Nucleotide sequence and structure of the sevenless gene of Drosophila melanogaster

David D.L. Bowtell, Michael A. Simon, and Gerald M. Rubin
Howard Hughes Medical Institute and Department of Biochemistry, University of California, Berkeley, California 94720 USA

In the Drosophila melanogaster mutant sevenless, the R7 photoreceptor in each of the repeating units, or ommatidia, of the compound eye fails to form. We have determined the nucleotide sequence and structure of the sevenless transcription unit. Overlapping cDNA clones from an eye imaginal disc library were isolated, and these together with corresponding genomic regions were sequenced. The positions of the major and two minor transcription start sites were mapped. The gene encodes a putative cell-surface receptor, the unmodified form of which is predicted as 288 kD, bearing a cytoplasmic tyrosine kinase domain. Several structural features distinguish sevenless from other tyrosine kinase receptors, most notably the large size of its extracellular domain. Moreover, unlike the insulin or epidermal growth factor (EGF) receptors, a putative amino-terminal signal sequence does not immediately follow the initiating methionine codon. Rather, a 21-amino-acid hydrophobic sequence is found 56 amino acids after the likely initiating methionine codon, suggesting that this sequence may function as an amino-terminal anchor. If so, the sevenless protein would be expected to have its 2001-amino-acid extracellular domain anchored as a loop at either end in the membrane. Finally, we have generated a number of transformant lines that carry either a 16.3-kb genomic fragment that extends ~950 bp upstream of the transcription start sites or constructs in which a subset of the introns present in the genomic sevenless DNA have been removed. The degree of rescue of the sevenless mutant phenotype has been assayed in these lines using morphological and behavioral assays.

[Key Words: Developmental biology; gene structure; positional information; protein tyrosine kinase; receptor]

Received March 18, 1988; accepted April 14, 1988.

The development of the compound eye of Drosophila requires the formation of a precise pattern of differentiated cell types. The formation of this pattern begins in the eye imaginal disc of third instar larvae, where cells are progressively recruited from an unpatterned epithelium to form the individual units, or ommatidia, of the eye (Ready et al. 1976). Each ommatidium contains 20 cells; a central core of eight photoreceptors is surrounded by pigment cells that insulate them optically and is covered by a crystalline cone that directs light onto them (Dietrich 1909). The recruitment and differentiation of these cells occur in response to positional cues, generated, sensed, and responded to by cells in the developing field. Components of these signaling pathways can, in principle, be identified by isolating mutant flies in which the developmental process is interrupted at specific stages. The best characterized mutant of this type is sevenless, in which the R7 photoreceptor is absent (Harris et al. 1976). In the absence of sevenless function, the precursor of the R7 cell forms a nonneuronal cell, a cone cell (Tomlinson and Ready 1986; Tomlinson and Ready 1987a). Mosaic analyses have shown the sevenless gene product to be cell autonomous, required either for recep-
cleotides predicts a protein of 288 kD. The sequence of the amino-terminal region suggests that the sevenless protein may have an unusual orientation in the membrane, the extracellular domain may be anchored at the amino terminus and carboxy terminus, producing a large loop in the extracellular domain, with the carboxyterminal kinase domain and a short amino-terminal sequence in the cytoplasm of the cell. The carboxy terminus of sevenless is homologous to many known tyrosine kinases but resembles the kinase domain of c-ros most closely (Birchmeier, et al. 1986; Neckameyer et al. 1986; Hafen et al. 1987). Comparison of the sevenless and partially determined human c-ros extracellular domain sequences (C. Birchmeier and M. Wigler, unpubl.) demonstrates that this homology extends at least part way into the extracellular domain. Finally, we have investigated the completeness of rescue of the sevenless phenotype in a large number of lines transformed with a 16.3-kb genomic fragment, shown previously to rescue the phenotype in a single line (Hafen et al. 1987), and with various constructs in which a subset of introns present in genomic sevenless DNA have been removed. These experiments indicate that the genomic fragment, which includes the transcribed region and only 950 bp of the promoter region of sevenless, is sufficient for rescue of the sevenless phenotype, as judged by antibody staining of eye imaginal discs, examination of the adult pseudopupil, and the behavioral response of adult flies to ultraviolet light.

**Results and discussion**

**Structure and expression of the sevenless gene**

The structure of the transcribed region was determined by the isolation and sequencing of 11 overlapping cDNA fragments and the sequencing of the majority of the corresponding 16.3-kb genomic DNA fragment (see Fig. 1A). The cDNA clones were initially mapped in relation to the genomic fragment and to each other by hybridization to restriction fragments. Then the sequences of clone 10, which extended the farthest 5', and of clone 1.2, which overlapped clone 10 by 185 nucleotides and extended 3' to the polyadenylation site (Fig. 2), were obtained on both strands. The genomic sequence was determined using a combination of shotgun sequencing and specific oligonucleotide primers (see Materials and methods). The structure of the transcription unit deduced from these data is given in Figure 1A. The composite cDNA and flanking genomic sequences for the sevenless gene and its predicted amino acid sequence are shown in Figure 2. The sevenless transcript is comprised of 12 exons and spans the majority of the 16.3-kb EcoRI fragment. The sequences surrounding the splice acceptor and donor sites (Fig. 1B) conform to splice consensus sequences identified previously (Mount 1982). The splicing pattern deduced from clones 10 and 1.2 was compared with the sequences of an additional nine cDNA clones that cover most of the transcription unit (exons 3 to 12, Fig. 1A). Their sequences essentially confirm the splicing pattern defined by clones 10 and 1.2, with the partial exceptions of clones 1 and 7; the first 54 nucleotides of intron 3 are present at the 5' end of clone 1, and in clone 7 the sixth intron is not removed (Fig. 1A). However, in each case, the presence of these additional sequences would lead to premature termination of translation. Therefore, these cDNAs most likely reflect incompletely processed RNA precursors of the mature sevenless mRNA.

An RNA blot of poly(A)^+ RNA from heads and bodies of adult Canton S flies and from third instar imaginal discs was probed with cDNA clone 10 and revealed a single transcript of 8.6 kb [Fig. 3, top panel], consistent with previous findings (Banerjee et al. 1987a; Hafen et al. 1987]. The estimated size of the transcript detected by RNA blotting [8.6 kb] was slightly smaller than the combined length of clones 10 and 1.2 [8.7 kb, excluding their overlap and the poly(A) tail of clone 1.2]. This difference probably reflects the relatively poor resolution of transcripts in this size range on a heavily loaded RNA gel. Contrary to previous reports (Banerjee et al. 1987a, Hafen et al. 1987), sevenless expression was not head specific but rather was expressed in poly(A)^+ body RNA at about one-third its abundance in head RNA. This result is not due to contamination of the body RNA preparation with head RNA, as a head-specific signal was obtained with a rhodopsin probe (Fig. 3, middle panel). The reason for the failure to detect the transcript in the body in previous reports is unclear. Differences in the methods by which RNA was prepared (guanidine HCl vs. phenol/chloroform extraction) could account for this result, particularly if the sevenless gene is expressed in tissues with high concentrations of endogenous nucleases. The sevenless phenotype is apparently limited to effects on eye development (Harris et al. 1976, Tomlinson and Ready 1986). The role of sevenless in the adult head and body, if any, remains to be established.

Three nonoverlapping oligonucleotide primers, complementary to sequences near the end of clone 10, were used in primer extension experiments to map the transcription start sites in head, body, and imaginal disc RNA. The lengths of the extended products obtained with each primer were in agreement with there being a major start site [48 bp] and two minor sites [76 bp and 137 bp] 5' to the end of clone 10, respectively (Fig. 4). The relative use of the three start sites was the same in head, body, and imaginal disc. To determine whether any additional introns lay between the 5' end of clone 10 and the start sites identified by primer extension, RNase protection was performed using a probe generated from a genomic EcoRV fragment that spanned the end of clone 10 and the putative start sites (see Fig. 2 for the position of this EcoRV fragment). In the absence of further splicing, protected fragments should result that are approximately 4 nucleotides smaller than the extended products obtained with the primer sevl. Poly(A)^+ head RNA protected three fragments from RNase digestion of the expected sizes, confirming the positions of the putative transcription start sites (Fig. 4D). The role of these alternative start sites is unclear; there is no apparent tissue specificity to their use, and no methionine codons
Figure 1. (See facing page for legend.)
appear between them and the major site that could represent alternate short translated sequences. Three polyadenylated cDNAs were isolated, each ending at the dinucleotide CT [nucleotide 8760, Fig. 2] (19 bp), 3′ to the sequence AATATA. This atypical polyadenylation signal has been reported in a number of other genes (for review, see Bimstiel et al. 1985), including the insulin receptor (Ullrich et al. 1985).

The sequence surrounding the major start site, P3 [ATCAGGT], is a perfect match with the consensus transcription site sequence for several Drosophila genes [ATCAGC–TCTG, Hultmark et al. 1986], except that the second, rather than the first A, appears to be the start site (Fig. 4A). The sequence surrounding P1 [TTCAGTT] is a 6/7 match with the consensus sequence, and again transcription initiates at the middle A. Between 17 and 29 nucleotides 5′ of the major start site, there is an AT-rich sequence flanked by GC bases, which is a 10/12 match with the proposed TATA box of the dopadecarboxylase gene [GCTTTAAAAGCC, Bray and Hirsh 1986]. Surrounding the transcriptional start site are several regions with repeated sequence motifs, including an ATCACGCTC motif that is repeated 12 times in a row and a TCGGT motif repeated 6 and 5 times in two blocks, respectively (Fig. 5). Downstream of the sevenless transcriptional start sites are other repeated elements, including one motif, ATCCAG, repeated five times (Fig. 5). Repeated sequences are found near the transcription start sites of a number of promoters, e.g., the sequence AAAGC is repeated seven times in the Drosophila c-abl gene [Henkermeyer et al. 1988]. Such repeated elements have been shown to bind factors that regulate transcription in several cases. For example, the Sp1 factor, which binds the repeat GCCGG, is important in the regulation of the SV40 early region promoter (reviewed McKnight and Tjian 1986). It is noteworthy that the sequence GCGGTTCGGT, present (17 bp) 3′ of the transcription start site of another gene expressed early in eye development, chaoptin (Reinke et al. 1988), is similar to one of the repeat elements found in the sevenless promoter, TCGGT.

Primary structure of the deduced sevenless protein

The previously deduced carboxy-terminal sequence of the sevenless protein identified a putative transmembrane domain followed by a domain with a high degree of homology to several protein tyrosine kinase receptors [Hafen et al. 1987]. Translation of the compiled cDNA sequence obtained here revealed a long open reading frame of 7677 nucleotides (2559 amino acids), extending to and encompassing this sequence (Fig. 2). It was anticipated that the presence of an amino-terminal signal sequence would help identify the likely initiating methionine codon. Examination of a hydrophobicity profile for the entire open reading frame revealed two distinct hydrophobic regions, one near the carboxy-terminal end, representing the putative transmembrane region adjacent to the kinase domain, and a second near the amino terminus (Fig. 6). The average hydrophathy of these sequences is +2.5 and +2.6, respectively, sequences with an average hydropathy of greater than +1.6 are considered to have a high probability of membrane insertion [Kyte and Doolittle 1982]. The open reading frame extends for 106 amino acids upstream of the amino-terminal hydrophobic sequence and includes three, in-frame methionine codons. However, the closest is 56 amino acids from the hydrophobic domain (Fig. 2). We favor this residue as the most likely initiation codon, as the sequence surrounding this codon most closely corresponds to the Drosophila consensus sequence for translation (Cavener 1987). Use of this initiation codon may result in a novel topology of the sevenless protein in the cell membrane (see below). The apparently untranslated leader of ~800-bp contains 10 methionine codons, of which 4 are surrounded by relatively good consensus sequences for the initiation of translation but are followed by stop codons within 4–11 amino acids. The remaining six methionine codons are surrounded by weak consensus sequences.

The sequence surrounding the carboxy-terminal end of the first hydrophobic sequence conforms to the proposed rules for formation of a signal cleavage site [von Heijne 1983, 1986a; For example, cleavage at the last cysteine in the underlined sequence would obey the −3,−1 rule, Fig. 2]. However, several features of this hydrophobic core, particularly the distance between a potential initiating AUG and the hydrophilic region, resemble more closely the amino-terminal anchor sequence of proteins such as the transferrin or asialoglycoprotein receptors (Schneider et al. 1984; Spiess and Lodish 1986) than the cleaved signal sequences of other tyrosine kinase receptors. It conforms to the comparison of amino-terminal anchor sequences [von Heijne 1986b] by having a relatively long hydrophobic core compared with cleaved signal peptides and a positively charged residue (histidine, residue 106) at the amino-terminal end of the sequence. Assuming translation initiates at the third methionine codon (residue 50), the amino-ter-

Figure 1. Intron-exon structure of the sevenless gene deduced from the sequence of 11 cDNA fragments and most of the 16.3-kb genomic EcoRI fragment. [A] Restriction map of the genomic fragment used to rescue the sevenless phenotype [Hafen et al. 1987] [E, EcoRI; E, EcoRV; H, HindIII; X, XhoI]. Bold lines indicate the portion of this fragment sequenced. Below is the deduced structure of the transcription unit and the 11 cDNA fragments from which it was derived. Numbering of these fragments refers to the cDNA sequence shown in Figure 2, beginning at the major transcriptional start site and ending at the polyadenylation site [nucleotide 8760]. Numbers in boldface type identify each clone. Clones 7 and 1 are partial exceptions to the deduced structure; clone 1 includes 54-bp of intron 3, and clone 7 retains intron 6. However, these clones probably represent partially processed RNA precursors (see text). [B] Genomic sequences surrounding the splice acceptor and donor sites and the length of each intron. Bases conforming to the consensus sequences [Mount 1982] for these sites are shaded. [C] Schematic representation of the various domains of the sevenless protein. [TM] Putative amino-terminal and carboxy-terminal transmembrane domains; [UT] 5′- and 3′-untranslated regions; [kinase] putative tyrosine kinase domain.
Figure 2. (See p. 626 for legend.)
Figure 2. (Cont’d.)
**Figure 2.** Composite nucleotide sequence of *sevenless* cDNA and flanking genomic DNA and the complete open reading frame of the deduced amino acid sequence. The EcoRI sites at either end of the sequence correspond to the 5' and 3'EcoRI sites flanking the genomic fragment diagramed in Figure 1A. Nucleotides in italics are derived from genomic sequences flanking the major site of deduced amino acid sequence. The EcoRI sites at either end of the sequence correspond to the 5'- and 3'EcoRI sites flanking the transcriptional unit, including clones 10 and 1.2, the ends of which are indicated. Intervening initiation of transcription (P3, nucleotide 1) and polyadenylation (nucleotide 8761). The sequence between nucleotides 49 and 8761 was derived from cDNAs that span the transcriptional unit, including clones 10 and 1.2, the ends of which are indicated. Intervening sequences are not included. Exon boundaries, determined from comparison of genomic and cDNA sequences, are indicated. Transcriptional start sites, determined by primer extension and RNase protection, are indicated (P1-P3). The sequence GGTTAAAAAGCC {TATA}_{box} of the dopa decarboxylase gene (GCTTTAAAAGCC; Bray and Hirsh 1980) which is a 10/12 match with the proposed TATA_{box} of the *sevenless* gene. The site of polyadenylation and preceding polyadenylation signal (AATATA) are underlined. Because the initiating methionine has not been determined, the translation of the complete open reading frame is shown, including three in frame methionine codons immediately preceding the hydrophobic sequence. However, careful sequencing of this region using 7-deaza-dGTP or dITP of both the cDNA and corresponding genomic sequences failed to reveal any sequencing errors that could put these codons in frame with the hydrophobic domain. Cysteine residues and potential sites of amino-linked glycosylation are shaded. Between residues 2043 and 2051, a stretch of arginines that may function as an endopeptidase cleavage site is underlined, followed by the putative tyrosine kinase domain. Nine differences between the cDNA and genomic sequences that result in amino acid differences are underlined, and the corresponding amino acid changes noted. Each of these changes was detected in at least two independent cDNA clones and is therefore unlikely to reflect cloning artifacts.

Amino-terminal region would be of similar length to other proteins anchored at their amino terminus. An amino-terminal anchor could result in the formation of a large loop in the extracellular domain, attached at the amino terminus and at the transmembrane domain adjacent to the kinase domain, with a short amino-terminal sequence and carboxy-terminal kinase domain on the cytoplasmic face of the cell. Several proteins appear to traverse the cell membrane multiple times, e.g., rhodopsin [Hargrave et. al. 1983; Nathans and Hogness 1983] and murine band 3 [Kopito and Lodish 1985]. However, few are thought to form a single loop, the aspartate receptor in bacteria [Russo and Koshland 1983] being one example. The putative 56-amino-acid amino-terminal cytoplasmic sequence of the *sevenless* protein includes 12 serines and 4 threonine residues that could act as sites of phosphorylation, perhaps modulating receptor function in a manner similar to that achieved by the phosphorylation of the epidermal growth factor (EGF) receptor by protein kinase C [Hunter et al. 1984]. Although the amino-terminal hydrophobic sequence is suggestive of a transmembrane anchor, we cannot exclude the possibility that this sequence is cleaved, in which case we would expect the *sevenless* protein to adopt a similar conformation in the membrane to other tyrosine kinase receptors. We are currently investigating the topology of *sevenless* in the membranes of tissue culture cells that express the protein under the control of heterologous promoters.

The unmodified protein is predicted to be ~288-kD, considerably larger than any other tyrosine kinase receptors. This predicted size is in good agreement with published
estimates of the protein detected in eye discs and larval brains (Banerjee et al. 1987b) and in Schneider L2 cells expressing sevenless protein (M. Simon and D. Bowtell, unpubl.). The proposed extracellular domain contains 17 potential sites for amino-linked glycosylation. The subunit structure of the protein has not been determined.

However, the presence of an additional 60-kD protein, specifically reactive with sevenless antisera on protein blots (Banerjee et al. 1987b; M. Simon and D. Bowtell, unpubl.), is consistent with sevenless being synthesized as a single polypeptide chain and cleaved into an α- and smaller β-subunit similar to the insulin receptor (Ullrich et al. 1985). If so, the stretch of arginine residues near the transmembrane domain [residues 2043–2051, Fig. 2] may function as the endopeptidase cleavage site, generating a β-subunit of ~57.5 kD. There are 43 cysteine residues in the predicted protein, and although 10 are clustered in one region near the amino terminus, there does not appear to be an obvious cysteine-rich domain that resembles that seen in several other receptors, e.g., the EGF or insulin receptors (Ullrich et al. 1984, 1985; Ebina et al. 1985).

Comparison of the genomic and cDNA sequences revealed a large number of polymorphisms, reflecting the strain differences in the origin of the cDNA and genomic libraries. These include 44 base substitutions throughout the 12 exons and, in the untranslated region (exon 1) of clone 10, an insertion of a T at nucleotide 194 and a deletion of 22 bp of a GT repeat at nucleotide 229 (see Fig. 2 for numbering). Only 9 of the 44 base substitutions resulted in changes in the amino acid sequence, with all changes being conservative ones [Fig. 2]. Interestingly, 13 nucleotide substitutions occurred in exon 2, 3' of the proposed initiation codon, but none resulted in changes in the amino acid sequence, suggesting strong constraints on the amino acid sequence in this region.

The carboxy-terminal region of the predicted sevenless protein shows a high degree of homology with all known protein tyrosine kinase domains but is most closely related to the chicken c-ros (47% homology, including conservative changes; Hafen et al. 1987) and human c-ros kinase domains (51% homology, Birchmeier et al. 1986). Comparison of the sevenless sequence with the available human c-ros sequence (C. Birchmeier and M. Wigler, unpubl.) shows that the first 550 amino acids of the extracellular domains of these proteins are 33% homologous, including conservative changes. The positions of cysteine residues and potential sites of amino-linked glycosylation do not appear to be tightly conserved in this region. Homology between the sevenless protein and the Drosophila insulin receptor [Nishida et al. 1986; Fernandez-Almonacid and Rosen 1987] was identified in the extracellular domain of the two proteins (28% amino acid homology, including conservative changes) but required the insertion of large gaps in the two sequences [see Materials and methods].

**Rescue of the sevenless phenotype by P-element transformation**

The sevenless phenotype has been rescued previously by P-element-mediated germ-line transformation with the 16.3-kb genomic EcoRI fragment [Hafen et al. 1987]. However, only one line was obtained; and although this appeared to rescue the sevenless phenotype morphologically it was not tested for rescue of R7 cell function, i.e.,
Figure 4. Identification of the transcriptional start sites of the sevenless gene by primer extension and RNase protection. Poly(A)+ RNA from wild-type adult heads, bodies, and larval imaginal discs was annealed separately to three nonoverlapping sevenless primers and to control rhodopsin and actin 5C primers and then extended with reverse transcriptase, as described in the Materials and methods. (A) Comparison of the extended products obtained using a sevenless primer, sevl, annealed to head and body RNA and a dideoxy sequencing reaction of a single-stranded genomic EcoRV (Fig. 2) template primed with sevl. Because the sequencing reaction was primed using the sevl primer, the resulting sequence was the reverse complement of that shown in Figure 2. Therefore, to facilitate comparison with sequence in Figure 2, the labeling of the sequencing reactions has been changed (A→T, G→C, C→G, and T→A). The sequence around the start sites (5' to 3') is TTCAGTT (P1), TGGTAAG (P2), and ATCAGTT (P3). The P1 and P3 sequences correspond closely to the consensus start site ATCAGTT of Hultmark et al. (1986). (B) Confirmation of the start sites using two additional sevenless primers and comparison of the transcription pattern in different tissues. Head, body, and imaginal disc RNA (20 μg) analyzed with sevenless primers sevl, sev2, and sev3. The bands seen with primer sev3 [B] correspond to within 2 nucleotides of each of the sites seen with sevl when the relative position of these two primers is taken into account (see Fig. 2). Similarly, the result obtained with sev2 confirms the position of the major start site. (Due to the lower specific activity of sev2, the other start sites were not detectable.) The pattern of transcription start sites is the same in each RNA sample. None of these bands was apparent when the reaction was performed with sevenless head RNA [data not shown] (C) Control primers Rh1 and actin used to extend 5 μg of the same RNA samples as in B demonstrate the lack of significant contamination of body with head RNA and that equal amounts of RNA are present in each sample. (D) Comparison of the products obtained by primer extension with sevl with the 32P-labeled RNA fragments protected from RNase digestion after annealing to head RNA (10 μg) or control tRNA. The position of the EcoRV fragment [used in the generation of the RNA probe] relative to sevl is such that if no splicing occurs between the start sites detected by primer extension and the 5' end of clone 10, the protected fragments should be 4 nucleotides smaller than the products of primer extension [see Fig. 2]. Protected fragments corresponding to each of the bands seen by primer extension are apparent [primer extension/RNase protection: 182/176, 118/110, 91/86 nucleotides]. The fragment corresponding to the P2 site is slightly smaller than expected, probably due to the AT-rich sequences immediately 3' to this site, allowing nipping of the ends of the hybrid. Sizes were estimated by comparison with labeled MspI restricted pBR322.
Nucleotide sequence of sevenless

-487 GATCATGTTTTTATTAAGATTATATTTTAGATAGATACTCTTGAG
-367 GGCATTGGTTTCTTGAGATGGAAAATACATCTATCTATCTATGTATCTATAAGAACTTAAATATCTGGCGGCC
-247 TGCTTGCGATCTGCGATCAACCGATTGAGTTGAGCTTTGACCCATG6AAAGAGGAGCGCGGTGGGCTATCTACCCATGATATCCAGTTTAGGGGTTGCCCTGCCAATGACGTCATAGTTG
-127 ATG(TATCyATC^ATCGCATCGCTTCGGTTCGGTTTCAGTTCTGCTCAAGTTCTCCAGCTGGGGGGTTGGACTTGGACTTGGAGCTTGGGGCTTGGAGAAAAGCCGCCGCTCGCC
-7 G6CAAATCAGTTCTGCTGCGATCGCGCCACTGCGCACTACGTAACGCTTCGGGAATCGCACATCCAGATCCAGATTCAGATCCACATCCAGATATCTGTGACATTTTGTCGAGTGAGTTG
113 CGTTCTCACAAGTACAAGATACATCCGTGCGCGTATGTGTGTGTTTGCATGAATTTCGGCTCTTTTAATACCCTTTTTTTTCGCGTGCACTAAGTGTTCCTATGTGTGTGTGTGGAG
233 TTGCTTTTCAGTTGGGGCCGCAACTGCAACAATAAAATGCAATTAAGCCGCTCAAATCAACAATAAATTGAATTAAAGCGCGTGAACAGCGCTTAACAACTTCAACAGCGGCGATTCGAG

Figure 5. Identification of repeated sequences clustered in the promoter region of sevenless. Repeated sequences surrounding the start sites are indicated by arrows. Dots above or below a repeat indicate bases that differ from the basic repeat motif. The repeats can be depicted in several ways. For example, the repeat ATCCAG (beginning at nucleotide 54) is flanked at either end by the larger repeat CACATCCAG. Numbering is relative to the major (P3) transcription start site. Minor start sites (P1 and P2) are also indicated.

the behavioral response of adult flies to ultraviolet light. (The R7 cell serves as the major receptor for ultraviolet light.) A complex temporal and spatial pattern of sevenless protein expression has been demonstrated in eye imaginal discs [Banerjee et al. 1987b; Tomlinson et al. 1987]. The 16.3-kb genomic fragment includes only about 950 bp of the promoter region of the gene, raising the question as to whether this region is sufficient for normal expression of the sevenless gene. Therefore, we generated eight transformant lines using this fragment and assayed sevenless function by three criteria. First, the presence or absence of the R7 cell was judged by examining the pseudopupils of transformant flies. (The pseudopupil image is formed by transmission of light through the rhabdomeres of the individual photoreceptors; see Materials and methods). Second, R7 cell function and neuronal connection in the optic lobes was assayed in a behavioral test involving the response of adult flies to ultraviolet light. All eight lines had apparently normal pseudopupils, and their behavioral preference for ultraviolet light was indistinguishable from that of wild-type flies. Finally, the pattern of sevenless protein expression in these transformant lines was examined by staining eye imaginal discs with an antibody to

Figure 6. Hydrophobicity analysis of the predicted sevenless amino acid sequence. The plot was calculated according to Kyte and Doolittle (1982), with a window length of 21 amino acids. Positive values indicate more hydrophobic regions and include a putative amino-terminal anchor sequence [N] and more carboxy-terminal transmembrane domain [C]. Both sequences have a high probability of spanning the membrane [see text].
the sevenless protein (Tomlinson et al. 1987) and comparing the pattern to control wild-type flies. The pattern of sevenless expression in the transformant lines tested was the same as that in wild-type larvae (Fig. 7a,b). These experiments suggest that the 16.3-kb EcoRI genomic fragment encompasses the complete regulatory and transcriptional regions of the gene.

In addition, we constructed two minigenes in which some of the introns present in the genomic fragment were removed [see Materials and methods]. Transformation with the hybrid 1 construct, which lacks the first eight introns of the genomic fragment, failed to rescue the sevenless phenotype in any of the four lines tested by examination of pseudopupils. A second minigene construct was made [hybrid 2] in which the first intron (~1.4 kb) was retained but introns 2–7 were absent (Fig. 7d). Of 13 lines generated, 1 line had a normal pseudopupil and showed a wild-type response to ultraviolet light. The remaining lines were mutant in both their response to ultraviolet light and the morphology of their pseudopupils. The eye discs of several of these lines were stained with the sevenless antibody. The staining pattern obtained with the hybrid 2 line that had a normal pseudopupil was considerably weaker than with transformants of the wild-type gene (Fig. 7c), but the overall pattern of expression was similar: a sharp onset of sevenless protein expression at the morphogenetic furrow and the presence of the characteristic progression of stained cells, as described previously (Tomlinson et al. 1987). In contrast, sevenless protein expression in the

Figure 7. Light micrographs of eye imaginal discs from third instar larva stained with a sevenless antiserum. Anterior is to the bottom; magnification, 600×. (a) Whole-mount eye disc from a wild-type (Canton S), larva. (b) Eye disc from a larva transformed with the 16.3-kb genomic fragment containing the sevenless gene, in a sevenless^dd background. The staining pattern obtained with this and other lines obtained by transformation with this fragment is indistinguishable from the pattern seen in eye discs from Canton S larvae. In this example, pattern formation, which commences at the morphogenetic furrow, is further advanced than that in the Canton S disc shown in a. sevenless protein expression was undetectable in control sevenless^-^- larvae [data not shown], as described previously [Banerjee et al. 1987b]. (c) Whole-mount eye disc from a line transformed with a sevenless gene construct from which part of the genomic sequence has been replaced by cDNA sequences (hybrid 2). This line was sevenless^- as judged by the morphology of the pseudopupil and phototaxis to ultraviolet light. sevenless protein expression was clearly apparent in the eye discs of larvae of this line but was weaker than in larvae transformed with the complete genomic fragment. (d) Schematic representation of the cDNA/genomic constructs. Hybrid 1 lacks introns 1–7, and hybrid 2 lacks introns 2–7, relative to the 16.3-kb genomic fragment.
other hybrid 2 lines tested was barely detectable. The low level of protein expression obtained with the hybrid gene transformants suggests that poor expression, rather than a requirement for additional transcripts, resulted in the inconsistent rescue obtained with the cDNA constructs. The rescue obtained with the one line shown here presumably resulted from integration of the construct into a site more favorable for its expression. Reduced expression of the cDNA constructs relative to the complete genomic fragment may be due to the presence of sequences in the introns required for correct expression of the gene or incorrect processing of the primary transcript due to the alteration of the exon/intron structure.

Conclusion

We have described the structure of the sevenless gene and its predicted protein product. The promoter region of the sevenless gene, mapped both by primer extension and RNase protection, contains one major and two minor transcription start sites that are surrounded by a large number of repeated sequences. A number of lines were generated by P-element transformation with a 16.3-kb genomic fragment that included only about 950 bp of the sequence, 5' to the transcriptional start sites. Transformation with this fragment consistently rescued sevenless gene function, as determined by morphological and behavioral criteria, and was sufficient for the complex pattern of sevenless protein expression in eye imaginal discs. A transcription unit of 12 exons, which spans the majority of the genomic fragment, was identified. Several lines of evidence suggest that this is the major, if not the only, sevenless transcript. RNA blot analysis showed a single transcript of 8.6 kb expressed in the adult head, body, and imaginal discs. The size of this transcript is in good agreement with the size of the composite cDNA sequence obtained from a large number of cDNAs. By sequencing all of the cDNA fragments, even small differences, such as that occurring in the insulin receptor (cited Ullrich et al. 1986), should be detectable, if they represent a common form of the sevenless transcript. Finally, rescue of the sevenless phenotype was obtained with a construct in which all of the introns corresponding to the putative extracellular domain of the protein were deleted (hybrid 2). Although this was obtained in only 1 of 13 lines tested, analysis of these lines suggested that a low level of protein expression, rather than the absence of additional gene products, is responsible for the inconsistent rescue of the sevenless phenotype obtained with this minigene construct.

The putative sevenless protein differs from most other tyrosine kinase cell-surface receptors in the unusually large size of its extracellular domain (~2000 amino acids) and the possible presence of two transmembrane domains in the mature protein. These structural features, as well as the limited pattern of sevenless expression, may relate to its function as a receptor for local positional information. The precise nature of this relationship awaits further study.

Materials and methods

Nucleic acids

DNA fragments. The genomic sequence of sevenless was determined from a 16.3-kb EcoRI fragment isolated from a Canton S genomic library (Hafen et al. 1987). Probes generated from the middle, 5', and 3' end of this fragment were used to screen a cDNA library prepared from RNA isolated from Oregon R third instar larval eye imaginal discs (A. Cowman, unpublished). Eleven independent cDNA fragments were isolated from approximately 7 × 10^6 independent recombinant plaques. Construction of minigene lacking a subset of the introns present in the 16.3-kb genomic fragment was performed using standard methods. Hybrid 1 was constructed by replacing genomic sequences between a SspI site in exon 1 and an Sphi site in exon 8, with the corresponding cDNA fragment from cDNA clone 10, thereby removing the first 7 introns but retaining the genomic sequences flanking exons 1 and 8 (Fig. 7d). Hybrid 2 was constructed by replacing genomic sequences between M11 sites in exons 2 and 8 with the corresponding cDNA sequences from clone 10.

RNA isolation. RNA for blot analysis, primer extension, and RNase protection studies was isolated using guanidine hydrochloride. Briefly, RNA from adult Canton S or sevenless^7^ (Bancroft et al. 1987a) flies was isolated by first separating heads from bodies, as described previously (Oliver and Philips 1970), and then grinding them under liquid nitrogen in a mortar and pestle. The powdered tissue was dispersed in 6 M guanidine HCl and 0.1 M sodium acetate (pH 5.5), clarified twice at 15,000 g, and the supernatant was layered over a pad of 5.2 M cesium chloride and 10 mM EDTA. The RNA was pelleted by centrifugation at 100,000 g for 18 hr at 20°C, and poly[A] + RNA was isolated by oligo(dT) chromatography. RNA was similarly obtained from mass-isolated Oregon R third instar imaginal discs (Eugene et al. 1979), dispersed directly in guanidine HCl without freezing or grinding.

Mapping of transcripts

Primer extension. To map the initiation of transcription, primer extension was performed using three nonoverlapping oligonucleotides (sevl, sev2, and sev3), corresponding to sequences adjacent to the end of the cDNA that extended to the farthest 5' (see Fig. 2). Primers were labeled to high specific activity, using crude [γ-32P]ATP (7000 Ci/m mole) and T4 polynucleotide kinase, and then separated from unincorporated nucleotides by ammonium acetate precipitation. Labeled primers were annealed to 10-20 μg of poly[A] + RNA at 60°C for 90 min, extended at 37°C using AMV reverse transcriptase, and the products analyzed on 6% polyacrylamide/7.6 M urea gels, as described previously (McKnight and Kingsbury 1982). Primers complementary to the PI promoter of cytoskeletal actin (Bond and Davidson 1986) and the major rhodopsin, Rh1 (Zuker et al. 1985), annealed at 60°C to 5 μg of poly[A] + RNA and extended as above, were used as controls for the amount and quality of each RNA sample and for the degree of head contamination of body RNA, respectively.

RNase protection. RNase protection was performed using probes uniformly labeled with [32P]UTP (800 Ci/m mole), generated from a template plasmid that bore a 372-bp genomic EcoRV fragment (nucleotides -289 to +86; Fig. 2) cloned into the plasmid pSP64. Approximately 10^6 cpm were hybridized at 37°C for 18 hr to 10 μg of poly[A] + head RNA or yeast tRNA and then digested for 15 min at 37°C with RNase T1 and RNase.
DNA sequencing and analysis

All DNA sequencing was performed using the chain termina-


sequences were done using the Repeat, Compare, and Dotplot


P-element-mediated transformation

The 16.3-kb genomic fragment or the hybrid 2 construct was


Assessment of sevenless function

Pseudopupils. Rescue of the sevenless mutant phenotype can


Behavioral test. Sevenless function was examined in a behav-


Acknowledgments

We would like to thank Matthew Michael for technical assis-


362 GENES & DEVELOPMENT
imaginal discs. We would also like to thank our colleagues Ulrike Heberlein, Janice Fischer, Dan Kalderon, and Andrew Tomlinson for advice on some of the molecular and antibody techniques used and members of this laboratory for critical reading of the manuscript. D.D.L.B. is a C.J. Martin fellow, and M.A.S. is a Damon Runyon–Walter Winchell fellow (grant DRG-926).

Note
Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00218.

References


Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement
of the dideoxy chain termination method of cDNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Res.* **14**: 1319–1324.


Nucleotide sequence and structure of the sevenless gene of Drosophila melanogaster.

D D Bowtell, M A Simon and G M Rubin

*Genes Dev.* 1988, 2:
Access the most recent version at doi:10.1101/gad.2.6.620

This article cites 54 articles, 15 of which can be accessed free at:
http://genesdev.cshlp.org/content/2/6/620.full.html#ref-list-1

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here.](#)