The EMBO Journal vol.7 no.4 pp.1087 - 1096, 1988

Shaker encodes a family of putative potassium channel proteins in the nervous system of *Drosophila*

O.Pongs, N.Kecskemethy, R.Müller, I.Krah-Jentgens, A.Baumann, H.H.Kiltz, I.Canal[1], S.Llamazares[1] and A.Ferrus[1]

Ruhr-Universität Bochum, Lehrstuhl für Biochemie, 4630 Bochum, FRG and [1]Instituto Cajal, Consejo Superior de Investigaciones Cientificas, Madrid, Spain

Communicated by B.Sakmann

The *Shaker* locus of *Drosophila* contains a very large transcription unit. It is expressed predominantly in the nervous system by multiple, differential as well as alternative, splicing mechanisms into different, but functionally related proteins. The structure of the *Shaker* transcription unit and the properties of the encoded *Shaker* protein family provide a molecular basis for A channel diversity in excitable cells.

Key words: Shaker/ion channels/nervous system

Introduction

By physiological and pharmacological criteria, a variety of potassium channels have been identified, which vary in distribution among different cellular regions and cell types displaying a spectrum of signaling capability (Hille, 1984). Despite this diversity, the 'transient outward current' (I_A) can be considered a characteristic feature of excitable cells since it regularly occurs in neuronal and non-neuronal cells (Ragowski, 1985). Voltage-dependent potassium (I_A) channels in excitable membranes are essential for many and diverse cellular functions, including action potential repolarization (Tanouye *et al.*, 1981), neuron bursting (Connor and Stevens, 1971), and cardiac pacemaking (Josephson *et al.*, 1984). The molecular genetic analysis of the *Shaker* gene complex (*Sh*) (X;57.6) of *Drosophila* (Solc *et al.*, 1987; Tempel *et al.*, 1987; Baumann *et al.*, 1987; Kamp *et al.*, 1987; Wu and Haugland, 1985; Koto *et al.*, 1981; Tanouye and Ferrus, 1985; Salkoff and Wyman, 1981) indicates that it encodes I_A-channel components. Single channel voltage-clamp techniques revealed an A-type potassium channel in *Drosophila* which was affected by *Shaker* mutations (Solc *et al.*, 1987). *Sh* mutants exhibit altered A-currents in voltage-clamped muscle cells (Wu and Haugland, 1985) and action potentials with delayed repolarization in identified neurons (Koto *et al.*, 1981; Tanouye and Ferrus, 1985). I_A defects include altered closing kinetics in some *Sh* alleles (Salkoff, 1983). Therefore, it is not surprising that a simple one to one correspondence between a specific *Sh* gene product and a particular potassium channel does not exist (Tanouye and Ferrus, 1985; Tanouye *et al.*, 1986).

We have now determined the primary structures of four different cDNAs (Shα, β, δ and ε) encoded in the *Sh* gene complex (Figure 1). The corresponding *Sh* transcripts originate from the same large transcription unit embedded in *Sh*. The transcripts apparently encode a family of ion channel proteins. The properties of this protein family provide a molecular basis for the diversity of A-currents in excitable membranes.

Results

A large transcription unit is embedded in Shaker

Several chromosomal rearrangements with cytogenetically separable breakpoints cause Sh physiology (Tanouye *et al.*, 1981; Salkoff, 1983). The *Sh* complex is limited, at present, by the most proximal (Df(1)S4010) and the most distal T(X;3)JC153 breakpoints that do not show Sh phenotypes. The complex can be subdivided (Tanouye *et al.*, 1986) into a haplolethal (HL), a viable (V) and a maternal effect (ME) region (Figure 1). The HL region is limited by the breakpoints of T(X;Y)V7 and T(X;Y)W32 while the ME region is limited by those of T(X;Y)B55 and T(X;Y)JC153. The V region, limited by T(X;Y)B55 and T(X;Y)W32, is not essential for viability. These chromosomal breakpoints of genetically mapped rearrangements with a Sh phenotype were located on the physical map (Baumann *et al.*, 1987).

Four different cDNAs were isolated from *Drosophila* cDNA libraries with DNA probes derived from the V region. Two of the cDNA sequences (Shα and Shδ) have been described previously (Tempel *et al.*, 1987; Baumann *et al.*, 1987). Shβ, Shδ and Shε cDNAs contained polyadenylated 3′ ends including polyadenylation signals within 30 nt 5′ of their poly(A) sequences (Figure 2). Shα cDNA, on the other hand, contained a truncated 3′-end and lacked a polyadenylation signal. The alignment of the four cDNA sequences with genomic DNA revealed 21 regions of colinearity which apparently belong to 21 exons from which mature Sh transcripts are derived (Figure 1). Most exons are separated by intervening sequences which are spliced out to generate transcripts with alternative exons. Exons 4/5 and 12/13, however, are not separated by intervening sequences. In this case, the differential use of a splice site generates either exon 4 or exon 4/5 and exon 12 or exon 12/13, respectively. The 21 exons are spread out over a transcription unit of >113 kb. It is the largest transcription unit described so far for *Drosophila*. For comparison, we have included in Figure 1 a fifth cDNA (Shγ) reported by Papazian *et al.* (1987) which probably consists of exons 3, 7−15, 19−21. the alignment of this cDNA sequence with the genomic DNA sequence is tentative since a Shγ sequence has not yet been published.

The largest introns separate exons 1 and 7 in Shβ and exons 2 and 7, in Shδ. The former is 86 kb and the latter 60 kb in length. The 5′ ends of Shβ and Shδ cDNA sequences were found, respectively, 10 and 50 kb proximal to T(X;Y)W32 inside the HL region (Figure 1). The 3′ ends of Shα and Shγ cDNA were located 10 kb distal to the distal T(X;Y)B55 breakpoint. This means that the *Sh* transcription unit has its 5′ end proximal to T(X;Y)W32 and the 3′ end distal to T(X;Y)B55. T(X;3)Sh^{LC} and T(X;Y)W32 mutants interrupt

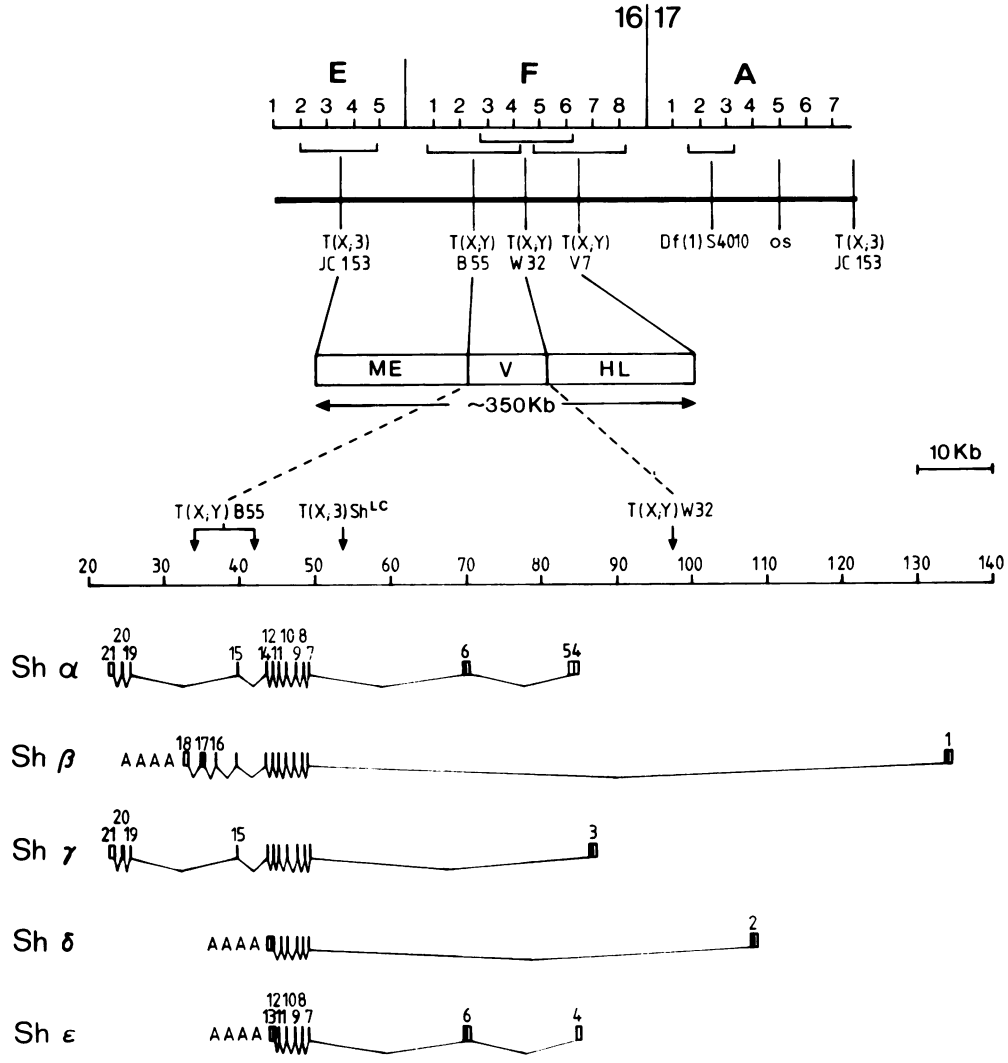
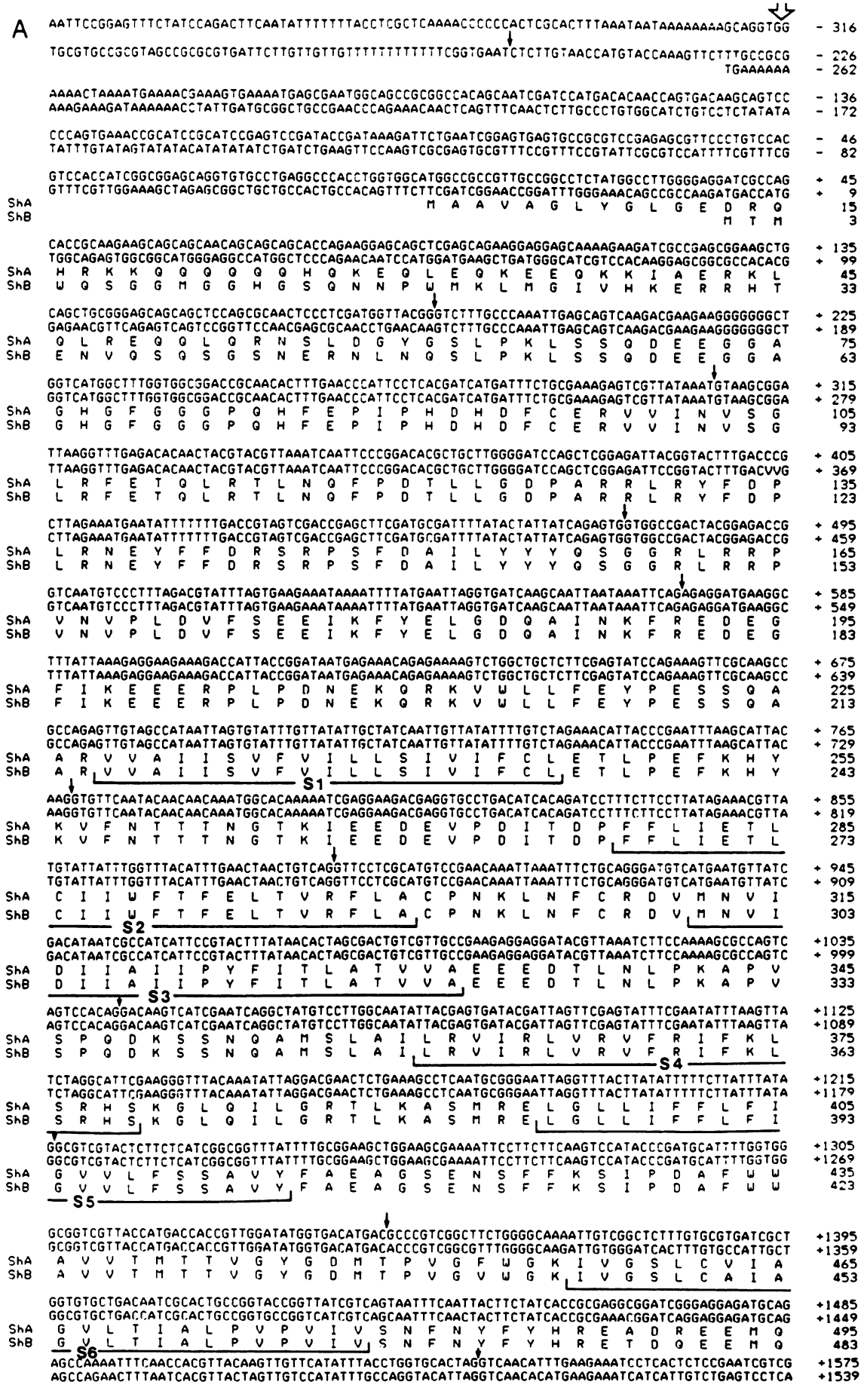


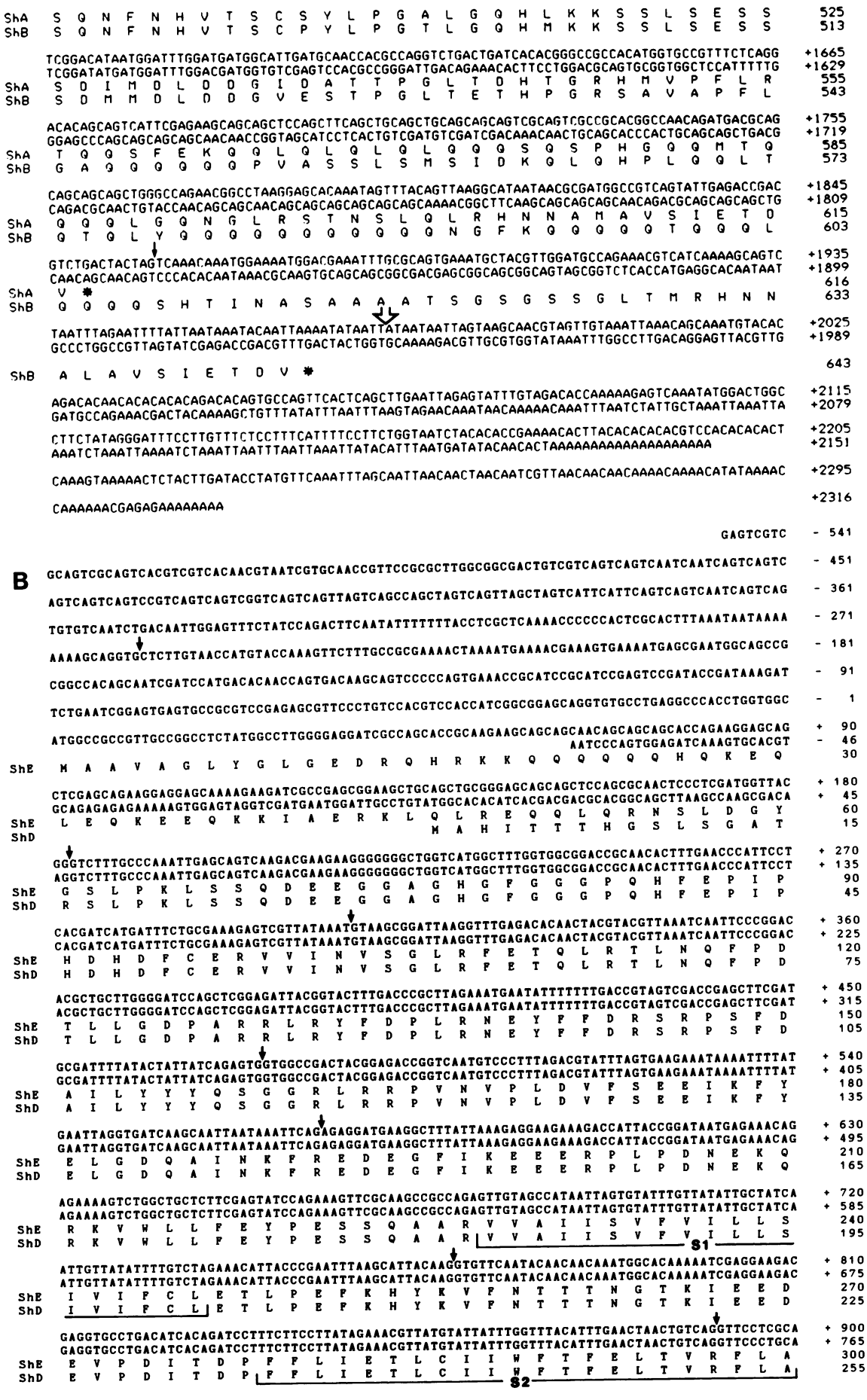
Fig. 1. Cytogenetic map of the *Shaker* (*Sh*) gene complex. **Top**. A schematic drawing of the salivary gland chromosome for this region is shown. The breakpoints of five chromosomal rearrangements used to localize *Sh* cytogenetically (Tanouye *et al.*, 1981) are indicated by brackets. The vertical bar gives the location of the mutation outstretched (os). The centromere is to the right. The genetic organization of the *Sh* gene complex into a maternal effect (ME), a viable (V) and haplolethal (HL) region is indicated by open bars. The regions span ~350 kb DNA (unpublished results). The location of translocation breakpoints of T(X;Y)B55, T(X;3)Sh^{LC} and T(X;Y)W32 are indicated by vertical arrows on the physical map of the *Sh* locus. The coordinates on the scale are as defined in Baumann *et al.* (1987). Calibration is in kb. **Bottom**. Structures of Sh transcripts aligned to the physical map of the genomic *Sh* region from which they derive. The 21 exons are indicated as boxes or bars whilst the introns are diagrammed as thin lines. Open boxes are untranslated regions, bars correspond to derived open reading frames of Sh cDNAs. Exons are numbered from right to left in the direction of transcription. AAAA indicates 3′ termini of Shβ, Shδ and Shε transcripts. Exons were aligned with genomic DNA by sequence comparison. Details of the genomic sequence will be published elsewhere. Sequences of Shα, Shβ and Shε are described in this paper. Details of the structure of Shδ are given in Baumann *et al.* (1987). The structure of Shγ was inferred from Tempel *et al.* (1987).

the *Sh* transcription unit (Figure 1). These data provide the molecular basis for their Sh phenotype. Synthesis of functional Sh proteins should be impaired and, hence, the A current is absent in these mutants (Tanouye *et al.*, 1981; Salkoff, 1983).

The alternate and/or differential use of exon/intron splice sites within the *Sh* transcription unit generates related Sh transcripts of differing lengths. The transcripts can be assigned to two classes based on their exon compositions. Class I consists of Shα, Shβ and Shγ transcripts, class II of Shδ and Shε transcripts. Both transcript classes share a central core made up of exons 7−12 which map within 5 kb of the *Sh* transcription unit (Figure 1). Class I transcripts extend beyond the T(X;Y)B55 breakpoint, class II transcripts do not. The differential use of the donor splice site at exon 12 leads to the synthesis either of class I or class II transcripts. If the donor splice site of exon 12 is not used, this exon becomes extended by 128 nt (exon 12/13 in Figure 1) yielding an alternative polyadenylation site. All Sh sequences have different and, as yet, incomplete 5′ ends. Perhaps, Shα and Shε have complete 5′ ends. The different 5′ ends are generated in Sh transcripts by alternative uses either of exon 1 (Shβ), exon 2 (Shα), exon 3 (Shγ) exons 4 and 6 (Shε), or of exons 4, 5 and 6 (Shα) (Figure 1). Exons 4 and 5, like exons 12/13, are not separated by an intron. Exon 5 extends at its 5′-side by 57 nt to exon 4/5. A differential use of donor sites generates either an exon 4 as was found in Shε or an exon 4/5 as in Shα.

Genomic DNA probes of the V- and the HL region hybridized in Northern blot experiments to several


A
S1
S2
S3
S4
S5
S6


ShA S Q N F N H V T S C S Y L P G A L G Q H L K K S S L S E S S 525
ShB S Q N F N H V T S C P Y L P G T L G Q H M K K S S L S E S S 513
TCGGACATAATGGATTTGGATGATGGCATTGATGCAACCACGCCAGGTCTGACTGATCACACGGGCCGCCACATGGTGCCGTTTCTCAGG +1665
TCGGATATGATGGATTTGGACGATGGTGTCGAGTCCACGCCGGGATTGACAGAAACACTTCCTGGACGCAGTGCGGTGGCTCCATTTTTG +1629
ShA S D I M D L D D G I D A T T P G L T D H T G R H M V P F L R 555
ShB S D M M D L D D G V E S T P G L T E T H P G R S A V A P F L 543
ACACAGCAGTCATTCGAGAAGCAGCAGCTCCAGCTTCAGCTGCAGCTGCAGCAGCAGTCGCAGTCGCCGCACGGCCAACAGATGACGCAG +1755
GGAGCCCAGCAGCAGCAGCAACAACCGGTAGCATCCTCACTGTCGATGTCGATCGACAAACAACTGCAGCACCCACTGCAGCAGCTGACG +1719
ShA T Q Q S F E K Q Q L Q L Q L Q L Q Q Q S Q S P H G Q Q M T Q 585
ShB G A Q Q Q Q Q Q P V A S S L S M S I D K Q L Q H P L Q Q L T 573
CAGCAGCAGCTGGGCCAGAACGGCCTAAGGAGCACAAATAGTTTACAGTTAAGGCATAATAACGCGATGGCCGTCAGTATTGAGACCGAC +1845
CAGACGCAACTGTACCAACAGCAGCAACAGCAGCAGCAGCAGCAGCAAAACGGCTTCAAGCAGCAGCAGCAACAGACGCAGCAGCAGCTG +1809
ShA Q Q Q L G Q N G L R S T N S L Q L R H N N A M A V S I E T D 615
ShB Q T Q L Y Q Q Q Q Q Q Q Q Q Q N G F K Q Q Q Q Q T Q Q Q L 603
GTCTGACTACTAGTCAAACAAATGGAAAATGGACGAAATTTGCGCAGTGAAATGCTACGTTGGATGCCAGAAACGTCATCAAAAGCAGTC +1935
CAACAGCAACAGTCCCACACAATAAACGCAAGTGCAGCAGCGGCGACGAGCGGCAGCGGCAGTAGCGGTCTCACCATGAGGCACAATAAT +1899
616
ShA V *
ShB Q Q Q Q S H T I N A S A A A A T S G S G S S G L T M R H N N 633
TAATTTAGAATTTTATTAATAAATACAATTAAAATATAATTATAATAATTAGTAAGCAACGTAGTTGTAAATTAAACAGCAAATGTACAC +2025
GCCCTGGCCGTTAGTATCGAGACCGACGTTTGACTACTGGTGCAAAAGACGTTGCGTGGTATAAATTTGGCCTTGACAGGAGTTACGTTG +1989
643
ShB A L A V S I E T D V *
AGACACAACACACACAGACACAGTGCCAGTTCACTCAGCTTGAATTAGAGTATTTGTAGACACCAAAAAGAGTCAAATATGGACTGGC +2115
GATGCCAGAAACGACTACAAAAGCTGTTTATATTTAATTTAAGTAGAACAAATAACAAAAACAAATTTAATCTATTGCTAAATTAAATTA +2079
CTTCTATAGGGATTTCCTTGTTTCTCCTTTCATTTTCCTTCTGGTAATCTACACACCGAAAACACTTACACACACACGTCCACACACACT +2205
AAATCTAAATTAAAATCTAAATTAATTTAATTAAATTATACATTTAATGATATACAACACTAAAAAAAAAAAAAAAAAAA +2151
CAAAGTAAAAACTCTACTTGATACCTATGTTCAAATTTAGCAATTAACAACTAACAATCGTTAACAACAACAAAACAAAACATATAAAAC +2295
CAAAAAACGAGAGAAAAAAAA +2316
GAGTCGTC - 541
B GCAGTCGCAGTCACGTCGTCACAACGTAATCGTGCAACCGTTCCGCGCTTGGCGGCGACTGTCGTCAGTCAGTCAATCAATCAGTCAGTC - 451
AGTCAGTCAGTCCGTCAGTCAGTCGGTCAGTCAGTTAGTCAGCCAGCTAGTCAGTTAGCTAGTCATTCATTCAGTCAGTCAATCAGTCAG - 361
TGTGTCAATCTGACAATTGGAGTTTCTATCCAGACTTCAATATTTTTTTACCTCGCTCAAAACCCCCCACTCGCACTTTAAATAATAAAA - 271
AAAAGCAGGTGCTCTTGTAACCATGTACCAAAGTTCTTTGCCGCGAAAACTAAAATGAAAACGAAAGTGAAAATGAGCGAATGGCAGCCG - 181
CGGCCACAGCAATCGATCCATGACACAACCAGTGACAAGCAGTCCCCCAGTGAAACCGCATCCGCATCCGAGTCCGATACCGATAAAGAT - 91
TCTGAATCGGAGTGAGTGCCGCGTCCGAGAGCGTTCCCTGTCCACGTCCACCATCGGCGGAGCAGGTGTGCCTGAGGCCCACCTGGTGGC - 1
ATGGCCGCCGTTGCCGGCCTCTATGGCCTTGGGGAGGATCGCCAGCACCGCAAGAAGCAGCAGCAACAGCAGCAGCACCAGAAGGAGCAG + 90
AATCCCAGTGGAGATCAAAGTGCACGT - 46
ShE M A A V A G L Y G L G E D R Q H R K K Q Q Q Q Q Q H Q K E Q 30
CTCGAGCAGAAGGAGGAGCAAAAGAAGATCGCCGAGCGGAAGCTGCAGCTGCGGGAGCAGCAGCTCCAGCGCAACTCCCTCGATGGTTAC + 180
GCAGAGAGAGAAAAAGTGGAGTAGGTCGATGAATGGATTGCCTGTATGGCACACATCACGACGACGCACGGCAGCTTAAGCCAAGCGACA + 45
ShE L E Q K E E Q K K I A E R K L Q L R E Q Q L Q R N S L D G Y 60
ShD M A H I T T T H G S L S G A T 15
GGGTCTTTGCCCAAATTGAGCAGTCAAGACGAAGAAGGGGGGGCTGGTCATGGCTTTGGTGGCGGACCGCAACACTTTGAACCCATTCCT + 270
AGGTCTTTGCCCAAATTGAGCAGTCAAGACGAAGAAGGGGGGGCTGGTCATGGCTTTGGTGGCGGACCGCAACACTTTGAACCCATTCCT + 135
ShE G S L P K L S S Q D E E G G A G H G F G G G P Q H F E P I P 90
ShD R S L P K L S S Q D E E G G A G H G F G G G P Q H F E P I P 45
CACGATCATGATTTCTGCGAAAGAGTCGTTATAAATGTAAGCGGATTAAGGTTTGAGACACAACTACGTACGTTAAATCAATTCCCGGAC + 360
CACGATCATGATTTCTGCGAAAGAGTCGTTATAAATGTAAGCGGATTAAGGTTTGAGACACAACTACGTACGTTAAATCAATTCCCGGAC + 225
ShE H D H D F C E R V V I N V S G L R F E T Q L R T L N Q F P D 120
ShD H D H D F C E R V V I N V S G L R F E T Q L R T L N Q F P D 75
ACGCTGCTTGGGGATCCAGCTCGGAGATTACGGTACTTTGACCCGCTTAGAAATGAATATTTTTTTGACCGTAGTCGACCGAGCTTCGAT + 450
ACGCTGCTTGGGGATCCAGCTCGGAGATTACGGTACTTTGACCCGCTTAGAAATGAATATTTTTTTGACCGTAGTCGACCGAGCTTCGAT + 315
ShE T L L G D P A R R L R Y F D P L R N E Y F F D R S R P S F D 150
ShD T L L G D P A R R L R Y F D P L R N E Y F F D R S R P S F D 105
GCGATTTTATACTATTATCAGAGTGGTGGCCGACTACGGAGACCGGTCAATGTCCCTTTAGACGTATTTAGTGAAGAAATAAAATTTTAT + 540
GCGATTTTATACTATTATCAGAGTGGTGGCCGACTACGGAGACCGGTCAATGTCCCTTTAGACGTATTTAGTGAAGAAATAAAATTTTAT + 405
ShE A I L Y Y Y Q S G G R L R R P V N V P L D V F S E E I K F Y 180
ShD A I L Y Y Y Q S G G R L R R P V N V P L D V F S E E I K F Y 135
GAATTAGGTGATCAAGCAATTAATAAATTCAGAGAGGATGAAGGCTTTATTAAAGAGGAAGAAAGACCATTACCGGATAATGAGAAACAG + 630
GAATTAGGTGATCAAGCAATTAATAAATTCAGAGAGGATGAAGGCTTTATTAAAGAGGAAGAAAGACCATTACCGGATAATGAGAAACAG + 495
ShE E L G D Q A I N K F R E D E G F I K E E E R P L P D N E K Q 210
ShD E L G D Q A I N K F R E D E G F I K E E E R P L P D N E K Q 165
AGAAAAGTCTGGCTGCTCTTCGAGTATCCAGAAAGTTCGCAAGCCGCCAGAGTTGTAGCCATAATTAGTGTATTTGTTATATTGCTATCA + 720
AGAAAAGTCTGGCTGCTCTTCGAGTATCCAGAAAGTTCGCAAGCCGCCAGAGTTGTAGCCATAATTAGTGTATTTGTTATATTGCTATCA + 585
ShE R K V W L L F E Y P E S S Q A A R V V A I I S V F V I L L S 240
ShD R K V W L L F E Y P E S S Q A A R V V A I I S V F V I L L S 195
S1
ATTGTTATATTTTGTCTAGAAACATTACCCGAATTTAAGCATTACAAGGTGTTCAATACAACAACAAATGGCACAAAAATCGAGGAAGAC + 810
ATTGTTATATTTTGTCTAGAAACATTACCCGAATTTAAGCATTACAAGGTGTTCAATACAACAACAAATGGCACAAAAATCGAGGAAGAC + 675
ShE I V I F C L E T L P E F K H Y K V F N T T T N G T K I E E D 270
ShD I V I F C L E T L P E F K H Y K V F N T T T N G T K I E E D 225
GAGGTGCCTGACATCACAGATCCTTTCTTCCTTATAGAAACGTTATGTATTATTTGGTTTACATTTGAACTAACTGTCAGGTTCCTCGCA + 900
GAGGTGCCTGACATCACAGATCCTTTCTTCCTTATAGAAACGTTATGTATTATTTGGTTTACATTTGAACTAACTGTCAGGTTCCCTGCA + 765
ShE E V P D I T D P F F L I E T L C I I W F T T F E L T V R F L A 300
ShD E V P D I T D P F F L I E T L C I I W F T T F E L T V R F L A 255
S2

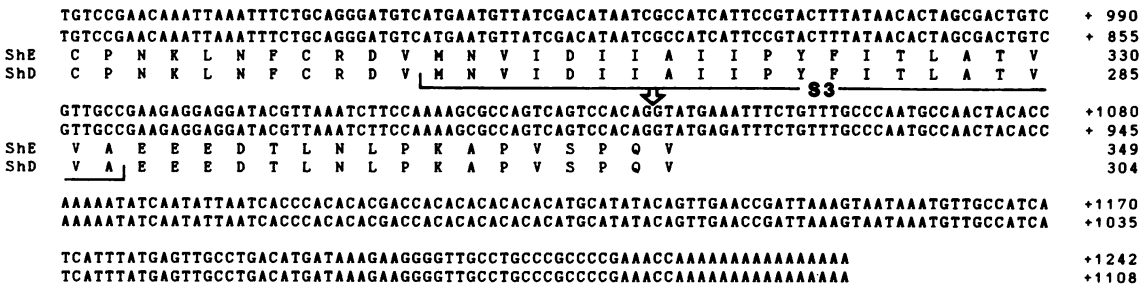


Fig. 2. A. Nucleotide sequences of cloned Shα (top) and Shβ cDNA (bottom) and alignment of the deduced amino-acid sequences (in one letter code) of ShA and ShB proteins. **B**. Nucleotide sequences of cloned Shδ (Baumann *et al.*, 1987) (bottom) and Shε cDNA (top) and alignment of the deduced amino-acid sequences (in one letter code) of ShD and ShE proteins. Nucleotides are numbered in the 5′−3′ direction, beginning with the first residue of the ATG triplet encoding the initiative methionine and the nucleotides on the 5′ side of residue 1 are indicated by negative numbers. The number of the nucleotide residue at the right end of each line is given. The deduced amino-acid sequences of Sh proteins are shown below the nucleotide sequences. Amino acid residues are numbered beginning with the initiative methionine. Numbers of the last residues are given in the right-hand side. Nonsense codons at the ends of ORFs are marked by an asterisk. The proposed transmembrane segments S1−S6 of ShA and ShB and S1−S3 of ShD and ShE are indicated. The termini of each segment are tentatively assigned on the basis of the hydropathy profiles (Figure 3A). Polyadenylation signals are found at nucleotides 2109−2114 (ATTAAA) in Shβ, at nucleotides 1193−1198 (GATAAA) in Shε and at nucleotides 1058−1063 in Shδ, respectively. Splice junctions of introns are indicated by arrows. Open arrows indicate splice junctions which are not used (A) in Shα or (B) in Shδ and Shε cDNA. Sh-cDNA sequences were identical to the corresponding genomic (Canton-S) exon sequences. Exon−intron splice sites were determined by sequencing genomic DNA between map positions (Figure 1) +25 to +50, +69 to +71, +83 to +85, +108 to +109 and +132 to +136. Sequence differences between Shα and a similar, 200 nt longer sequence in (Tempel *et al.*, 1987) were at nucleotides −46 (T→C), +1390 (G→A) and +1463 (A→G). These differences resulted in one amino acid replacement (Val 464 to Ile).

Drosophila poly(A)$^+$ RNAs (Kamb *et al.*, 1987). This result supports the proposition that the *Sh* transcription unit generates various Sh transcripts by alternative/differential splicing mechanisms of a large primary transcript.

Sh encodes a family of putative ion-channel proteins

The different Sh cDNA sequences code for related proteins. It is likely that the *Sh* transcription unit creates also mRNAs with differing 3′ and/or 5′ sequences without altering a particular ORF. Therefore, we have named the cDNAs with Greek letters and protein sequences derived from corresponding open reading frames with Roman letters. Shα is very similar to a cDNA sequence previously described (Tempel *et al.*, 1987). The Shα sequence has within its longest ORF only two differences which are A→G exchanges. This leads to one conservative amino acid replacement (Val 464 to Ile 464) in the derived protein sequence. The Shα sequence is 2722 nt long (Figure 2A). Comparison of the cDNA sequence with the corresponding genomic region revealed 12 points of non-colinearity, indicative of 12 intervening sequences in the primary transcript (Figures 1 and 2A). The splice junctions for these intervening sequences are located at positions −317, +182, +306, +475, +572, +768, +890, +1044, +1216, +1346, +1537 and +1858. The 3′ end of Shα (exon 21) contained a short poly(dA) track which was also found in the genomic DNA sequence. Therefore, it is probably not complete. The derived ShA protein sequence is 616 amino acids long with a molecular mass of 70 272 daltons. Possible I_A channel components of similar size (65−75 kd) have been identified (Halliwell *et al.*, 1986) in vertebrate brain by cross-linking studies with dendrotoxin, an I_A channel blocker.

The Shβ sequence is 2401 nt long aside from the poly(dA) track (Figure 2A). Comparison of this cDNA sequence with the corresponding genomic region revealed 11 points of non-colinearity, indicative of 11 intervening sequences in the primary transcript (Figures 1 and 2A). The splice junctions for these intervening sequences are located at position +146, +270, +429, +536, +732, +854, +1008, +1180, +1310, +1501 and +1939. The 3′ end of Shβ (exon 18) contained a poly(dA) track which was not found in the genomic DNA sequence. Therefore, this 3′ end should be the 3′ terminus of the Shβ transcript A protein sequence (ShB) was derived from translating the longest ORF which is 643 amino acids long with a molecular mass of 72 508 daltons. Hydropathy analysis (Kyte and Doolittle, 1982) (Figure 3A) shows that the ShB protein has a central hydrophobic core which is identical with that of ShA protein in number and kind of possible membrane-spanning helices. The first five are encoded in exons 10−15 common to Shα and Shβ cDNAs. The sixth putative membrane-spanning helix of ShA is derived from exon 19, that of ShB from exon 16. The alternative use of exons 16 or 19 does not alter the sequence of the presumed transmembrane spanning helix except for one conservative amino-acid replacement (Val 463 in ShA to Ala 451 in ShB). It is not clear how this might effect the properties of Sh protein. On the other hand, the 5′- and 3′-terminal sequences of Shα and Shβ and, therefore, the hydrophobic amino and carboxyl termini of ShA and ShB are different. The amino terminus of ShA is more hydrophilic than that of ShB. However, the carboxyl terminus of ShA is less hydrophilic than that of ShB. Both protein sequences contain polyglutamine tracks (*opa* element) (Wharton *et al.*, 1985). They appear either at the amino (ShA) or at the carboxyl terminus (ShB) (Figure 2A).

The derived ShA and ShB protein sequences have the properties of integral membrane proteins containing six identical, potentially membrane-spanning segments (Figure 3A). Their topology as well as amino-acid sequences are strikingly similar to that of vertebrate voltage-dependent sodium channels (Noda *et al.*, 1984, 1986) and of the rabbit dihydropyridine receptor (Tanabe *et al.*, 1987) (Figure 3B). Therefore, we maintained for the proposed Sh transmembrane segments the nomenclature that has been used for vertebrate voltage-dependent channels. We assume in analogy to the proposed model for the sodium channel (Noda *et al.*, 1986) that the six presumably α-helical segments S1−S6 are similarly oriented in a pseudosymmetric fashion

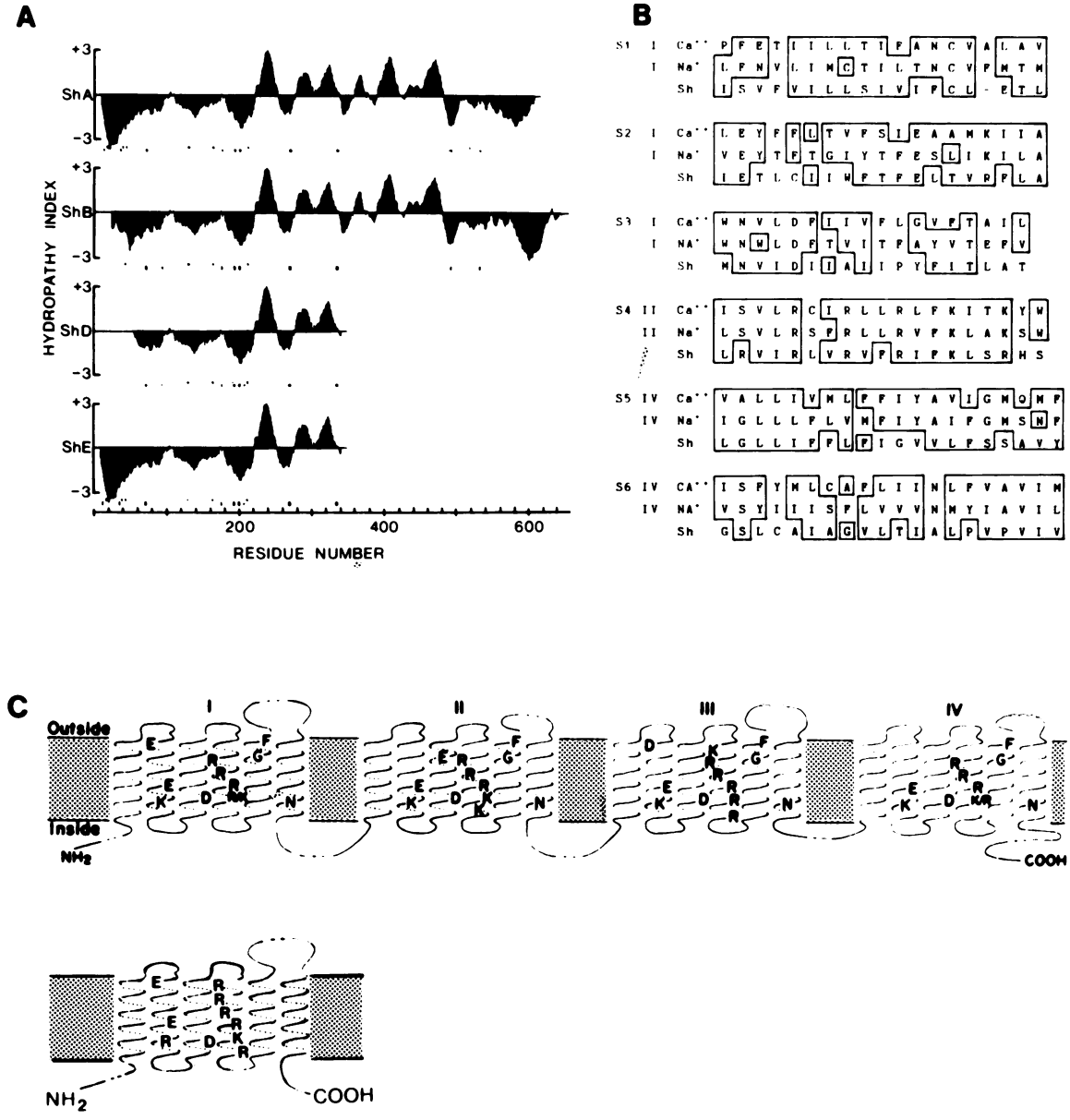


Fig. 3. A. Hydropathy plots of *Sh* proteins. Hydropathy indices are plotted versus amino-acid residues. The plot was generated by the program of Kyte and Doolittle (1982) with a window size of 19 amino acids. Positive index values indicate hydrophobic groups, negative index values hydrophilic groups. The positions of positively charged residues (Lys and Arg) and negatively charged residues (Asp and Glu) are indicated by upward and downward vertical lines, respectively. **B.** Sequence similarities between proposed transmembrane segments S1–S6 between the dihydropyridine receptor (Ca^{2+}) (Tanabe *et al.*, 1987), the voltage-dependent sodium channel II of rat brain (Na^{+}) (Noda *et al.*, 1986) and *Sh* protein (Sh). The amino-acid sequence is given in one-letter code: segment S1, amino acids 52–70 (repeat I of dihydropyridine receptor), 119–141 (repeat I of rat brain sodium channel II), 220–237 (ShB); segment S2, amino acids 89–107 (repeat I of dihydropyridine receptor), 158–176 (repeat I of rat brain sodium channel II), 269–288 (ShB), segment S3, amino acids 122–139 (repeat I of dihydropyridine receptor), 191–208 (repeat I of rat brain sodium channel II), 300–317 (ShB); segment S4, amino acids 524–542 (repeat II of dihydropyridine receptor), 847–865 (repeat II of rat brain sodium channel II), 349–367 (ShB); segment S5, amino acids 1270–1289 (repeat IV of dihydropyridine receptor), 1663–1682 (repeat IV of rat brain sodium channel II), 384–403 (ShB); segment S6, amino acids 1362–1381 (repeat IV of dihydropyridine receptor), 1757–1776 (repeat IV of rat brain sodium channel II), 447–466 (ShB). Amino acids have been allotted to the following similarity groups: I, V, L, M/Y, F, W/S, T, A, G, P/E, D, Q, N/K, R. **C.** Schematic representation of structural characteristics common to the dihydropyridine receptor, the voltage gated sodium channel and *Sh* proteins. Data for vertebrate ion channels are taken from Tanabe *et al.* (1987). The proposed transmembrane α-helical segments S1–S6 (from left to right) are displayed linearly. The dihydropyridine receptor and the sodium channel have four units of homology spanning the membrane, which are assumed to surround the ionic channel. *Sh* proteins only have one such unit. Amino acids conserved in all segments of the dihydropyridine (Tanabe *et al.*, 1987), the three known sodium channels (Noda *et al.*, 1984, 1986) and the *Sh* proteins are shown in one-letter code. The amino acid numbers of ShB are as follows: Glu residues 271, 281 and Arg residues 285 in segment S2; Asp residue 304 in segment S3; Arg residues 353, 356, 359, 365 and Lys residue 362 in segment S4.

across the membrane. The amino and carboxyl termini are on the cytoplasmic side of the membrane (Figure 3C). This model is consistent with two of the three *N*-glycosylation sites (Hubbard and Ivatt, 1981) (Asn 247 and Asn 251 in ShB) and the two potential cAMP-dependent phosphorylation sites (Krebs and Beavo, 1979) (Ser 380 and Ser 507 in ShB) being located on the extracellular and on the cytoplasmic side, respectively.

Vertebrate ion channels contain four internal repeats with similar topology (Tanabe *et al.*, 1987). Each internal repeat has five hydrophobic segments (S1, S2, S3, S5 and S6) and one positively charged segment (S4). The Sh proteins do not have a repeat structure (Figure 3C). The proposed membrane-spanning sequences are encoded in separate exons; segment S1 in exon 10, S2 in exon 11, S3 in exon 12, S4 in exon 14, S6 in exons 16 or 19. Segment 5, interestingly, is encoded in exons 14 and 15. Since the proximal T(X;Y)B55 breakpoint was located between exons 14 and 15 (Figure 1), ShA and ShB proteins of this mutant cannot contain an intact segment S5 and lack segment S6 altogether. The alignment of positive charges in S4 is thought to carry the gating current (Catterall, 1985). Since the segment S4 has been proposed to be the voltage sensor in voltage-dependent channels (Noda *et al.*, 1986), it is conceivable that the altered topology in Sh proteins of T(X;Y)B55 could account for the residual A current found in this mutant (Salkoff, 1983).

In addition to the reported Shδ (Baumann *et al.*, 1987), Shε is another class II cDNA. The Shε sequence is 1774 nt long (Figure 2B). A protein sequence (ShE) is derived from the longest ORF which is 349 amino acids long and has a molecular mass of 40 497 daltons. The first 16 amino acids of ShD (exon 2) are replaced by 61 amino acids in ShE (exon 6) including an *opa*-element. ShD is the only Sh protein so far which does not contain a polyglutamine sequence. The hydropathy analysis of ShE revealed three potentially membrane-spanning sequences at amino-acid residues 228–246, 279–300 and 312–332 (Figures 2B and 3A). Class II transcripts do not have segments S4, S5 and S6. Since the donor splice site of exon 12 is not used, the open reading frames of class II transcripts terminate one amino acid after the end of exon 12 in exon 12/13 (Figure 2B). Except for the last amino acid Val, ShE is identical to the first 348 amino acids of ShA.

The comparison of ShA, ShB, ShD and ShE sequences shows that the *Sh* transcription unit expresses a family of related proteins. Probably, the Sh family has more members than the four derived protein sequences shown in Figure 2. Members of this family differ in their amino- and in their carboxyl-terminal sequences, in the location of *opa*-elements, as well as in number and kinds of proposed transmembrane helices as summarized in Figure 4. ShA and ShB have the six segments S1–S6, ShD and ShE only the three segments S1–S3. These structural differences imply different functional properties of ShA/ShB versus ShD/ShE. The sequence criteria suggest that all *Sh* cDNA clones contain complete ORFs, although the sequences surrounding the first AUG codons in the ORFs differ from the consensus for a ribosome binding site compiled from known initiation sequences (Cavener, 1987). The biochemical analyses of isolated Sh proteins will provide the definitive assignments.

Tissue expression of Sh transcripts

A riboprobe originated from Shδ cDNA was prepared in an expression vector (see Materials and methods). Shδ cDNA was selected as a hybridization probe because it does not contain an *opa*-element. Sense and antisense riboprobes were hybridized to serial sections from embryo, third instar larvae, white pupae and adults (Figure 5). The Sh transcripts seem to localize in the somata of the nervous system continuously from the mature embryo to the adult stages. Data

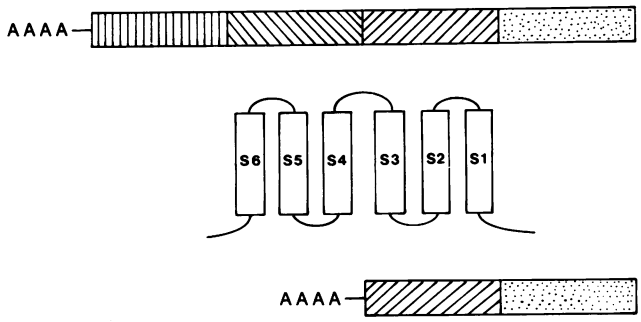


Fig. 4. Schematic structure of *Sh* transcripts to illustrate their diversity. The direction of transcription is from right to left. 3′ ends of *Sh* transcripts are indicated by AAA. Class I transcripts (**top**) have variable 5′ ends, a central core with the proposed arrangement of transmembrane segments S1–S6, and variable 3′ ends. Class II transcripts (**bottom**) have variable 5′ ends, a central core with the proposed arrangement of transmembrane segments S1–S3, and a common 3′ end. Segments S1–S6 are indicated by vertical open boxes. Stippled horizontal boxes indicate the 5′ ends which occur interchangeably in class I and class II transcripts; diagonally striped boxes indicate the two central cores, coding for segments S1–S3 and S4–S6, respectively. The hatched box represents the variable 3′ ends of class I transcripts. It should be noted that segment S6 is derived from alternative exons which leads to replacement of Val 463 in ShA by Ala 451 in ShB (Figure 3A).

from the optic ganglia anlage in the late third instar larvae indicate that Sh transcripts accumulate in the mature cells rather than in the newly divided ones. In addition to the central nervous system, a significant signal of hybridization was found in adult retina. Although there is electrophysiological evidence of I_A channels in larval and adult muscles (Solc *et al.*, 1987; Wu and Haugland, 1985; Salkoff, 1983), *in situ* hybridization did not reveal a prominent signal in this tissue. More sensitive, immunological methods would resolve the expression of Sh products in muscles. The ovaries appear to be deprived of Sh transcripts with the possible exception of the muscle envelopes. Also, at the early stages of embryogenesis there is no indication of transcriptional activity from the V region of *Sh*.

To control for possible cross-hybridization with other transcripts of similar sequences, $Df(1)B55^{D}/W32^{P}$ males which are deficient for the V region (Figure 1), were hybridized with the riboprobe used (Figure 5G). These flies did not show any sign of hybridization. Since the riboprobe used contained exons 7–12, which are common to all Sh transcripts, no differential expression could be detected. However, since the most significant signal was found in the nervous system, it can be said that all Sh transcripts are functionally related to the biology of the nervous system.

Discussion

Structural implications

Taking together the physiological phenotypes of *Sh* mutants (Wu and Haugland, 1985; Tanouye and Ferrus, 1985; Tanouye *et al.*, 1986; Salkoff, 1983) and the deduced Sh protein sequences strongly suggest that Sh proteins are subunits of potassium channels. The similarities in topology and amino-acid sequence between sodium channel (Noda *et al.*, 1984, 1986), dihydropyridine receptor (Tanabe *et al.*, 1987) and Sh proteins also indicate that voltage-dependent ion channels share many functional properties (Hille, 1984) because of common structural features. The location of

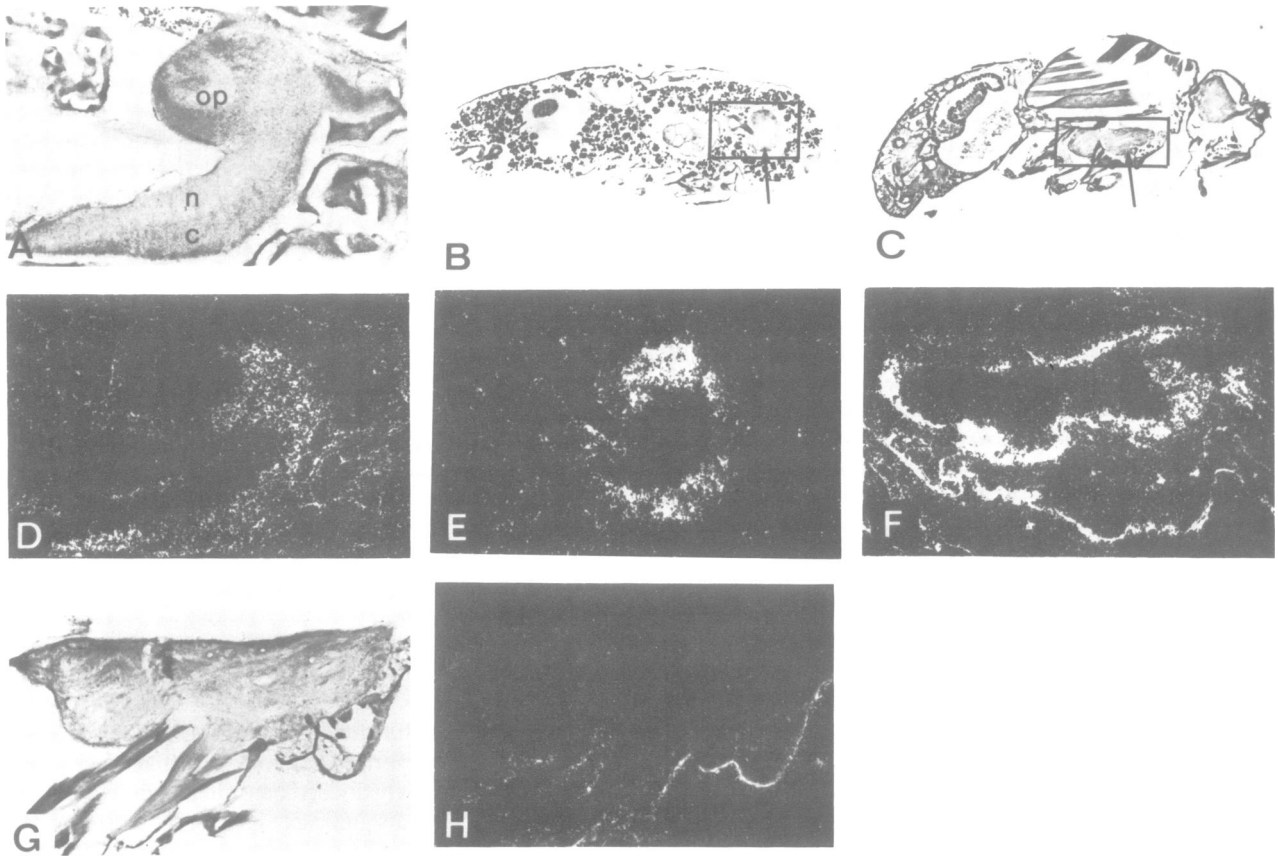


Fig. 5. *In situ* hybridization of a riboprobe from Shδ cDNA to tissue sections. **A**. Bright field view of a third larval instar CNS in saggital section. Anterior is to the right and dorsal is up. n, neuropile; c, cortex; op, developing optic ganglia. **B**. Saggital section of a white pupa. Anterior is to the right and dorsal is up. During metamorphosis most larval tissues undergo histolysis. This appears as dark spots all over the body. The nervous system does not show substantial cell death but general reorganization of neural projections (see text). Arrow points to the brain within inset. **C**. Saggital section of a mature adult. Anterior is to the right and dorsal is up. Arrow points to the thoracic ganglia within inset. **D**. Dark field view of A. The grains of radioactive signal are located in the cortex of the entire CNS except the developing optic ganglia where the cells are still undergoing divisions and are not yet mature neurons. Note also the absence of signal in imaginal tissue (upper right corner). **E**. Dark field view of inset in B. The signal locates abundantly in the cortex demonstrating that Sh products are expressed during metamorphosis when most cells undergo developmental changes and, presumably, manifest physiological properties different from those shown by mature, terminally differentiated states. **F**. Saggital section of 7-day adult thoracic ganglia equivalent to that shown in inset of C with arrow. Anterior is to the left and dorsal is up. The signal is as abundant in the somata as in the larval CNS. **G**. Saggital section of an adult thoracic ganglia from a $Df(1)B55^{D}/W32^{P}$ viewed in bright field. Anterior is to the right and dorsal is up. **H**. Dark field view of G. No signal can be detected demonstrating that the hybridization found in D, E and F corresponds to structural sequences coded only in the V region of the *Sh* complex.

charged residues in segments S2, S3 and S4 is conserved in the individual repeats of the dihydropyridine receptor and the three known voltage-dependent sodium channels (Tanabe *et al.*, 1987). The charge distribution in the proposed membrane-spanning helices of Sh proteins is remarkably similar (Figure 3C). Therefore, the conservative alignment of charged residues in voltage-dependent channels certainly fulfils an important role in channel function, but cannot be the molecular basis of channel specific ion selectivity. Segment S2 has been proposed to line the inner channel wall (Noda *et al.*, 1986). It is tempting to speculate that segments S1−S3 are the building units to form the pore of the channel and segments S4−S6 the building units to regulate the gating of the channel by a voltage sensing mechanism. Accordingly, ShD and ShE proteins would not have a voltage sensor. The proposed membrane-spanning sequences are encoded in separate exons. Their sequences (Figure 3B and C) are similar to those of the presumed transmembrane helices of sodium channels (Noda *et al.*, 1984, 1986) and dihydropyridine receptor (Tanabe *et al.*, 1987). The best fit was found for *Sh* S1−S3 segments versus repeat I of vertebrate ion channels, *Sh* S4 segment versus repeat II and *Sh* S5 and S6 versus repeat IV. This finding suggests that the proposed segments which line the inner wall of the channels evolved from a common ancestor via a different route than the proposed segments which mediate voltage-dependent channel gating. The latter could be replaced by units to regulate channel gating by different mechanisms such as a rise in intracelular calcium concentration (Hille, 1984).

Functional implications

It has been proposed on the basis of genetic and electrophysiological experiments that *Sh* is a complex of closely linked functions (Tanouye and Ferrus, 1985). The molecular analysis of *Sh* now demonstrates that it contains a very large transcription unit of >110 kb which can be expressed into a family of different, but functionally related proteins by multiple differential as well as alternative splicing mechanisms. Additionally, alternative primary transcripts could be generated during development and in different

tissues or subset of cells. The protein sequences of sodium channels (Noda *et al.*, 1984, 1986) and dihydropyridine receptor (Tanabe *et al.*, 1987) contain four homologous internal repeats (Figure 3C). The V region of *Sh* encodes a family of products that contain only one such repeat. Given the structural homologies between ion channels (Figure 3B and C) it is conceivable that multimeric aggregates of four Sh products would from an I_A channel.

Gene dosage experiments (Tanouye *et al.*, 1981; Salkoff and Wyman, 1983) had shown that the viable *Sh* mutations belong to the antimorph type. Genotypes in which several doses of the normal gene are combined with a single dose of the mutated *Sh* locus, show the mutant phenotype albeit with reduced severity. A simple interpretation of this type of mutation would suggest that the mutation causes an abnormal gene product with a novel biological activity that interferes with the normal function. However, when the biological function under study is brought about by multimeric structures, other interpretations are also possible (Herskowitz, 1987). The mutation could abolish the expression of a structural product but, it this product were to be assembled with others to form a multimere, the resulting multimere deprived of this component could operate with an antimorphic behaviour. It should be realized that the gene dosage experiment is a functional, as opposed to a structural, test for the biological activity in question.

Our proposition that this family of products forms homo- or heteromultimeres offers a molecular basis for the diversity of I_A-channels as well as the antimorphic nature of *Sh* mutations. This does not exclude the alternative formation of multimeres with components coded in other parts of the *Sh* complex or even outside of it (Wu and Haugland, 1985). In all probability, different multimeres would perform different biological activities.

Materials and methods

Libraries

Our genomic *Drosophila* Canton S DNA library was prepared and kindly provided by Maniatis *et al.* (1978). The other library was constructed from partial *Sau*3a digests of embryonic OregonR DNA inserted into EMBL4 vector (Murray, 1983). This library was kindly provided by H.Jäckle (Tübingen, FRG). gt10 and gt11 cDNA libraries were a gift of T.Kornberg (UCSF, USA) and of P.Salvaterra (City of Hope, USA). Restriction fragments of recombinant phages were subcloned into pAT153, M13mp18/19 or Bluescript by standard cloning techniques (Maniatis *et al.*, 1982). Recombinant DNA was propagated in ER1 host–vector system under L1 containment conditions, as defined in the guidelines of the Federal German Government for recombinant DNA research.

Isolation and labelling of DNA

Recombinant DNA was isolated according to Maniatis *et al.* (1982). Isolation and characterization of Shδ cDNA was described previously (Baumann *et al.*, 1987). cDNA clones were isolated either from a cDNA library made from head poly(A)$^+$ RNA (Shα), or from pupae poly(A)$^+$ RNA (Shβ) or from 1st/2nd instar larvae poly(A)$^+$ RNA (Shε) of *Drosophila melanogaster*. Libraries (300–800 000 phages) were screened with Shδ cDNA probes by the method of Benton and Davies (1977). Bacterial colonies were screened according to Grunstein *et al.* (1975). Hybridizations were carried out with DNA radioactivity labelled by nick translation with [α-^{32}P]dATP (800 Ci/mmol). Five Shα cDNA clones were isolated per 2×10^5 phages, two Shβ cDNA clones per 8×10^3 phages and one Shε per 2×10^5 phages.

Restriction maps and sequencing

Restriction maps were derived by a combination of complete, single and double digests followed by gel electrophoresis of the resulting fragments on 0.7% agarose gels (Maniatis *et al.*, 1982). The colinearity of the restriction map and the genomic DNA was verified by hybridization to whole genome Southern blots. DNA sequencing was done using the Sanger protocol (Sanger *et al.*, 1977). Both strands were sequenced after subcloning cDNA inserts or genomic DNA into M13mp18/19 phages (Yanisch-Perron *et al.*, 1985) or Bluescript.

In situ hybridizations

10 μm sections from paraffin embedded Canton-S individuals or aneuploid males Df(1) B55^D/W32^P (Tanouye *et al.*, 1981) were obtained. [^{35}S]-UTP-labelled riboprobes were obtained following the procedure described in Romani *et al.* (1987). The antisense-strand RNA probes used in these *in situ* hybridizations were obtained from a T7 promoted transcript of Shδ cDNA cloned into the *Eco*RI site of pGEM-1 vector. Controls were carried out with sense-strand RNA probes from an SP6 promoted transcript from the same construct. These control hybridizations (data not shown) indicated that results obtained with the antisense probe were specific.

Acknowledgements

We are thankful to Dr T.Karsch for help with computer analysis and to Drs J.Modolell's and S.Campuzano's hospitality and teaching of the *in situ* technique. This work was supported by the Deutsche Forschungsgemeinschaft (O.Pongs) and by grants from CAYCIT and R.Pascual Fund (A.Ferrus). We are especially indebted to the Volkswagenstiftung for its support of this collaboration. A German–Spanish exchange program 'acciones integradas' provided travel funds.

References

Baumann,A., Krah-Jentgens,I., Müller,R., Müller-Holtkamp,F., Seidel,R., Kecskemethy,N., Casal,J., Ferrus,A. and Pongs,O. (1987) *EMBO J.*, **6**, 3419–3429.

Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.

Catterall,W.A.T. (1985) *Neuroscience*, **8**, 39–41.

Cavener,D.R.N. (1987) *Nucleic Acids Res.*, **15**, 1353–1358.

Connor,J.A. and Stevens,C.F.J. (1971) *Physiology (Lond.)*, **213**, 31–53.

Grunstein,M. and Hogness,D.S. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 3961–3965.

Halliwell,J.C., Othmann,I.B., Pelchen-Matthews,A. and Dolly,J.O. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 493–497.

Herskowitz,I. (1987) *Nature*, **329**, 219–222.

Hille,B. (1984) *Ionic Channels of Excitable Membranes*. Sinauer, Sunderland, MA.

Hubbard,S.C. and Ivatt,R.J.A. (1981) *Rev. Biochem.*, **50**, 555–583.

Josephson,I.R., Sanchez-Chapula,J. and Brown,A.M. (1984) *Circ. Res.*, **54**, 157–162.

Kamb,A., Iverson,L.E. and Tanouye,M.A. (1987) *Cell*, **50**, 405–413.

Koto,M., Tanouye,M.A., Ferrus,A., Thomas,J.B. and Wyman,R.J. (1981) *Brain Res.*, **221**, 213–217.

Krebs,E.G. and Beavo,J.A. (1979) *Annu. Rev. Biochem.*, **48**, 923–959.

Ryte,J. and Doolittle,D.F. (1982) *J. Mol. Biol.*, **152**, 105–132.

Maniatis,T., Hardison,R.C., Lay,E., Lauer,J., O'Connel,C., Ouon,D., Sim,G.K. and Efstratiadis,A. (1978) *Cell*, **15**, 687–701.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Murray,N.E. (1983) In Hendrix,R.W., Roberts,J.W., Stahl,F. and Weisberg,R.A. (eds), *Lamba II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Noda,M., Shimizu,S., Tanabe,T., Takai,T., Kayano,T., Ikeda,T., Takahashi,H., Nakayama,H., Kanaoka,Y., Minamino,N., Kangawa,K., Matsuo,H., Raftery,M.A., Hirose,T., Inayama,S., Hayashida,H., Miyata,T. and Numa,S. (1984) *Nature*, **312**, 121–127.

Noda,M., Ikeda,T., Kayano,T., Suzuki,H., Takeshima,H., Kurasaki,M., Takasashi,H. and Numa,S. (1986) *Nature*, **320**, 188–192.

Papazian,D.M., Schwarz,T.L., Tempel,B.L., Nung Jan,Y. and Jan,L.Y. (1987) *Science*, **237**, 749–753.

Ragowski,M.A. (1985) *Trends Neurosci.*, **8**, 214–219.

Romani,S., Campuzano,S. and Modolell,J. (1987) *EMBO J.*, **6**, 2085–2092.

Salkoff,L. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **48**, 221–231.

Salkoff,L. and Wyman,R. (1981) *Nature*, **293**, 228–230.

Salkoff,L. and Wyman,R.J. (1983) *Trends Neurosci.*, **6**, 128–133.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Solc,C.K., Zagotta,W.N. and Aldrich,R.W. (1987) *Science*, **236**, 1094–1098.

Tanabe,T., Takeshima,H., Mikami,A., Flockerzi,V., Takahashi,H., Kan-

gawa,K., Kojima,M., Matsuo,H., Hirose,T. and Numa,S. (1987) *Nature*, **328**, 313–318.
Tanouye,M.A. and Ferrus,A. (1985) *J. Neurogenet.*, **2**, 253–271.
Tanouye,M.A., Ferrus,A. and Fujita,S.C. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6548–6552.
Tanouye,M.A., Kamb,C.A. and Inverson,L. (1986) *Annu. Rev. Neurosci.*, **9**, 225–276.
Tempel,B.L., Papazian,D.M., Schwarz,T.L., Nung Jan,Y. and Jan,L.Y. (1987) *Science*, **237**, 770–775.
Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55–62.
Wu,C.F. and Haugland,F.N. (1985) *J. Neurosci.*, **5**, 2626–2640.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on December 30, 1987; revised on January 26, 1988

Note added in proof

These sequence data will appear in the EMBL/GenBank/DDBJ/Nucleotide Sequence Databases under the accession number Y00847. Similar sequence data were recently reported by Schwartz *et al.* (1988) *Nature*, **331**, 137–142. Their data shows a longer open reading frame for ShB protein due to a sequencing error. The relevant sequence of Schwartz *et al.* — CTAGAGCGCTG — should correctly be CTAGAGCGGCTG (see position −31 to −20 of Shβ in Figure 2A).