Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 5351–5353, July 1990
Genetics

Comparison of the sevenless genes of *Drosophila virilis* and *Drosophila melanogaster*

(tyrosine kinase/evolution/retinal development/receptor)

W. MATTHEW MICHAEL, DAVID D. L. BOWTELL, AND GERALD M. RUBIN

Howard Hughes Medical Institute and Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720

Contributed by Gerald M. Rubin, April 18, 1990

ABSTRACT **The sevenless gene of *Drosophila melanogaster* encodes a transmembrane tyrosine kinase receptor required for normal eye development. We report here the isolation and DNA sequence analysis of the sevenless gene from *Drosophila virilis*. The predicted amino acid sequences of the sevenless proteins from these two species, which diverged ≈ 60 million years ago, are compared.**

The sevenless (*sev*) gene of *Drosophila melanogaster* is required for proper development of a single cell type in the developing retina, the R7 photoreceptor. In the absence of the gene or its product the R7 photoreceptor progenitor fails to differentiate normally and becomes, instead, a nonneuronal cell (1). The *sev* gene has been cloned, and its DNA sequence has been determined (2–4). Conceptual translation of the nucleotide sequence (2–4) and biochemical analyses of the sev protein (5) indicate that *sev* encodes a 288-kDa transmembrane protein with a tyrosine kinase domain (2–4). These results, together with the phenotype of sevenless mutants, suggest that the sev protein functions as a receptor for an extracellular signal required to instruct a cell to differentiate into an R7 photoreceptor.

The extracellular domain of the sev protein is much larger than that of most tyrosine kinase receptors. Although much can be inferred about the relationship between structure and function in the kinase domain of sev protein from work done on other tyrosine kinases (6, 7), we have no basis for making predictions about the functional organization of its large extracellular domain. As a first step in determining what features of the primary sequence of the sev protein are likely to be functionally important, we have isolated and determined the DNA sequence of the coding region of the *sev* homologue in the species *Drosophila virilis*.* Because these species diverged approximately 60 million years ago (8), sequence conservation should exist only in functionally relevant regions of the protein. We present an amino acid sequence comparison and discuss how the rates of divergence within the protein vary in relation to the presumed functional domains.

METHODS

A *D. virilis Sau*3A genomic library constructed in EMBL4 (gift of M. Scott; University of Colorado, Boulder, CO) was screened using a probe made from the *D. melanogaster* cDNA clone cED3.1 (3). Twelve clones were recovered and one, λ 11, was restriction mapped and partially sequenced. Because we were unable to isolate clones extending 5′ of λ 11 from this library, two genomic libraries were constructed in λ Dash (Stratagene)—one containing *Eco*RI and the other containing *Bam*HI fragments. Screening of these libraries yielded clones that extended the cloned region further 5′. To complete the cloning a *Mbo* I genomic library constructed in EMBL3 (gift of R. Blackman, Harvard University, Cambridge, MA) was screened, and a clone, λ 3.1, was recovered, which hybridized to sequences from the 5′ end of the *D. melanogaster* coding region.

DNA sequence determinations were done as described (9). DNA sequences were analyzed by using the GEL, ALIGN, and FASTDB programs of IntelliGenetics.

RESULTS AND DISCUSSION

The *D. virilis* homologue of the *D. melanogaster* sevenless gene was isolated by screening a *D. virilis* genomic library at reduced hybridization stringency with a *D. melanogaster* sevenless cDNA probe. The sequence of ≈19 kilobases (kb) of *D. virilis* genomic DNA containing the *D. virilis sev* homologue was determined, and the intron–exon structure of the gene was deduced by comparison with the *D. melanogaster* sequence (Fig. 1). Fig. 2 displays the conceptual translation of the *D. virilis* coding sequence aligned with that of the *D. melanogaster* gene. The *D. virilis sev* gene encodes a putative protein of 2595 amino acids compared with the 2552 amino acids for the *D. melanogaster* protein (3, 4).

In the *D. melanogaster* protein are three closely spaced in-frame methionine codons near the beginning of the long open reading frame (3, 4); it has not been established which of these codons functions as the initiating codon for translation *in vivo*. The first of these three codons is not conserved in *D. virilis*, making it unlikely that this first codon is the initiation codon. The third methionine codon (*D. melanogaster* residue 43 in Fig. 2) is in a context that fits the consensus for translation initiation codons (10) better than the second methionine codon (residue 1 in Fig. 2). Moreover, transcriptional gene fusions made to the sevenless genomic sequence between the second and third methionine codons can produce functional sevenless protein *in vivo* (11). However, there is considerable amino acid conservation between the second and third methionine codons of these proteins, suggesting that the second methionine codon (residue 1 in Fig. 2) is used to initiate translation.

The overall amino acid identity between the two proteins is 63%, and the similarity is 70% when conservative substitutions are included. These figures are lower than those of several other published comparisons of *D. melanogaster* and *D. virilis* protein sequences (12–15). A possible reason for this lower degree of identity is the large size of the extracellular domain relative to the number of residues that might be functionally important, leaving a significant fraction of the sequence under little selective pressure and, therefore, free to diverge. In contrast, most protein sequences compared between *melanogaster* and *virilis* have been of soluble globular proteins.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M34543, M34544, and M34545).

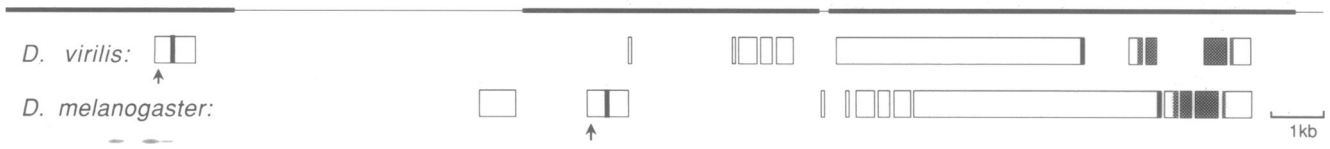

FIG. 1. Structure of *D. virilis* sevenless gene. Horizontal line at top represents genomic *D. virilis* DNA; the thick line represents the DNA regions sequenced. Below the line, positions of exons 2–12 of the *virilis* gene are indicated. Exon 1 lies beyond the left end of the region shown, but its precise position was not determined. For comparison, exons 1–12 of *D. melanogaster* sevenless gene (3, 4) are diagramed to the same scale and arbitrarily aligned with *D. virilis* gene at the end of exon 12. In general, exons in both species are the same size, and the intron–exon boundaries are similar; however, introns tend to be larger in *D. virilis*. The two potential transmembrane domains in each sequence are shown as solid black bars, kinase domains are stippled, and arrows mark the start of the long translational open reading frame.

FIG. 2. (*Figure continues on the opposite page.*)

Fig. 2. Comparison of the deduced amino acid sequences of the *D. virilis* and *D. melanogaster* sevenless proteins. The amino acid sequences (in one-letter code) deduced from the *D. melanogaster* sevenless genomic and cDNA sequences (3, 4) and *D. virilis* genomic sevenless sequence are aligned. Amino acid 1 corresponds to the first methionine in *D. virilis* open reading frame and the second methionine in *D. melanogaster* open reading frame. Positions of exon boundaries are indicated above the *D. melanogaster* sequence and below the *D. virilis* sequence, and the exons are numbered. Position of the kinase domain and the two putative membrane spanning domains are indicated by underlining. It is unclear whether the putative membrane-spanning region near the N terminus is present as a transmembrane domain in the native protein or serves as a cleaved signal sequence (5). Sequences that match the consensus for amino-linked glycosylation (N-X-S/T) are indicated by shading, as are the conserved arginine residues (positions 2036–2044 of *D. melanogaster* sequence) that may serve as an endopeptidase cleavage site. The 42 cysteines residues conserved in both sequences are indicated by dots.

Fig. 3 describes the amino acid conservation in the various subdomains of the protein. In the *D. melanogaster* protein 17 potential amino-linked glycosylation sites (Asn-Xaa-Ser/Thr) exist in the putative extracellular domain, and all are conserved in *D. virilis*. The rate of cysteine residue conservation is also quite high; 42 of the 43 of the cysteines residues found in the *melanogaster* sequence are conserved in *virilis*. However, the *virilis* protein has more cysteines (52 vs. 43) than the *melanogaster* protein. The kinase domains of these proteins are 83% identical, and the transmembrane domain separating the kinase and extracellular domains is completely conserved. In contrast, the extracellular domains themselves are only 60% identical. The second hydrophobic region, which is located near the N terminus of the protein and functions either as a cleaved signal sequence or a second transmembrane domain (3–5), is 86% identical between the two proteins. The sevenless protein is thought to be cleaved into α and β subunits at an endopeptidase cleavage site consisting of nine consecutive arginine residues (5; position 2036–2044 of the *melanogaster* sequence). Seven of these arginines are conserved in the *virilis* protein.

When the sequence of the extracellular domains of either the *D. melanogaster* or *D. virilis* sevenless proteins is used to search protein databases, similarities to a number of other proteins are found. The extent of these similarities indicates them to be of borderline significance (20–25% identity; see ref. 16 for discussion of evaluating sequence similarity). Knowledge of the sevenless sequence from two species proved extremely valuable in assessing the relevance of these similarities; similarities to different proteins were seen when the *virilis* and *melanogaster* extracellular domains were used to search the databases, strongly suggesting that the observed similarities are not biologically significant. Several short regions of DNA sequence similarities were observed within the intron separating exons 3 and 4 of the *virilis* and *melanogaster* genes (data not shown). This region has been shown as important for regulation of *D. melanogaster* sevenless gene (17, 18), and these conserved elements may serve as sites for DNA-binding proteins that regulate sevenless transcription.

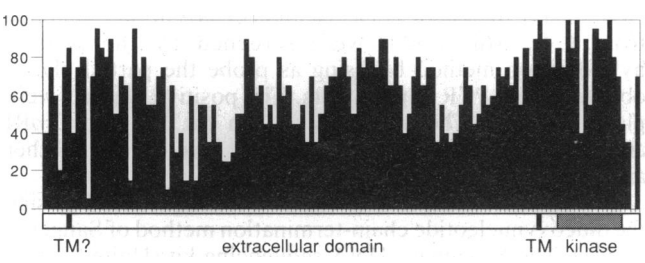

FIG. 3. Percentage sequence identity as function of protein position. *D. melanogaster* sequence was divided into 20 amino acid blocks, and the percentage identity to *D. virilis* sequence within each block, in the alignment of Fig. 2, is plotted in a histogram. A diagram of the domain structure of sevenless protein is shown for reference. TM, transmembrane.

Note Added in Proof. We have previously noted similarity between the sevenless protein and those portions of the ros protein the sequence of which was available for comparison (2, 3). Two recent reports describe the complete sequence of the human (19) and rat (20) *ros1* gene and demonstrate that the similarity between sev and ros extends throughout these proteins.

1. Tomlinson, A. & Ready, D. F. (1987) *Dev. Biol.* **123,** 264–275.
2. Hafen, E., Basler, K., Edstroem, J. E. & Rubin, G. M. (1987) *Science* **236,** 55–63.
3. Bowtell, D. D. L., Simon, M. A. & Rubin, G. M. (1988) *Genes Dev.* **2,** 620–634.
4. Basler, K. & Hafen, E. (1988) *Cell* **54,** 299–311.
5. Simon, M. A., Bowtell, D. D. L. & Rubin, G. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8333–8337.
6. Hunter, T. & Cooper, J. A. (1985) *Annu. Rev. Biochem.* **54,** 897–930.
7. Yarden, Y. & Ullrick, A. (1988) *Annu. Rev. Biochem.* **57,** 443–478.
8. Beverly, S. M. & Wilson, A. C. (1984) *J. Mol. Evol.* **21,** 1–13.
9. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
10. Hultmark, D., Klemenz, R. & Gehring, W. J. (1986) *Cell* **44,** 429–438.
11. Bowtell, D. D. L., Simon, M. A. & Rubin, G. M. (1989) *Cell* **56,** 931–936.
12. Kassis, J. A., Poole, S. J., Wright, D. K. & O'Farrell, P. H. (1986) *EMBO J.* **5,** 3583–3589.
13. Treier, M., Pfeifle, C. & Tautz, D. (1989) *EMBO J.* **8,** 1517–1525.
14. Colot, H. V., Hall, J. C. & Rosbash, M. (1988) *EMBO J.* **7,** 3929–3937.
15. Wilde, C. D. & Akam, M. (1987) *EMBO J.* **6,** 1393–1401.
16. Doolittle, R. (1986) *Of URFs and ORFs* (University Science Books, Mill Valley, CA), pp. 3–48.
17. Bowtell, D. D. L., Kimmel, B. E., Simon, M. A. & Rubin, G. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6245–6249.
18. Basler, K., Siegrist, P. & Hafen, E. (1989) *EMBO J.* **8,** 2381–2386.
19. Birchmeier, C., O'Neill, K., Riggs, M. & Wigler, M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4799–4803.
20. Matsushime, H. & Shibuya, M. (1990) *J. Virol.* **64,** 2117–2125.