

Genomic Organization and Deduced Amino Acid Sequence of a Putative Sodium Channel Gene in *Drosophila*

LAWRENCE SALKOFF,* ALICE BUTLER, AGUAN WEI, NANCY SCAVARDA, KELLEEN GIFFEN, CATHERINE IFUNE, RICHARD GOODMAN, GAIL MANDEL

The deduced amino acid sequence of a *Drosophila* gene isolated with a vertebrate sodium channel complementary DNA probe revealed an organization virtually identical to the vertebrate sodium channel protein; four homologous domains containing all putative membrane-spanning regions are repeated in tandem with connecting linkers of various sizes. All areas of the protein presumed to be critical for channel function show high evolutionary conservation. These include those proposed to function in voltage-sensitive gating, inactivation, and ion selectivity. All 24 putative gating charges of the vertebrate protein are in identical positions in the *Drosophila* gene. Ten introns interrupt the coding regions of the four homology units; introns with positions conserved among homology units bracket a region hypothesized to be the selectivity filter for the channel. The *Drosophila* gene maps to the right arm of the second chromosome in region 60D-E. This position does not coincide with any known mutations that confer behavioral phenotypes, but is close to the *seizure* locus (60A-B), which has been hypothesized to code for a voltage-sensitive sodium channel.

THE STRUCTURE OF THE VOLTAGE-GATED PORE THAT MAKES up the *Electrophorus* (eel) electroplax sodium channel is a large single alpha subunit (molecular size, 250 kilodaltons). The primary sequence of the sodium channel protein has been deduced from the complementary DNA (cDNA) sequence (1). There appear to be three similar genes expressed in rat brain (2). When messenger RNA (mRNA) transcribed from a full-length cDNA encoding one of the brain channels (type II) was injected into *Xenopus* oocytes a normal voltage-activated sodium current was produced similar to that described in neurons (3-5). This result suggests that the alpha subunit is itself capable of insertion into the membrane and voltage-gated ion conductance even though other accessory peptides may be associated with the large peptide in vivo (6).

The structures of the alpha subunits in eel and rat are similar and seem to have evolved from a smaller gene that underwent two rounds of intragenic duplication to produce the modern structure containing four homologous domains. These homologous domains are referred to as homology domains a to d. Since it is likely that ion channel evolution occurred before the separation of vertebrate and invertebrate species (7), we assume that a similar channel protein exists in invertebrate species. We therefore used the eel cDNA as a probe to screen a *Drosophila* genomic library for related sequences.

We now describe the structure and sequence of a *Drosophila* gene that is homologous to the eel and rat sodium channel alpha subunits. Like the vertebrate channel, the four homologous domains of the deduced *Drosophila* channel polypeptide are about 300 amino acids each and are connected to each other by dissimilar linker regions of different lengths. The hydrophilic linker segments are thought to be cytoplasmic while the four homologous domains are hypothesized to contain all of the portions of the channel residing within the membrane (2, 8, 9). Conservation of the homologous domains is very high while conservation of two of the three cytoplasmic linkers is very low. The short, highly conserved linker between homology domains c and d, as discussed below, may function as the inactivation gate of the channel (9). Each of the homologous domains has several hydrophobic or amphipathic strings of amino acids that may be membrane-spanning alpha helices or other intramembrane structures (1, 2, 8, 9). Noda *et al.* (2) have suggested that there may be six membrane-spanning structures (S1 to S6) in each homology domain and their nomenclature is used to identify these segments. Guy and Seetharamulu (8) have proposed that there are two additional intramembrane segments between S5 and S6. Sequences that distinguish each of the four homologous domains from each other and are absolutely conserved between the fly and vertebrate proteins show that the channel protein evolved more than 600 million years ago, before the separation of vertebrate and invertebrate species. This was suggested previously on the basis of biophysical data (7).

Structure of the *Drosophila* channel gene. We isolated the *Drosophila* gene by screening a *Drosophila* genomic library at low stringency with a cDNA probe encoding the eel sodium channel (legend to Fig. 1). All four of the homology domains (a to d) were sequenced from genomic DNA while two of the four (c and d) were also sequenced from cDNA (Fig. 1, 3-1A1) for analysis of intron-exon junctions.

A method of determining the presence of highly conserved DNA sequences was used for both the initial screening of positive clones as well as the mapping of exons within the genomic DNA of the *Drosophila* channel gene (Fig. 1) (10).

The method for eliminating false positive clones was to determine the degree of interspecies DNA conservation in each clone. Because it was likely that the sodium channel gene would be conserved much

L. Salkoff, A. Butler, A. Wei, N. Scavarda, K. Giffen, and C. Ifune are in the Department of Anatomy and Neurobiology, Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, MO 63110. L. Salkoff is jointly in the Department of Genetics at Washington University School of Medicine. R. Goodman and G. Mandel are at the Tufts New England Medical Center Department of Medicine, 171 Harrison Avenue, Boston, MA 02114.

*To whom correspondence should be addressed.

more highly than most DNA found in a randomly selected clone, the rapid identification of clones containing large stretches of highly conserved DNA allowed us to focus attention on the correct clones. Each phage clone insert was nick-translated and used to probe DNA from *Drosophila mercatorum*, a species that is distant from *D. melanogaster*. The stringency of hybridization was adjusted so that unselected DNA from *D. melanogaster* when used as a probe against whole genomic DNA from *D. mercatorum* would give no hybridization signal above background. At the same level of stringency, however, the coding region for a highly conserved gene, such as the sodium channel, gave a prominent signal against *D. mercatorum* DNA. The actual conditions of hybridization when optimized for this purpose were the same as those given above for the initial screen.

The restriction map of the *Drosophila* gene is shown in Fig. 1A and the positions of all of the exons constituting the four homology domains are indicated. The positions of the introns with regard to the membrane-spanning regions are summarized in Fig. 1B and

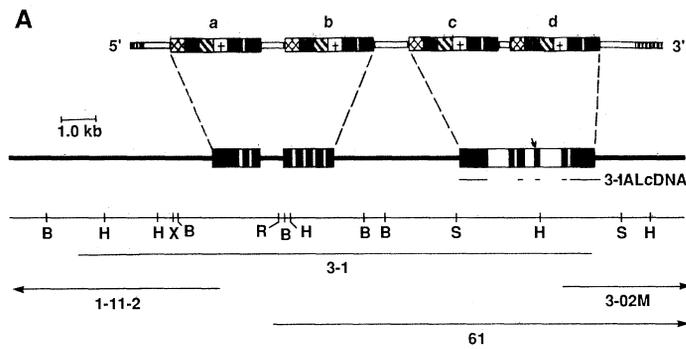


Fig. 1. (A) Genomic map showing all exons (filled boxes) of the four homologous domains (a, b, c, and d) of the *Drosophila* channel gene. Introns are shown in open boxes. Uncharacterized regions are shown by the solid horizontal bar. The corresponding mRNA showing each of the homologous regions subdivided into six membrane-spanning segments is presented at the top. The single exon in the conserved linker c-d is indicated by an arrow. Exons of the linker regions a-b and b-c are not indicated. Lambda charon 4A clones shown below the restriction map are from a genomic library. Restriction enzyme recognition sites for Bam HI (B), Eco RI (R), Hind III (H), Sal I (S), and Xho I (X) are shown. The cDNA 3-1AL lacks the second exon of homology domain c. The deletion occurred at the intron-exon splice junctions and is presumed to be the result of a splicing error of native origin because the reading frame is shifted. Two possible translational start sites have been located upstream from homology domain a but the intron-exon organization of this upstream region has not yet been fully determined. Below the restriction map several overlapping genomic clones are indicated (3-1, 1-11-2, 3-02M, 61). An eel sodium channel cDNA probe was used to screen a *Drosophila* genomic library provided by T. Maniatis. The eel cDNA included the region from S4d to the 3' untranslated terminus of the mRNA. Hybridization was in sodium chloride-sodium phosphate-EDTA solution ($5\times$ SSPE with 30 percent formamide at 42°C). Final washing conditions were $1\times$ SSC (standard saline citrate) at 50°C . In the process of isolating the *Drosophila* channel gene homologous to the eel sodium channel, numerous false positive clones had to be eliminated. For sequencing, m13 single-strand cloning vectors mp18 and mp19 were used and overlapping deletions in single-stranded DNA were constructed with the single-strand exonuclease activity of T4 polymerase (IBI Cyclone kit). Alternatively, synthetic primers were used. **(B)** Summary of intron positions in all four homology domains, a, b, c, and d. The six putative membrane-spanning units (S1 to S6) are indicated. The blank box between S5 and S6 indicates the region which is hypothesized to contain the selectivity filter for the channel (8, 9). The sizes and positions of all introns relative to this region and the putative membrane-spanning units are indicated.

their exact positions are shown in the deduced amino acid sequence in later figures. All introns follow the GT-AG rule for donor and acceptor splice junctions (11). Ten introns interrupt the coding regions of the four homology domains. Homology domain b has four introns while a, c, and d are each interrupted by two introns. Introns account for about half of the genomic sequence of the four homology regions.

The positions of two introns that flank both ends of the area separating S5 and S6 are present in homology domains a and b and at one of these sites in c. In a computer-generated molecular model of the vertebrate sodium channel (8, 9), the region between these introns contains the selectivity filter for the channel. Most of the residues constituting the putative selectivity filter, which include negatively charged residues lining the channel pore are conserved in the *Drosophila* gene in all four of the homology domains.

None of the other five introns are at similar positions among the four homology domains nor do the introns clearly demarcate domains. Introns are not found in the putative membrane-spanning regions S3, S5, and S6, which are very hydrophobic and very highly conserved regions. However, introns are found within S1 and S2 in homology domain d. These regions are amphipathic and less well conserved than other putative membrane-spanning regions. The most puzzling site of an intron is in S4b, in the center of a highly conserved string of gating charges.

There is one exon between homology domains c and d (arrow, Fig. 1A). Although much of the nucleotide sequence of the genomic DNA separating homology domains a and b and b and c was

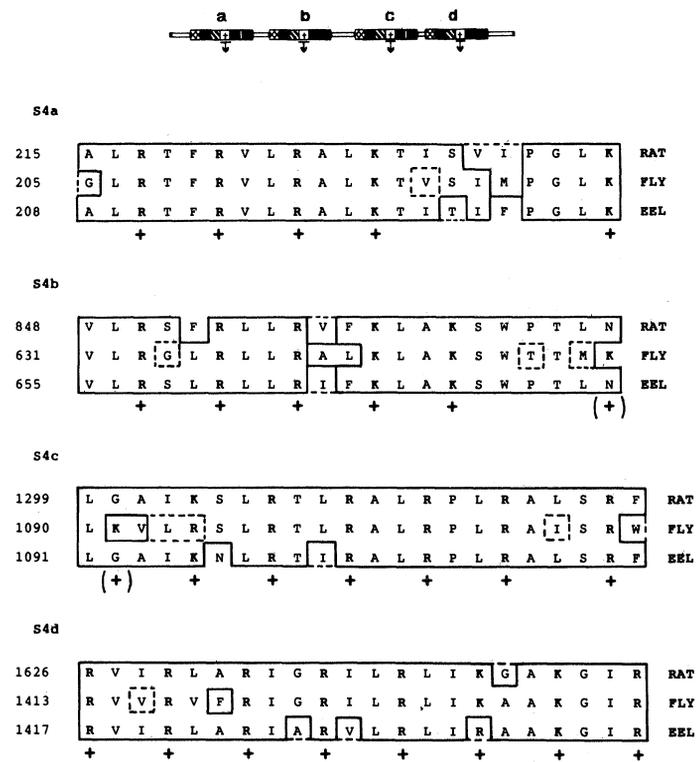


Fig. 2. Deduced amino acid sequence of S4 region of homology domains a, b, c, and d. Positive residues presumed to be gating charges are labeled with a "+." A parenthesis indicates a charge present only in the *Drosophila* sequence. Middle *Drosophila* (fly) sequence is compared with rat (top) and eel (bottom). Solid lines enclose identical residues; dotted lines enclose conservative substitutions. Conservative amino acid substitutions are defined as pairs of residues belonging to one of the following groups: S,T,P,A,G; N,D,E,Q; H,R,K; M,I,L,V; F,Y,W. Single letter abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

determined, the exons constituting the nonconserved cytoplasmic regions in these areas were not fully determined.

The deduced amino acid sequence of different regions of the *Drosophila* gene and its homology to vertebrate channel genes from both rat and eel are shown in Figs. 2 and 3. In general, analogous regions in the four homology domains show a similar degree of interspecies conservation. As with the vertebrate homology do-

main, b and d are slightly larger than a and c because of extra residues separating S5 and S6.

Conservation of presumed gating charges. The voltage-dependent activation of the sodium channel requires the outward movement of protein-bound positive charges across the plane of the membrane. The requirement for large charge movement within the channel protein has strongly focused attention on four strings of

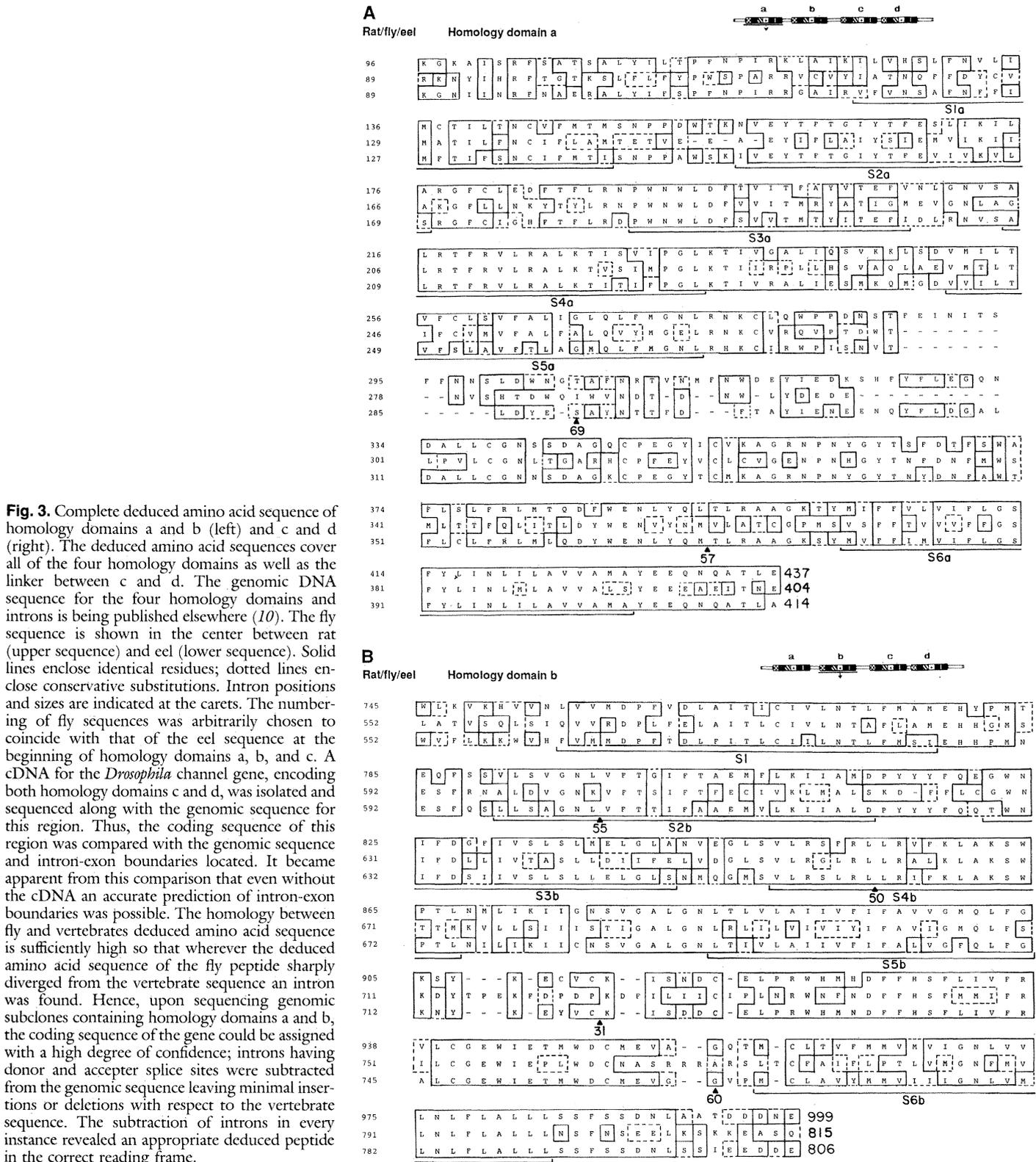


Fig. 3. Complete deduced amino acid sequence of homology domains a and b (left) and c and d (right). The deduced amino acid sequences cover all of the four homology domains as well as the linker between c and d. The genomic DNA sequence for the four homology domains and introns is being published elsewhere (10). The fly sequence is shown in the center between rat (upper sequence) and eel (lower sequence). Solid lines enclose identical residues; dotted lines enclose conservative substitutions. Intron positions and sizes are indicated at the carets. The numbering of fly sequences was arbitrarily chosen to coincide with that of the eel sequence at the beginning of homology domains a, b, and c. A cDNA for the *Drosophila* channel gene, encoding both homology domains c and d, was isolated and sequenced along with the genomic sequence for this region. Thus, the coding sequence of this region was compared with the genomic sequence and intron-exon boundaries located. It became apparent from this comparison that even without the cDNA an accurate prediction of intron-exon boundaries was possible. The homology between fly and vertebrates deduced amino acid sequence is sufficiently high so that wherever the deduced amino acid sequence of the fly peptide sharply diverged from the vertebrate sequence an intron was found. Hence, upon sequencing genomic subclones containing homology domains a and b, the coding sequence of the gene could be assigned with a high degree of confidence; introns having donor and acceptor splice sites were subtracted from the genomic sequence leaving minimal insertions or deletions with respect to the vertebrate sequence. The subtraction of introns in every instance revealed an appropriate deduced peptide in the correct reading frame.

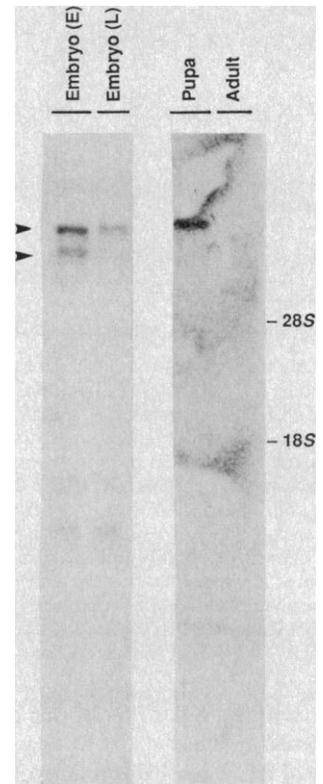
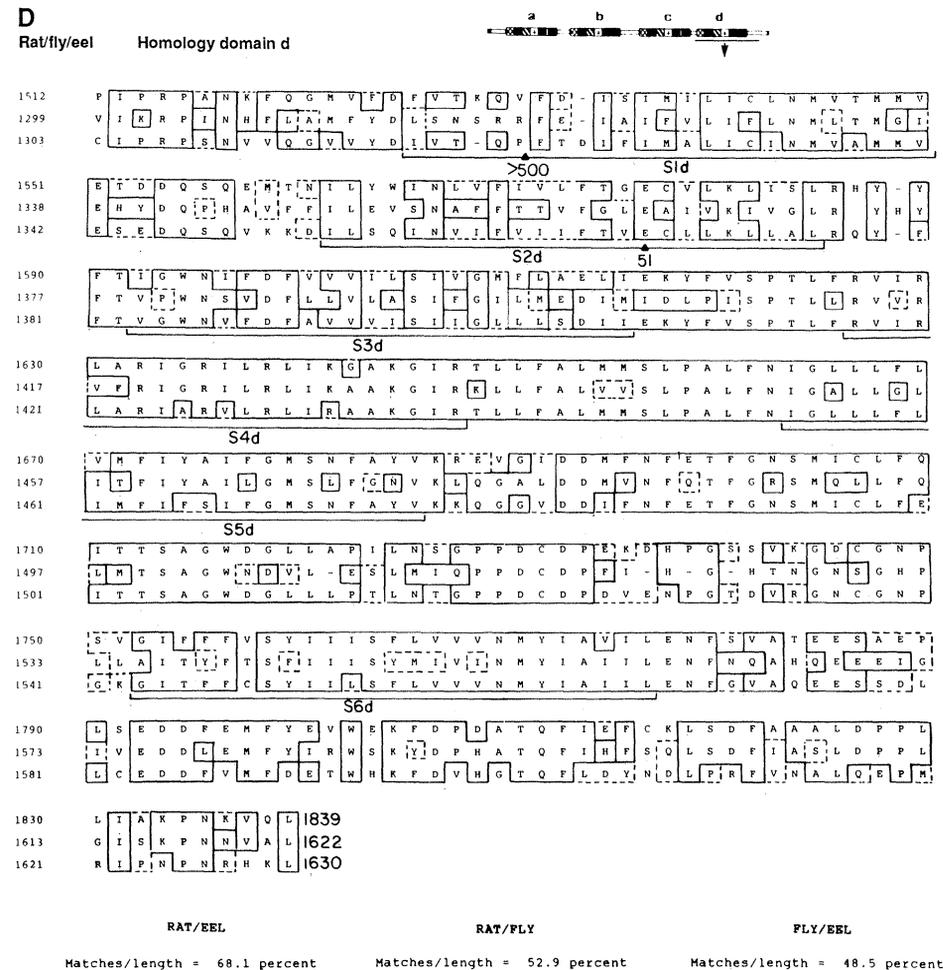
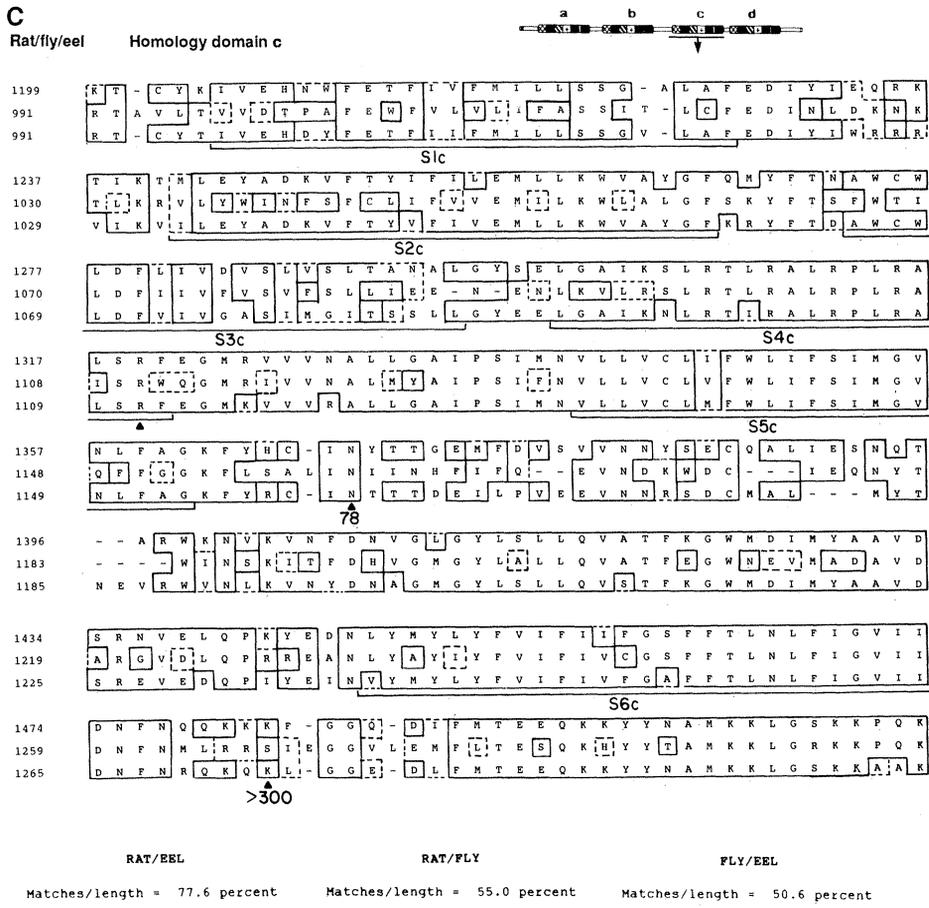


Fig. 4. Northern blot analysis of poly(A)-containing mRNA purified from flies of different developmental stages. The probe used is cDNA 3-1AL. Embryo (E), 0- to 9-hour embryos; embryo (L), 9- to 18-hour embryos; pupa, 30- to 96-hour post-pupariation pupae; adult, 1- to 2-week-old adults. Northern blot with high stringency hybridization conditions (see below). Large (>10 kb) messenger RNA's are present under these conditions. Messenger RNA's of two different sizes are present in early embryos and may represent alternative splicing products of the same gene. Conditions of high stringency hybridization: $5\times$ SSPE at 65°C ; washing in $0.1\times$ SSC at 55°C .

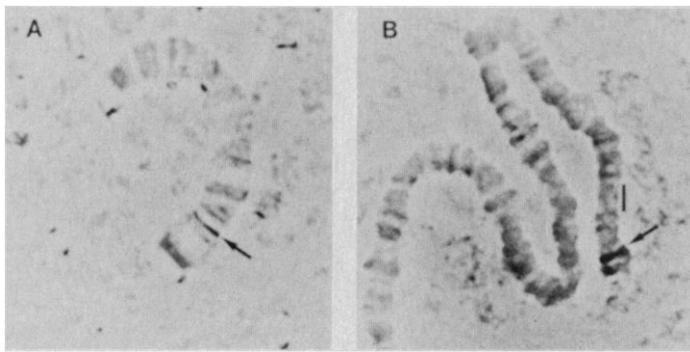


Fig. 5. Localization of two independently isolated clones to the right end of chromosome 2 (region 60D-E) by in situ hybridization to polytene chromosomes (21). The conditions for isolating these nonoverlapping clones are given in the legend to Fig. 1. In each part the arrow indicates the stained chromosome band from a biotin-labeled DNA probe (21). (A) Clone 3-1, the *Drosophila* channel gene. (B) Clone 2-6, a second clone mapping to the same region. Clone 2-6 was independently isolated and weakly cross hybridizes with clone 3-1. Clone 2-6 has not been sequenced. Because the DNA in clone 2-6 is poorly conserved as determined in interspecies blotting experiments, it is suspected that the clone may contain a pseudogene. Double label experiments with a mixed probe of clones 3-1 and 2-6 produce a stain signal that cannot be resolved into two bands. The bar indicates the approximate position of the *seizure* locus.

positive charges (S4a to S4d), one present in each of the homologous domains. Transmembrane strings of positive charges have been postulated as a mechanism of voltage-dependent gating before sequence data were available (12). Each positive charge in a string alternates with two uncharged residues. All 24 charges present in the vertebrate channel protein are present at identical positions in *Drosophila* (Fig. 2). The regularity and absolute conservation of charge distribution implicates this structure in channel gating. One model of channel gating (8, 9) proposes that these structures form channel-spanning alpha helices that rotate during gating to move positive charges across the plane of the membrane toward the outside.

The occurrence of an intron in the center of the highly conserved string of S4b gating charges is unexpected. Since introns commonly fall on the surface of proteins (13, 14), could this indicate that the central portions of the S4 structures are at the extracellular or intracellular surface of the protein? This seems unreasonable if present-day molecular models of sodium channels are accurate; these models are based on the assumption that the S4 structures are membrane-spanning alpha helices. It is also difficult to imagine how a surface placement of the S4 regions would permit these structures to respond to changes in the transmembrane voltage gradient. Instead of separating functional domains, the intron in S4 is situated centrally within an apparent single functional domain; this could suggest that the S4 regions evolved from still smaller functional domains.

In the deduced *Drosophila* channel peptide, S6 and S5 are the most hydrophobic of the presumed membrane-spanning regions, and are extremely similar between *Drosophila* and the vertebrate species. The high degree of conservation is consistent with the model of the sodium channel (8, 9) that places these structures in the interior core of the protein; it is presumed that amino acid side chains interacting with other protein structure constrains variation in side-chain size and shape because these interactions often must be precise. Comparative studies of protein evolution show that residues on the surface of proteins, whether exposed to water (13, 14) or membrane lipid (15, 16), are less well conserved than residues in the interior of proteins that interact with other protein structure. Signal sequences, hydrophobic peptides that interact with lipid, are poorly conserved in evolution (16).

One cytoplasmic linker is highly conserved. In contrast to the two larger linker regions a-b and b-c, the smaller c-d linker region shows high evolutionary conservation. The complete deduced amino acid sequence of the linker region is presented in Fig. 3, C and D. Armstrong and Bezanilla (17) proposed that inactivation is caused by a positively charged cytoplasmic domain that blocks the open channel from the cytoplasmic side. Conserved residues in the linker region c-d can account for many of the properties predicted by experimental observations. The presumed cytoplasmic region includes a highly positively charged domain containing six lysines in conserved positions. These could well be the targets of trypsin, which is known to eliminate inactivation by its action on a cytoplasmic domain of the channel (18). In addition, paired tyrosine residues at conserved positions in the center of this domain offer an additional target to other reactants that eliminate inactivation [see Hille, 1984 (7)].

In contrast to linker c-d, the linker regions a-b and b-c are poorly conserved; very large insertions and deletions occur even between the two vertebrate species. Some cytoplasmic domains of ion channels may not play a critical role in the biophysical properties of channel function. Experiments deleting portions of the putative cytoplasmic regions of the acetylcholine receptor have no detectable effect on channel conductance, activation of channels by agonist, or binding of antagonists (19). We have sequenced portions of linker regions a-b and b-c in the *Drosophila* gene and have found little homology to either of the vertebrate sodium channel genes.

Expression and cytology of the gene. The expression of message was indicated by Northern blot analysis, which showed that there is an mRNA larger than 10 kilobases and that it hybridizes to the cDNA probe 3-1AL (Fig. 1). This mRNA is detected in embryonic and pupal stages, and at an extremely low level in adult flies (Fig. 4). The message may be alternatively spliced at different developmental stages; two messages are present in early (0 to 9 hours) embryos, but only the larger of the two messages is present in late (9 to 18 hours) embryos and pupae. The message is barely detectable in adults.

The *Drosophila* channel gene maps cytogenetically to the right arm of the second chromosome. The position shown (arrow, Fig. 5) is judged to be in 60D-E. This is near the *seizure* locus at 60A-B (20) where mutations confer a temperature-sensitive paralytic phenotype. The *seizure* gene has been postulated to code for a sodium channel. Although these positions are intriguingly close, it is not known if there is any relationship between the *seizure* gene and the gene described here.

Ancient origin of the sodium channel. Each of the four homology domains in the vertebrate channel has structural features that distinguish it from the other three. In homology domain d, the structure at residue 1514 (fly) Pro-Pro-Asp-Cys-Asp-Pro is present only in homology domain d. This characteristic structure is present in all three species, rat, eel, and fly, and suggests that the intragenic duplication and specialization of the homology domains occurred before the separation of vertebrates from invertebrates in evolution.

Our study in *Drosophila* suggests that the major structural domains determining function in voltage-gated ion channels are very highly conserved. Hille (7) suggests that all voltage-gated ion channels are descendants of a common ancestral cation channel, and that ion selectivity was one of the last functional properties to evolve. The sodium channel gene seems to be the result of an intragenic duplication that produced the modern structure, namely, that four similar genes are welded together by linkers into a larger single gene. Apparently the ancestral gene resembled a single one of these homology domains and perhaps functioned as a homotetramer. The question as to whether ion selectivity to Ca^{2+} and K^{+} evolved before or after the events of intragenic duplication bears directly on the question of whether these other channel genes will

resemble either the large four-in-one sodium channel gene, a single one of the homology domains, or, as in the case of the acetylcholine receptor, four separate homologous genes.

REFERENCES AND NOTES

1. M. Noda *et al.*, *Nature (London)* **312**, 121 (1984).
2. M. Noda *et al.*, *ibid.* **320**, 188 (1986).
3. M. Noda *et al.*, *ibid.* **322**, 826 (1986).
4. W. Stühmer, C. Methfessel, B. Sakmann, M. Noda, S. Numa, *Eur. Biophys. J.* **14**, 131 (1987).
5. A. L. Goldin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7503 (1986).
6. W. A. Catterall, *Annu. Rev. Biochem.* **55**, 953 (1986).
7. B. Hille, *Ionic Channels of Excitable Membranes* (Sinauer, Sunderland, MA, 1984).
8. H. R. Guy and P. Seetharamulu, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 508 (1986).
9. H. R. Guy, in preparation.
10. L. Salkoff *et al.*, *Nucleic Acids Res.*, in press (contains the nucleotide sequence of the gene and the experimental details).
11. R. A. Padgett, P. J. Grabowski, M. M. Konarska, S. Seiler, P. A. Sharp, *Annu. Rev. Biochem.* **55**, 1119 (1986).
12. C. M. Armstrong, *Physiol. Rev.* **61**, 644 (1981).
13. C. S. Craik, W. J. Rutter, R. Fletterick, *Science* **220**, 1125 (1983).
14. C. S. Craik, S. Sprang, R. Fletterick, W. J. Rutter, *Nature (London)* **299**, 180 (1982).
15. E. Hafen, K. Basler, J.-E. Edstroem, G. M. Rubin, *Science* **236**, 55 (1987).
16. C. S. Zuker, A. F. Cowman, G. M. Rubin, *Cell* **40**, 851 (1985).
17. C. M. Armstrong and F. Bezanilla, *J. Gen. Physiol.* **70**, 567 (1977).
18. ———, E. Rojas, *ibid.* **62**, 375 (1973).
19. M. Mishima *et al.*, *Nature (London)* **313**, 364 (1985).
20. F. R. Jackson, J. Gitschier, G. R. Strichartz, L. M. Hall, *J. Neurosci.* **5**, 1144 (1985).
21. P. R. Langer-Sofer, M. Levine, D. C. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4381 (1982).
22. Supported by NIH grant 1 R01 NS24785-01, NSF grant BNS-8311024, and a grant from the Esther A. and Joseph Klingenstein Fund. We thank J. Garcia, D. Pauron, and C. Smith for technical aid and helpful comments. 24 April 1987; accepted 22 June 1987

Cloning of Genomic and Complementary DNA from *Shaker*, a Putative Potassium Channel Gene from *Drosophila*

DIANE M. PAPAZIAN,* THOMAS L. SCHWARZ, BRUCE L. TEMPEL,
YUH NUNG JAN, LILY Y. JAN

On the basis of electrophysiological analysis of *Shaker* mutants, the *Shaker* locus of *Drosophila melanogaster* has been proposed to encode a structural component of a voltage-dependent potassium channel, the A channel. Unlike sodium channels, acetylcholine receptors, and calcium channels, K⁺ channels have not been purified biochemically. To facilitate biochemical studies of a K⁺ channel, genomic DNA from the *Shaker* locus has been cloned. Rearrangements in five *Shaker* mutants have been mapped to a 60-kilobase segment of the genome. Four complementary DNA clones have been analyzed. These clones indicate that the *Shaker* gene contains multiple exons distributed over at least 65 kilobases of genomic DNA in the region where the mutations mapped. Furthermore, the gene may produce several classes of alternatively spliced transcripts. Two of the complementary DNA clones have been sequenced and their sequences support the hypothesis that *Shaker* encodes a component of a K⁺ channel.

EXCITABLE MEMBRANES OF NERVE AND MUSCLE ARE ENDOWED with a variety of potassium channels that are activated by voltage or calcium, or controlled by neurotransmitters (1). These channels conduct several kinetically distinct types of K⁺ currents, which have been characterized electrophysiologically. They determine, in large part, the rate of repolarization after action potentials and whether a cell will be quiescent or spontaneously active (1). Since even small alterations in the duration of action potentials can lead to large variations in the amount of transmitter released from nerve terminals, the regulation of K⁺ channels is likely

to modulate the strength of synaptic interactions in the nervous system (2). This regulation may be involved in complex behaviors, such as learning and memory. A well-studied example is the involvement of a serotonin-sensitive K⁺ channel in sensitization of the gill withdrawal reflex in *Aplysia* (3). Potassium channels also function in cells and tissues not generally considered to be excitable. For example, mitogenesis in T lymphocytes may involve a delayed rectifier-type K⁺ channel (4).

Although K⁺ channels are widely distributed, they have not yet been purified because no tissues rich in K⁺ channels have been identified, and until recently (5) no high-affinity ligands for them were known. Until now, genes for ion channels have been isolated only after channel proteins have been purified and characterized. For example, partial amino acid sequences of the Na⁺ channel and the subunits of the acetylcholine receptor have been used to design oligonucleotide probes for identifying the corresponding complementary DNA (cDNA) clones (6). This strategy has not been applicable to genes encoding K⁺ channels.

The genetics of *Drosophila melanogaster* provides an alternative approach for the molecular analysis of rare or uncharacterized proteins. This approach requires that mutations in the gene encoding the protein have been identified, and that the precise chromosomal location of the gene is known. Cloned genomic DNA from that region is isolated, and the gene is localized by mapping the molecular positions of mutations. Coding sequences can then be identified by using the genomic DNA to identify transcripts on

The authors are at the Howard Hughes Medical Institute and the Department of Physiology, University of California, San Francisco, CA 94143-0724.

*To whom correspondence should be addressed.