wild-type sized forked RNAs. Truncation of transcripts is also found in gypsy induced mutants of the hsp2 gene (Dorsett et al. 1991) and the Hairy wing (Hu) mutant (Campuzano et al. 1985) where similarly oriented gypsy insertions into an intron or exon of the transcribed RNA respectively, result in truncation of the RNA from the host gene within the gypsy LTR. In addition, the gypsy retrotransposon may influence the normal transcription of forked RNAs. In the f mutant, levels of the RNAs generated from the 1.9- and 1.1-kb promoter are reduced compared to
Canton S. The coding regions of these RNAs are downstream of the \textit{gypsy} insertion site and are not disrupted by the insertion. Transcription of the larger \textit{forked} RNAs may similarly be influenced by regulatory sequences within the \textit{gypsy} element, as is the case for the \textit{yellow} and \textit{achaete} genes. In the case of \textit{yellow}, interactions between sequences within the \textit{gypsy} element that bind the sw(Hw) protein and enhancers of the \textit{yellow} gene in the \textit{y} \textsuperscript{f} allele are responsible for the mutant phenotype (CORCES and GEYER 1991). Similarly, the truncated \textit{achaete} RNA is elevated 5–20-fold over wild-type levels in the \textit{Hw} \textsuperscript{f} allele (CAMPUZANO et al. 1985), suggesting that the \textit{achaete} promoter is being stimulated by regulatory sequences within the \textit{gypsy} element. These observations, together with the results obtained from the analysis of \textit{gypsy}-induced \textit{forked} alleles, suggest that this retrotransposon might cause mutant phenotypes by a combination of mechanisms that include effects on the rate of transcription initiation, RNA stability, and termination of transcription.

We would like to thank S. PARKHURST and R. BURATOWSKI for help during the initial characterization of the \textit{forked} locus, and N. PETERSEN for communicating results prior to publication. We are indebted to M. SEPANSKI for help with the electron microscope, and A. CHRISTENSEN, D. DERER, T. GERASIMOVA, A. KIM and D. KUHN for gifts of plasmids and Drosophila strains. Our special gratitude goes to M. GREEN for many illuminating discussions; without his persistence, we would never have cloned \textit{f}. In spite of this persistence, we never characterized \textit{f} \textsuperscript{W}. This work was supported by U. S. Public Health Service grant GM35463 from the National Institutes of Health. The nucleotide sequence data reported here will appear in the EMBL, GenBank and DDBJ databases under the accession number X09871.

LITERATURE CITED


JACK, J., D. DORSETT, Y. DELLOTTO and S. LIU, 1991 Expression of the \textit{cut} locus in the \textit{Drosophila} wing margin is required for
The Drosophila forked Gene

525


WHARTON, K., K. JOHANSEN, T. XU and S. ARVANTAVIS-TSAKONAS, 1985a Nucleotide sequence from the neurogenic locus Notch implies a gene product that shares homology with proteins


Communicating editor: M. J. Simmons