The EMBO Journal vol.7 no.3 pp.611 - 618, 1988

Conservation of neural nicotinic acetylcholine receptors from *Drosophila* to vertebrate central nervous systems

Blaise Bossy, Marc Ballivet[1] and Pierre Spierer

Department of Molecular Biology and [1]Department of Biochemistry, University of Geneva, 30 quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland

Communicated by A.Tissières

Nicotinic acetylcholine receptors (nAChR) are found both in vertebrate and insect central nervous systems. We have isolated a *Drosophila* gene by crosshybridization with a vertebrate probe. Structural conservation of domains of the deduced protein and of intron/exon boundaries indicate that the *Drosophila* gene encodes an nAChR α-like subunit (ALS). That the *Drosophila* gene product most resembles the neuronal set of vertebrate nAChRs α-subunits is also indicated by the failure of an ALS-β-galactosidase fusion protein to bind α-bungarotoxin on blots in contrast to vertebrate endplate α-subunit constructions. The ALS encoding gene exceeds 54 kb in length and the transcript has a very long and unusual 5′ leader. As we found previously for a gene whose product is also involved in cholinergic synapses, acetylcholinesterase, the leader encodes short open reading frames, which might be involved in translation control. We also note the presence of *opa* repeats in the gene, as has been found for various *Drosophila* genes expressed in the nervous system.

Key words: acetylcholine receptors/α-bungarotoxin/*Drosophila melanogaster*/protein homology

Introduction

Physiological responses to a limited set of neurotransmitters can be varied by increasing the diversity of receptors for a given neurotransmitter. Acetylcholine (ACh) is the neurotransmitter that best illustrates this situation. It is recognized by two unrelated receptor families, the nicotinic and muscarinic ACh receptors, that coexist in vertebrates as well as in insects (Dudai, 1980). In vertebrates, the number of known members of both receptor families has recently increased (Goldman *et al.*, 1987; Nef *et al.*, 1988; Bonner *et al.*, 1987).

Nicotinic acetylcholine receptors (nAChRs) have been shown to constitute a receptor family whose members are expressed in different regions of the vertebrate central and peripheral nervous systems, including different mammalian brain areas (Goldman *et al.*, 1987; Whiting *et al.*, 1987). According to pharmacological and structural criteria (Martin, 1986), this receptor family can be divided into two groups. The members of the first group are specifically antagonized by the neurotoxin α-bungarotoxin (α-BTX), whereas in the second group of nAChRs α-BTX does not block function. This pharmacological criterion has been found to be determined by a toxin-binding amino acid sequence overlapping the ACh binding site located close to a conserved pair of vicinal cysteines found in all the nAChRs α-subunits that have been analyzed (Wilson *et al.*, 1985; Neumann *et al.*, 1986; Barkas *et al.*, 1987).

The first group includes the endplate nAChR, the electroplax receptor and possibly a brain α-BTX binding component (Conti-Tronconi *et al.*, 1985). Endplate and electroplax receptors are complexes built of four different subunits in the stoichiometry $\alpha_2\beta\gamma\delta$ (reviewed by Changeux, 1984; Hucho, 1986). Members of the second group are located in the central nervous system and in sympathetic ganglia (Clarke, 1985; Patrick and Stallcup, 1977). Three different receptor subtypes belonging to this second group have been described so far (Boulter *et al.*, 1986; Goldman *et al.*, 1987; Nef *et al.*, 1988). They may assemble from only two different subunits, one of which (the non-α) may be common to several receptor subtypes (Boulter *et al.*, 1987).

nAChRs have been analyzed cytologically, pharmacologically and biochemically in several insect species including *Drosophila* (Hall and Teng, 1975; Schmidt-Nielsen, 1977; Dudai, 1977; Rudloff *et al.*, 1978). These studies suggested the presence of nAChRs with properties similar to those of the vertebrate endplate receptor and including an α-BTX binding component. Purification of nAChR from locust (*Locusta migratoria*) ganglia by α-BTX affinity (Breer *et al.*, 1985), and subsequent receptor reconstitution in planar bilayer (Hanke and Breer, 1986) indicated that in this insect species, at least, one form of the nAChR is an oligomer composed of identical subunits, a structure predicted for an ancestral AChR protein. More recently, *Drosophila* cDNA clones encoding a protein (named ARD) homologous to the vertebrate nAChR non-α subunits were isolated and analyzed (Hermans-Borgmeyer *et al.*, 1986). We describe here the isolation and analysis of an insect nAChR α-subunit gene and of the corresponding cDNA clones. Interestingly, its deduced amino acid sequence has a closer homology to the vertebrate neural nAChR α-subunits than to the neuromuscular α-subunits, as is also the case for the *Drosophila* non-α subunit (Hermans-Borgmeyer *et al.*, 1986). Furthermore, a *Drosophila* nAChR-β-galactosidase (*Escherichia coli*) hybrid protein did not bind α-BTX *in vitro*. Thus the protein product of this *Drosophila* gene has properties analogous to those of the vertebrate neural nAChR α-subunits that do not bind α-BTX. The structure of the ALS gene differs from the structure of its vertebrate counterparts. First, it spans more than 50 kb of DNA and its major transcription product is 10.5 kb long. In addition, one intron conserved in all the vertebrate nAChR subunits is not found in this gene, whereas four remaining introns occur at precisely the same sites in vertebrates and in *Drosophila*. Three repetitive elements of the *opa* family are present in the 5′ and 3′ sequences flanking the main open reading frame (ORF) on the cDNAs. Such elements are also found in other *Drosophila* genes expressed in the nervous system (Wharton *et al.*, 1985).

Results

Genomic and cDNA clones isolation

A *Drosophila* genomic library was screened at low stringency with a fragment of the chick neuronal nAChR α2 gene (Nef *et al.*, 1988) encoding amino acids 96–195 of the α2 subunit. Genomic clone φ4 (Figure 1a) was the original isolate. The nucleotide sequence around a *Bam*HI restriction site (co-ordinate zero, Figure 1b) located within the crosshybridizing region confirmed the existence of an ORF having 66% conserved nucleotides with the probe.

Additional vertebrate receptor probes allowed the identification of some neighbouring exons on the same *Drosophila* genomic clone (data not shown). Restriction fragments containing these *Drosophila* exons were used to screen cDNA libraries (a gift of L.Kauvar). Three different cDNA clones were selected from a first and second instar larval library and sequenced. One clone (no. 3, Figure 1d) overlaps the other two (no. 8 and no. 16), which are separated by just one nucleotide. The combined nucleotide sequence from these three clones defines a single long ORF of 1701 nt (Figure 2) bracketed by 1282 nt and 514 nt in 5′ and 3′ respectively.

cDNA deduced amino acids sequence

The main ORF encodes 567 amino acids presenting structural domains homologous to those of the vertebrate nAChR subunits. The 20 first residues have the expected features of a putative signal peptide and the N terminus of the mature protein is probably Asn 1 at nucleotide position 1346 (Figure 2 and 3) as the residues at −1 and −3 conform to the rules of Von Heijne (1983). The first 218 residues of the mature protein are homologous to the extracellular domain of the vertebrate nAChRs with two conserved Cys at positions 128 and 142, and the Cys doublet identifying all the α subunits. This Cys doublet is found at position 201–202 instead of the expected 192–193 co-ordinates, because of a nine-residue insertion in the small variable region between residues 158 and 174. We therefore call this *Drosophila* amino acid sequence ALS (α-like sequence). Like the vertebrate neuronal α2, α3 and α4 subunits, ALS contains two potential glycosylation sites (Asn 24 and 212) instead of one in vertebrate endplate α-subunits. Moreover, the Asn 24 site is conserved in the ARD and vertebrate neuronal subunits.

Barkas *et al.* (1987) defined the interval between residues 160 and 216 as containing the α-BTX binding site of the mouse α-subunit. This sequence corresponds to co-ordinates 167–224 (Figure 3) in the ALS protein. Interestingly, the highest similarity in this interval is found between ALS and rat neuronal α3 (25 identical residues out of 57). Homology between ALS and the vertebrate endplate α subunits at the α-BTX binding site is lower (e.g. 19 identical amino acids out of 57 in the mouse muscle α subunit). Proper alignment of vertebrate neuronal with endplate α subunits in the region of the Cys doublet requires introducing a gap of one amino acid, and so does ALS.

The next structural domain contains three stretches of hydrophobic amino acids that define three putative transmembrane regions according to Kyte and Doolittle (1982), separated by stop transfer residues. This region (co-ordinates 219–308, Figure 3) probably constitutes the transmembrane segment of the cation channel in co-operative assembly with the homologous regions provided by the other subunits in the complex (Giraudat *et al.*, 1986). This region is highly conserved between insects and vertebrates with 58% identical residues between ALS and mouse muscle or *Torpedo* electroplax α-subunits, 70% with rat α3 or α4 neural subunits and up to 90% similarity when conservative amino acid exchanges are considered.

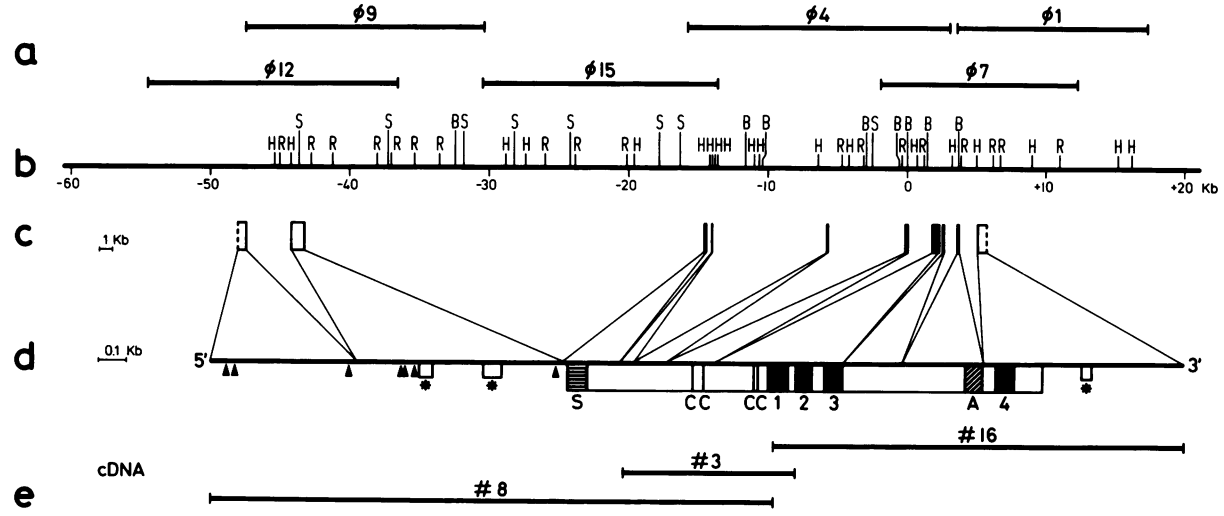

Fig. 1. Correspondence between the ALS gene and its cDNA products. (**a**) Chromosomal walk in the 96A region. The horizontal bars indicate the overlapping chromosomal fragments isolated from a *Drosophila* genomic library. (**b**) Restriction map of the cloned genomic interval. R, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; S, *Sal*I. (**c**) Exons localization. Filled boxes indicate the ALS encoding exons and open boxes indicate exons that code for ALS 5′ and 3′ flanking sequences. The dashed lines specify that the 5′- and 3′- most exons might extend further. The scale is identical for (1), (b), and (c), and is indicated in kb. (**d** and **e**) cDNA map. The horizontal line corresponds to the sequence of 3498 nt reconstituted from the three cDNA clones (nos. 8, 3, 16) mentioned in (e). The large box symbolizes the ALS-encoding region with the following landmarks: S, signal peptide; C, Cys at amino acids positions 132, 148, 201, 202; A, amphipathic region; 1, 2, 3, 4, the four hydrophobic regions. Starred boxes indicate the *opa* repetitive elements. Triangles localize the additional AUGs in the 5′ leader region. The scale is the same for (d) and (e). It has been magnified 20× relative to (a), (b), and (c). The correspondence between the exons in (c) and the cDNA sequence in (d) is visualized by oblique lines.

The next peptidic interval extends between positions 309–444. The sequence that is immediately adjacent to the third transmembrane region is well conserved over about 30 residues and then diverges considerably. This interval is generally thought to constitute the cytoplasmic domain of the receptor (Claudio *et al.*, 1983; Finer-Moore and Stroud, 1984; Ratnam *et al.*, 1986). Its length is variable, but the vertebrate α3 and α4 subunits are characterized by a cytoplasmic domain at least 30 amino acids longer than in the vertebrate endplate α subunits, an observation that also applies to ALS. ALS does not contain potential Ser phosphorylation sites in this interval.

The last region corresponds to the putative amphipathic α-helix (Finer-Moore and Stroud, 1984; Guy, 1984), followed by a fourth hydrophobic domain and the C-terminus of the subunit. Both the primary structure and the amphipathic nature of an α- helix are conserved between ALS or ARD and vertebrate nAChR subunits. This conservation between insects and vertebrates suggests a structural role for this domain in nAChRs, in as much as the amphipathic α helix is not conserved in vertebrate glycine receptor and $GABA_A$ receptor subunits that belong to the same receptor super-family (Grenningloh *et al.*, 1987; Schofield *et al.*, 1987). After a fourth hydrophobic region that is conserved in all nAChR subunits, ALS ends by a particularly long (residues 511–546) and lysine-rich C-terminal domain.

Forty to forty-four percent of the *Drosophila* ALS residues from the putative mature protein are conserved in vertebrate neuronal α subunits, 42% in the *Drosophila* ARD subunit, 35.5% in mouse α and 34% in electroplax α subunit. The best similarity (44% of ALS residues) has been found in the chick α2 subunit (Nef *et al.*, 1988). In the strychnine-binding subunit of the rat glycine receptor, 20% of the residues are identical to ALS and there is a similar amount of conservative exchanges. This percentage is very close to that obtained with ARD which exhibits the best homology found between an nAChR subunit and this rat glycine receptor subunit. The bovine $GABA_A$ receptor α-subunit is slightly more similar (19% identical residues) to ALS than is the β-subunit (16.5%) of the same receptor.

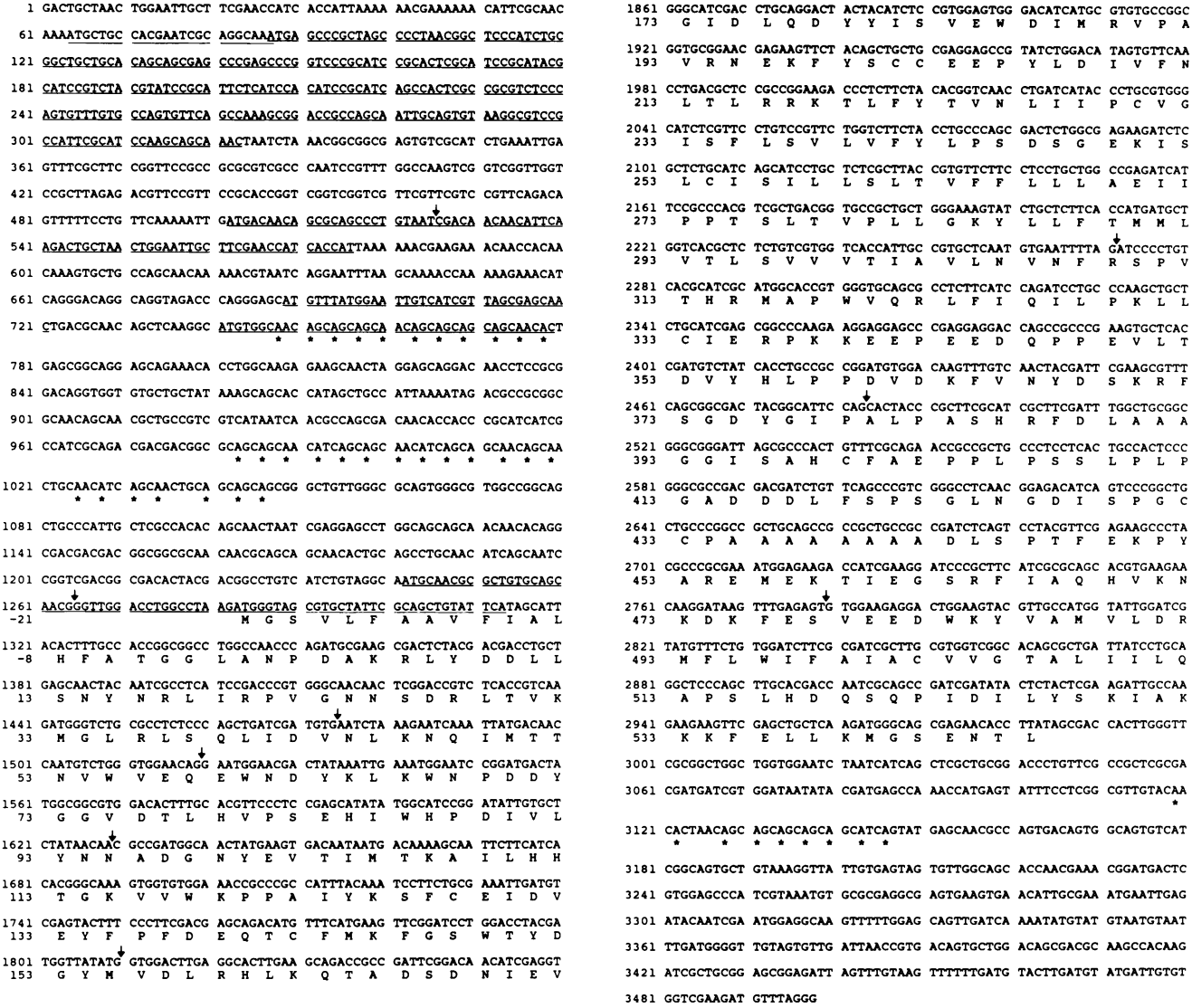

Fig. 2. Nucleotide and deduced amino acid sequence of overlapping cDNA clones 8, 3 and 16. Arrows localize intervening sequences. The small ORFs in the 5′ part of the transcript are underlined. Stars are put below the CAX codons of the *opa* repetitive elements. The cDNA clones nos. 8, 3 and 16, respectively, extend from nucleotides 1–2021, 1492–2093 and 2023–3498.

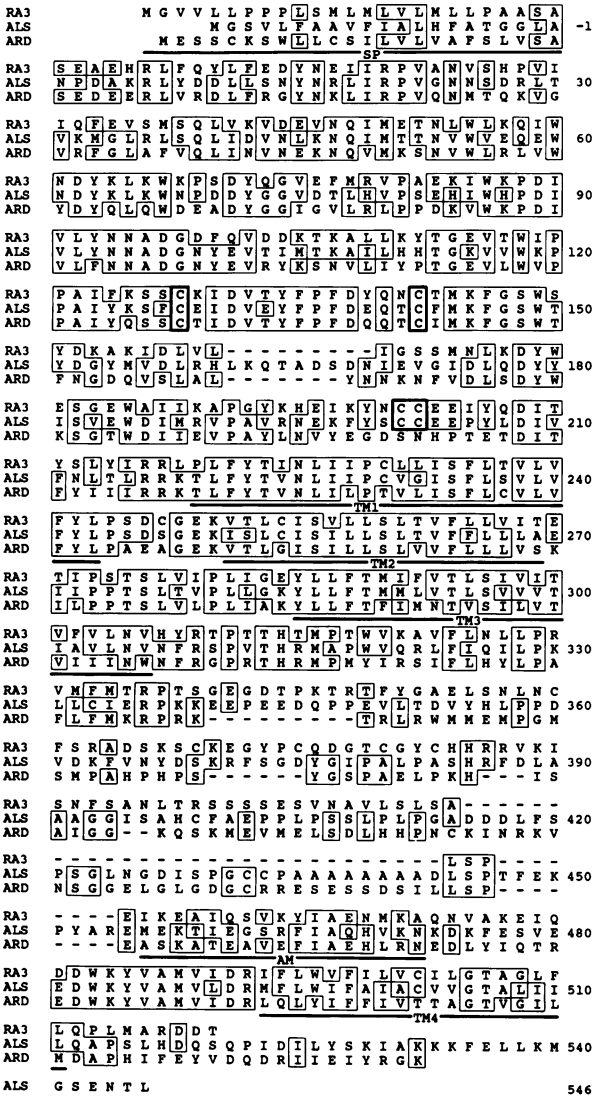

Fig. 3. Comparison of the deduced amino acid sequences of three acetylcholine receptor subunits. The sequences mentioned are: **top**, the rat neuronal α3 subunit (RA3) from the PC12 cell line (Boulter *et al.*, 1986); **middle**, the *Drosophila*, α-like sequence (ALS) deduced from Figure 2; and **bottom** the *Drosophila* non-α (ARD) sequence (Hermans-Borgmeyer *et al.*, 1986). Identical amino acids are boxed and ALS is numbered from its probable mature N terminus. Putative signal peptide (SP), transmembrane (TM 1–4), amphipathic helix (AM) sequences are underlined and bold boxes surround conserved cysteine residues.

A complex 5′ leader sequence

cDNA no. 8 contains one of the longest (1282 nt) 5′ leader ever mentioned in the literature. This leader sequence contains only 15% U and has many GC-rich stretches frequently separated by groups of 2–6 A (Figure 2). A 10 bp purely GC palindrome is present around the *Sac*II restriction site at position 896. Seven AUGs are located in the 5′ leader sequence, and they define putative start codons for six small ORFs (the fifth AUG being internal to the fourth frame) that code for between 8 and 79 amino acids (Figure 2). In addition to the probable start site for the receptor coding sequence, at least two AUG environments (positions 64 and 689 in Figure 2) are compatible with the *Drosophila* consensus sequence proposed by Cavener (1987) for translation initiation. One small ORF codes for 13 amino acids containing 10 adjacent glutamine residues. This polyglutamine pattern corresponds to a repetitive element termed *opa* or M repeat, which exists in 400–500 copies per haploid *Drosophila* genome (Wharton *et al.*, 1985). Two other *opa* repeats are found in the cDNA sequence, one again in the 5′ leader sequence, the other 3′ of the ALS coding region. These latter two repetitive elements consist of CAN triplets (usually CAG or CAA) which are not included in any ORF. They are also partially degenerated with some interspersed modified triplets (1 nt substituted). Moreover, the first and the third *opa* repeats are short, three times shorter than the *Notch opa* repeat (Wharton *et al.*, 1985).

cDNA no. 16 does not contain a poly(A) tail at its 3′ end, nor any polyadenylation consensus signal. It is therefore probable that the 3′ untranslated region of the mRNA extends beyond the 514 nt listed in Figure 2.

Expression of the ALS coding gene through Drosophila development

Poly(A)$^+$ RNA was purified from a series of *Drosophila* developmental stages and analyzed on Northern blots. In order to avoid crosshybridization to the *opa* repeats we used a single-stranded DNA probe which was complementary to the coding region 2023–2580 (5′ end of cDNA no. 16). One main band of 10.5 kb was present in all developmental stages from late embryogenesis to adulthood, but not during the first 2 h of development (Figure 4). A much fainter band of 7.7 kb is present at late embryogenesis and at the limit of detection at early pupal stage. In late embryonic and larval stages a larger and more diffuse band is also apparent. This larger band was also labelled with the actin 5C and *Ace* (acetylcholinesterase) probes used to standardize the transcript size (data not shown). It probably consists of genomic DNA with AT rich sequences that may have copurified with poly(A)$^+$ RNA during the oligo(dT) chromatography. The abundance of the 10.5 kb transcript seems to have reached its maximum at late embryogenesis and to stay relatively stable up to early pupal stage. A 2-fold decrease is seen at the late pupal stage (pupae with red eyes) and is followed by a further 3-fold decrease at adulthood (7 days after eclosion). Rehybridization of the same blot with an actin probe confirmed that equal amounts of poly(A)$^+$ RNA were loaded in each lane (data not shown).

Structure of the ALS coding gene and genomic localization

In order to verify whether the 3 cDNA clones are really transcripts from the same gene (cDNA nos. 16 and 3 only overlap by 70 nt) and whether the long 5′ leader sequence of cDNA no. 8 is not a cloning artifact, corresponding exons were localized on the *Drosophila* genome. Restriction fragments from the 5′ and 3′ ends of clone φ4 were used to select adjacent genomic clones by the walking method (Bender *et al.*, 1983). In parallel, a subclone from cDNA no. 8 containing only the first kilobase of the 5′ leader sequence was used to jump directly to the corresponding exon(s). The genomic fragments cloned by both strategies overlap very slightly (on less than 0.5 kb, see φ9 and φ15 in Figure 1a) and the size of a common *Sal*I fragment was determined on a Southern blot (data not shown). The restriction map spanning the 70 kb of cloned genomic DNA was determined (Figure 1b), and most of the exons localized on

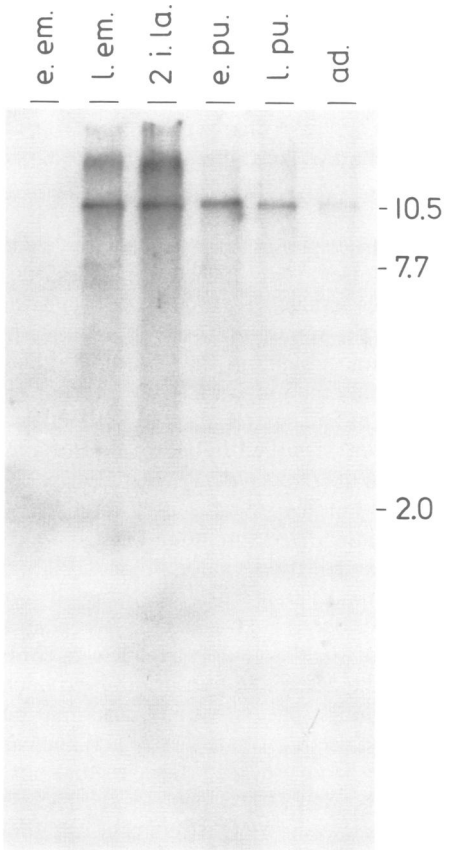

Fig. 4. Northern blot of poly(A)$^+$ RNA from 6 developmental stages (2 μg/lane): e. em., early embryos; l. em., late embryo; 2 i. la., second instar larvae; e. pu., early pupae; l. pu., late pupae; ad., adult flies. The size estimates are given in kb. The probe is a single-stranded cDNA fragment described in Materials and methods.

it were subcloned for sequencing (Figure 1c). Their correspondence to the cDNA sequence is given in Figure 1d and the intron localization is also indicated by arrows in Figure 2. Ten different exons spread over 54 kb were identified. The last exon was sequenced only up to its *opa* repeat. The first and third 5′ exons were only localized on the restriction map. Within the receptor coding exons, only one non-conservative substitution was found, in the C-terminal region at position 517, where a Tyr replaces a His.

The first two exons contain 5′ leader sequences, and they are separated from the eight coding exons by 28 kb of intervening sequence. The exon encoding the putative signal peptide is not split by an intron in contrast to all vertebrate nAChRs genes analyzed so far. On the other hand, four out of seven splice sites between homologous exons occur at precisely the same nucleotide in ALS (after positions 1474, 1519, 1629 and 2779) and its vertebrate equivalents. the splice site after position 1810 is 15 nt more 5′ than its equivalent in the human muscle α subunit. The first splice, which is located in the cytoplasmic domain coding sequence (after position 2271) is in the Arg codon that lies immediately after the third membrane-spanning region, 40 nucleotides more 5′ than in vertebrates. The next splice site, after position 2483, is in an unconserved region that does not allow comparison. The sequences corresponding to the first three membrane spanning regions are part of the same exon, a situation also found in all vertebrate neuronal genes (Nef

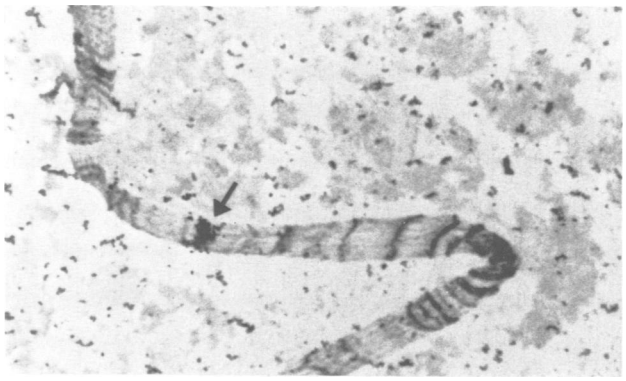

Fig. 5. *In situ* hybridization to polytene chromosomes. The nick-translated phage DNA φ4 probe is seen to hybridize to the 96A region of the 3R chromosome.

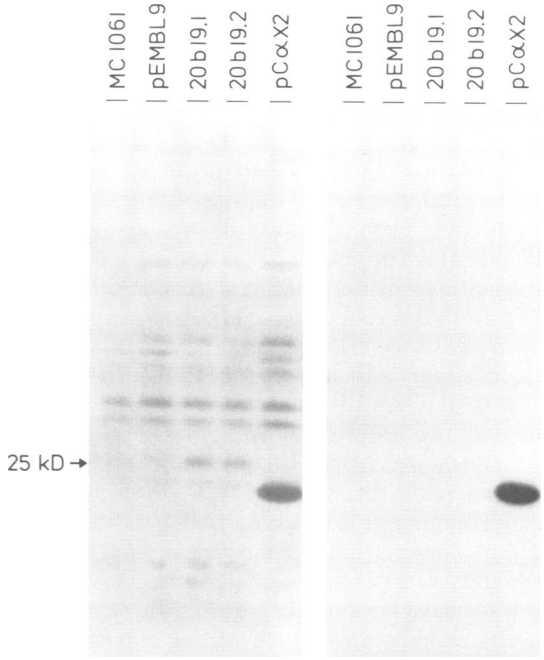

Fig. 6. α-Bungarotoxin binding assay. **Left panel**: protein blot on nitrocellulose filter colored with amido black. MC1061, *E.coli* pEMBL9, same strain transformed with pEMBL9; 20b19 1 and 2, same strain transformed with a hybrid ALS−*lacZ* gene cloned in pEMBL19. The two slots correspond to two independent but identical clones; pCαX2, same strain transformed with a hybrid mouse α-*lacZ* gene cloned in pUC8 (Barkas *et al.*, 1987). **Right panel**: same blot incubated with iodinated α-bungarotoxin and autoradiographed as described in Materials and methods.

et al., 1988) in contrast to the neuromuscular receptor genes, where they are split in two or three exons (Nef *et al.*, 1984).

In order to localize the ALS coding gene cytogenetically, genomic clone φ4 was used to probe the *Drosophila* genome on salivary gland polytene chromosomes. The ALS gene maps on the 3R chromosome at position 96A as shown in Figure 5.

α-BTX does not bind to a hybrid ALS-lac Z protein

A cDNA fragment generated by exonuclease III digestion of cDNA no. 8 was cloned in phase in a *lacZ* expression vector (pEMBL19) in order to create a *Drosophila*−*E.coli*

hybrid protein. This recombinant (clone 20b19) codes for 135 ALS amino acids (positions 91−225, Figure 3) in a sum of 228 residues creating a 25 kd fusion protein. This protein was one of the main proteins constitutively expressed in strain MC1061 (Figure 6). However, it was much less abundant than the chick neuromuscular α-subunit fragment cloned in phase in the recombinant pCαX2 (Barkas *et al.*, 1987) used as positive control. Iodinated α-BTX binds strongly to this positive control and not more than to the nitrocellulose background to the clone 20b19 (Figure 6), even after overexposure (data not shown). Barkas *et al.* (1987) demonstrated that the DNA encoding the 160−216 amino acid fragment of the mouse neuromuscular α subunit, when cloned in phase as a *lac*Z hybrid recombinant was able to bind specifically α-BTX on a nitrocellulose blot. Recently, the equivalent regions from the vertebrate neuronal genes (α2, α3, α4) were shown to be unable to bind α-BTX under similar conditions (Ballivet *et al.*, unpublished). The negative result obtained with the 20b19 clone supports but does not prove that the ALS subunit is unable to bind α-BTX.

Discussion

nAChRs from vertebrate skeletal muscle and *Torpedo* electroplax are cation-selective transmembrane channels that open transiently upon acetylcholine binding to their extracellular domain. The resulting depolarizing ion flow generates a specific response in the postsynaptic cell. Recently, nAChRs from the vertebrate central nervous system have been analyzed at the cDNA level. These neuronal nAChRs have a quaternary structure that probably differs from the $\alpha_2\beta\gamma\delta$ stoichiometry of muscle receptor. They apparently contain only two different subunit types, one of which (α) binds acetylcholine (ACh) (Boulter *et al.*, 1987).

The *Drosophila* subunit sequences described previously (Hermans-Borgmeyer *et al.*, 1986) and in this report exhibit an extensive conservation with the vertebrate neural subunits. Forty to forty-five percent of the *Drosophila* ARD or ALS amino acids have been conserved since the divergence between vertebrates and arthropods. Some sequence elements are particularly useful when attempting to classify the different subunits into subtypes. Such is the case for the putative ALS glycosylation site on Asn 24, a site that does not exist in non-neuronal vertebrate nicotinic receptors. In addition to the Cys at position 128 and 142 found in all nAChR subunits, the ALS Cys doublet at positions 201−202 corresponds in location and environment to the vertebrate 192−193 Cys doublet involved in ACh binding, a landmark identifying α subunits. We suggest that ALS is the α-like component of a neuronal AChR that does not bind α-BTX because the sequence extending between residues 169−224 (ALS numbering) is signficantly more conserved between ALS and rat neural α3 (45% similarity) than between ALS and all known endplate α subunits (33% similarity). This region constitutes the core element of the toxin-binding site in endplate α subunits, whereas reconstituted receptors incorporating rat α3 are not blocked by α-BTX (Boulter *et al.*, 1987). Within the same interval, another landmark resides in an additional residue (Pro) at position 194, adjacent to the Cys doublet in all the α-BTX binding α subunits. This residue is absent in the neural α2, α3, α4 and ALS subunits. Additionally, we constructed and expressed a *Drosophila*−*E.coli* hybrid protein containing the ALS sequence extending from residues 91−225. This polypeptide does not bind α-BTX on blots, under conditions where similar constructions from vertebrate endplate α subunits do specifically bind this ACh antagonist. Thus the evidence obtained so far suggests that ALS is a neuronal α subunit that does not bind α-BTX. It can therefore be considered as a *Drosophila* equivalent of the vertebrate α2, α3 or α4 subunits.

The ALS transcript and the ALS gene have a number of noteworthy features. First, the main transcript (10.5 kb) as well as the minor species (7.7 kb) detected on Northern blots greatly exceed the minimum of about 1.7 kb necessary to encode a nicotinic α subunit. Calf muscle δ-subunit mRNA (6 kb) is probably the longest mRNA known to code for a vertebrate nAChR subunit (Noda *et al.*, 1983a). A transcript of 7 kb has been described by Hermans-Borgmeyer *et al.* (1986) as one of the *Drosophila* ARD transcription products. The strychnine-binding subunit of the rat glycine receptor, which has a primary structure homologous to the nAChRs, also has a long major transcription product (9 kb) with some less abundant smaller transcripts (Grenningloh *et al.*, 1987). In the latter case, only 1.3 kb is required as coding sequence. The existence of such large poly(A)$^+$ RNAs corresponding to the bulk of the detectable transcription products of these genes is intriguing. The possibility that they may be incompletely processed, given their size and nuclear localization, remains to be investigated. If the mature mRNAs have a smaller size they have to be much less abundant than their precursors. That would suggest either a rate-limiting processing step or a labile fully processed mRNA.

In the 3.5 kb of ALS cDNA, 1.3 kb are located 5′ to the coding sequence. This leader sequence, although incomplete, is probably one of the longest ever described. Three copies of repetitive elements belonging to the *opa* family are present in the non-coding sequences of the ALS transcript. This family of $(CAX)_n$ repeats are dispersed in the *Drosophila* genome and copies were found in the *bithorax* and *Antennapedia* complexes and in the *engrailed* locus, all three involved in pattern formation, and in the two neurogenic loci *Notch* and *mastermind* (Wharton *et al.*, 1985; Weigel *et al.*, 1987). It may be noticed that the long 3′ non-coding sequence from calf muscle α-subunit transcripts also contains repetitive elements that are not related to *opa* repeats but to a repetitive sequence family interspersed in the bovine genome (Noda *et al.*, 1983b).

Furthermore, the ALS cDNA 5′ leader sequence is characterized by seven ATG triplets, two of them at least being in a favorable context for translation initiation in *Drosophila* according to statistical data (Cavener *et al.*, 1987). These leader peptides evoke the procaryotic upstream, ORFs involved in the attenuation process. In eucaryotes, 5′ leader AUGs and small upstream ORFs are implicated in the translation regulation of the yeast GCN4 gene by two *trans*-acting genes to modulate GCN4 translation in response to starvation (reviewed by Fink, 1986). The scanning hypothesis for translation initiation (Kozak, 1978) is not necessarily incompatible with mRNAs containing additional upstream AUG codons, and a termination-reinitiation process in the 5′ leader sequence has been described in mammalian cells (Peabody and Berg, 1986). Translation initiation can occur at an internal AUG if the translation starting upstream stops before or just after the internal initiation codon (Peabody *et al.*, 1986). In as much as only 5−10%

of eucaryotic mRNAs contain additional upstream AUGs (Kozak, 1983) and as both the transcripts that code for the *Drosophila Ace* and two putative nAChR subunits (ALS and ARD) are in this situation, this similarity in their 5′ leader sequences might not be fortuitous. Because acetylcholinesterase, nAChR α and non-α subunits have to participate together in cholinergic transmission in some central nervous system synapses, it is quite probable that the *Ace*, ALS and ARD genes are simultaneously expressed in some neurons. It is tempting to imagine a translation regulation of their mRNAs for a co-ordinated expression of their protein products.

The genomic structure of the ALS gene has also been studied. The three overlapping ALS cDNA clones are derived from 10 exons spread over 54 kb of genomic DNA. The first two exons contain most of the 5′ leader sequence and are separated by an intron of 28 kb from the 8 exons that encode the ALS subunit. These 8 coding exons map within a 21 kb chromosomal interval. The ALS gene must extend beyond this vast interval since we can only account for exons within the 3.5 kb of cloned cDNA. Four out of seven introns in the ALS coding sequence have splice site localizations that are perfectly conserved throughout evolution since they occur at precisely the same nucleotide in vertebrates. The three remaining introns are located in two poorly conserved coding regions which vary in length depending upon the subunit type. One intron which splits the signal peptide coding sequence into two exons in all known veretebrate nAChR genes is absent from the ALS gene. No introns separate the first three ALS transmembrane coding sequences, a situation that also occurs in vertebrate neuronal nAChR genes (Nef *et al.*, 1988), whereas in endplate nAChR genes introns do separate the coding sequences for transmembrane regions.

If we assume that the receptor incorporating the ALS subunit does not bind α-BTX, one or several additional α subunits should exist in the *Drosophila* nervous system. Previous genetic work on *Drosophila* nicotine-resistant mutants localized a nictotine-resistant locus on the X-chromosome (Hall *et al.*, 1978; Hall, 1980). Iodinated α-BTX has been described to bind to most of the *Drosophila* adult brain, to larval brain hemispheres and to the larval ventral ganglion (Hall and Teng, 1975; Rudolf *et al.*, 1978. Dudai, 1980). This extensive presence of α-BTX binding sites in the *Drosophila* central nervous system (CNS) suggests that if different nAChRs are simultaneously present in the CNS, one of them being the α-BTX binding component, they might be expressed in the same neurons. Such a situation may also occur in the mammalian CNS, where the α3 and α4 transcripts are both present in the substantia nigra pars compacta and in the medial habenula. These two brain areas contain both presynaptic and postsynaptic nicotinic receptors (Lichtensteiger *et al.*, 1982; Sakurai *et al.*, 1982). One possibility which has been proposed (Goldman *et al.*, 1987) is that the α3 and α4 subunits each participate separately in one of these nicotinic receptors. It will be interesting to discover whether the apparently simpler *Drosophila* CNS exhibits a neural nicotinic receptor diversity that is more restricted than in the vertebrate CNS or if the analogy mentioned above can be extended to other α-like subunits, possibly defining further receptor subtypes.

The molecular mapping of the ALS gene on the *Drosophila* genome described here will help develop genetic approaches to select new nictotinic receptor mutants and to map chromosomal rearrangements in this locus. Such mutants may contribute to a better understanding of the nAChR structure, functions and developmental regulation during synaptogenesis.

Materials and methods

Screening of genomic and cDNA libraries

A *Drosophila* Canton S genomic library (Maniatis, 1978) was screened at low stringency according to Benton and Davis (1977). The hybridization was performed in 4× SSCP (480 mM NaCl, 60 mM sodium citrate, 52 mM NaH_2PO_4, 1 mM Na_2EDTA, pH 6.5), 0.25% SDS, 4× Denhardt, 0.1 mg/ml denatured salmon sperm DNA at 55°C. Filters were washed twice at room temperature with 2× SSC, 0.1% SDS. The probe was a nick-translated cDNA fragment (pBB64) from the chick neuronal α2 subunit (Nef *et al.*, 1988). The further genomic and cDNA libraries screenings were performed at high stringency. In this case, prehybridization and hybridization were performed at 65°C, and the final washes were 0.1× SSC, 0.1% SDS, at 55°C. cDNA libraries from Oregon R strains were generously provided by L.Kauvar (Poole *et al.*, 1985) and the three cDNA clones we describe were obtained from the 1−2 instar larvae G2 pool.

DNA sequence analysis

Purified cDNAs were cloned in pEMBL18 and deletions were generated with exonuclease III according to Henikoff (1984). Deleted subclones were selected after plasmid minipreparation, restriction digestions and agarose gel electrophoresis. pEMBL single-stranded DNAs were prepared according to Dente *et al.* (1983) with either f1 or M13K07 (a gift of F.Karch) as helpers. The dideoxy nucleotide sequencing technique of Sanger (1977) was used according to Amersham International (1984). Both cDNA strands were sequenced, except for the first 25 5′ nt. The programs of Staden (1982) were used to compile sequence data from overlapping or complementary clones.

RNA purification and Northern blots

Drosophila Oregon R wild type eggs were collected after a 2 h laying period, following a 1 h prelaying lag. Eggs were then incubated at 20°C and embryos were collected at 25 h, larvae at the first or second instar stages, pupae at early and late pupation (red eyed pupae), adults 7 days after their eclosion. *Drosophila* from these successive stages were then frozen in liquid nitrogen and homogenized in a mortar. Poly(A)$^+$ RNA was purified and Northern blots were performed as detailed by Hall *et al.* (1983). The probe was a single-stranded cDNA fragment complementary to nucleotides 2023−2580 (Figure 2).

Polytene chromosome in situ hybridization

Salivary glands from *Drosophila* third instar larvae were squashed and their polytene chromosomes were hybridized with a nick-translated probe labelled with α[^{35}S]dATP, according to Spierer *et al.* (1983).

α-Bungarotoxin binding assay

Expression clone 20b19 is derived from an exonuclease III generated deletion of cDNA no. 8 bracketed by *Hin*dIII and *Eco*RI restriction sites. This cDNA fragment was recloned in phase in the β-galactosidase gene fragment of the plasmid vector pEMBL 19 (Dente *et al.*, 1983) and resequenced. Transformed *E.coli* strain MC1061 was grown at 37°C to an A_{600} of 1. After centrifugation, the cells were suspended in a lysis buffer containing 5.5% glycerol, 0.7 M β-mercaptoethanol, 1.65% SDS, 30 mM Tris/HCl pH 6.8 and 0.05% bromophenol blue. The sample was heated to 100°C for 5 min before loading on a 12.5% SDS−polyacrylamide gel. After gel separation, proteins were electroblotted to a nitrocellulose membrane (BA85; Schleicher and Schuell). The membrane was then washed in PBS, Tween 0.05%, 0.1 mM PMSF for 1 h at room temperature. Iodinated α-BTX was added at 2.5 nM and incubated at room temperature for 160 min in a sealed bag. Finally, the membrane was washed in 20 ml of PBS, 0.05% Tween for 20 min at room temperature and autoradiographed. Proteins were visualized on the membrane by staining in 0.1% amido black, 10% acetic acid, 45% methanol for 5 min, followed by a wash in 40% methanol, 2% acetic acid.

Acknowledgements

We thank A.Tissières for his continuing interest, M.Vollenweider, C.Fumey, I.Dunn for expert technical assistance and O.Jenni and Y.DeLotto for the

Figures. We are also grateful to L.Kauvar for providing samples of cDNA libraries. This work was supported by the Swiss National Science Foundation and the State of Geneva.

References

Amersham International plc (1984) *M13 cloning and sequencing handbook*. Amersham International plc, Amersham, UK.

Barkas,T., Mauron,A., Roth,B., Alliod,C., Tzartos,S.J. and Ballivet,M. (1987) *Science*, **235**, 77–80.

Bender,W., Spierer,P. and Hogness,D.S. (1983) *J. Mol. Biol.*, **168**, 17–33.

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

Bonner,T.I., Buckley,N.J., Young,A.C. and Brann,M.R. (1987) *Science*, **237**, 527–532.

Boulter,J., Evans,K., Goldman,D., Martin,G., Treco,D., Heinemann,S. and Patrick,J. (1986) *Nature*, **319**, 368–374.

Boulter,J., Connolly,J., Deneris,E., Goldman,D., Heinemann,S. and Patrick,J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7763–7767.

Breer,H. (1981) *Neurochem. Int.*, **3**, 43–52.

Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.

Changeux,J.P., Devillers-Thiéry,A. and Chemouilli,P. (1984) *Science*, **225**, 1335–1345.

Clarke,B.P.S., Schwartz,R.D., Paul,S.M., Pert,C.B. and Pert,A. (1985) *J. Neurosci.*, **5**, 1307–1315.

Claudio,T., Ballivet,M., Patrick,J. and Heinemann,S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1111–1115.

Conti-Tronconi,B.M., Dunn,S.M.J., Barnard,E.A., Dolly,J.O., Lai,F.A., Ray,N. and Raftery,M.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5208–5212.

Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.

Dudai,Y. (1977) *FEBS Lett.*, **76**, 211–213.

Dudai,Y. (1980) In Satelle,D.B., Hall,L.M., Hildebrand,J.G. (eds), *Receptors for Neurotransmitters, Hormones and Pheromones in Insects*. Elsevier North-Holland, Amsterdam, pp. 93–110.

Finer-Moore,J. and Stroud,R.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 155–159.

Fink,G.R. (1986) *Cell*, **45**, 155–156.

Giraudat,J., Dennis,M., Heidmann,T., Chang,J.-Y. and Changeux,J.-P. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2719–2723.

Goldman,D., Deneris,E., Luyten,W., Kochar,A., Patrick,J. and Heinemann,S. (1987) *Cell*, **48**, 965–973.

Grenningloh,G., Rienitz,A., Schmitt,B., Methfessel,C., Zensen,M., Beyreuther,K., Gundelfinger,E.D. and Betz,H. (1987) *Nature*, **328**, 215–220.

Guy,H. (1984) *Biophys. J.*, **45**, 249–261.

Hall,L.M. (1980) In Satelle,D.B., Hall,L.M. and Hildebrand,J.G. (eds), *Receptors for Neurotransmitters, Hormones and Pheromones in Insects*. Elsevier North-Holland, Amsterdam, pp. 111–124.

Hall,L.M. and Teng,N.N.H. (1975) In McMahon,D. and Fox,C.F. (eds), *Developmental Biology. Pattern formation. Gene regulation. ICN–UCLA Symposia on Molecular and Cellular Biology*. W.A.Benjamin, Inc., Menlo Park, CA, Vol. 2, pp. 282–289.

Hall,L.M.C. and Spierer,P. (1986) *EMBO J.*, **5**, 2949–2954.

Hall,L.M.C., Mason,P.J. and Spierer,P. (1983) *J. Mol. Biol.*, **169**, 83–96.

Hall,L.M., Von Borstel,R.W., Osmond,B.C., Hoeltzli,S.D. and Hudson,T.H. (1978) *FEBS Lett.*, **95**, 243–246.

Hanke,W. and Breer,H. (1986) *Nature*, **321**, 171–174.

Henikoff,S. (1984) *Gene*, **28**, 351–359.

Hermans-Borgmeyer,I., Zopf,D., Ryseck,R.P., Hovemann,B., Vetz,H. and Gundelfinger,E.D. (1986) *EMBO J.*, **5**, 1503–1508.

Hucho,F. (1986) *Eur. J. Biochem.*, **158**, 211–226.

Kozak,M. (1978) *Cell*, **15**, 1109–1123.

Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1–45.

Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.

Lichtensteiger,W., Hefti,F., Felix,D., Huwyler,T., Melamed,E. and Schlumpf,M. (1982) *Neuropharmacology*, **21**, 963–968.

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

Martin.B.R. (1986) In Conn,P.M. (ed.), *The Receptors*. Academic Press, Orlando, FL, Vol. 3, pp. 393–415.

Nef,P., Mauron,A., Stalder,R., Alliod,C. and Ballivet,M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7975–7979.

Nef,P., Oneyser,C., Alliod,C., Couturier,S. and Ballivet,M. (1988) *EMBO J.*, **7**, 595–601.

Neumann,D., Barchan,D., Safran,A., Gershoni,J.M. and Fuchs,S. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3008–3001.

Noda,M. , Takahashi,H., Tanabe,T., Toyosato,M., Kikyotani,S., Hirose,T., Asai,M., Takashima,H., Inayama,S., Miyata,T. and Numa,S. (1983a) *Nature*, **301**, 251–255.

Noda,M., Furutani,Y., Takahashi,H., Toyosato,M., Tanabe,T., Shimizu,S., Kikyotani,S., Kayano,T., Hirose,T., Inayama,S. and Numa,S. (1983b) *Nature*, **302**, 818–823.

Patrick,J. and Stallcup,W. (1977) *J. Biol. Chem.*, **252**, 8629–8633.

Peabody,D.S. and Berg,P. (1986) *Mol. Cell. Biol.*, **6**, 2695–2703.

Peabody,D.S., Subramani,S. and Berg,P. (1986) *Mol. Cell. Biol.*, **6**, 2704–2711.

Poole,S.J., Kauvar,L.M., Drees,B. and Kornberg,T. (1985) *Cell*, **40**, 37–43.

Ratnam,M., Le Nguygen,D., Rivier,J., Sargent,P.B. and Lindstrom,J. (1986) *Biochemistry*, **25**,. 2633–2643.

Rudloff,E. (1978) *Exp. Cell Res.*, **111**, 185–190.

Sakurai,Y., Takano,Y., Kohjimoto,Y., Honda,K. and Kamiya,H.O. (1982) *Brain Res.*, **242**, 99–106.

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

Schmidt-Nielsen,B.K., Gepner,J.I., Teng,N.N.H. and Hall,L.M. (1977) *J. Neurochem.*, **29**, 1013–1029.

Schofield,P.R., Darlison,M.G., Fujita,N., Burt,D.R., Stephenson,F.A., Rodriguez,H., Rhee,L.M., Ramachandran,J., Reale,V., Glencorse,T.A., Seeburg,P.H. and Barnard,E.A. (1987) *Nature*, **328**, 221–227.

Spierer,P., Spierer,A., Bender,W. and Hogness,D.S. (1983) *J. Mol. Biol.*, **168**, 35–50.

Staden,R. (1982) *Nucleic Acids Res.*, **10**, 4731–4751.

Von Heijne,G. (1983) *Eur. J. Biochem.*, **133**, 17–21.

Weigel,D., Knust,E. and Campos-Ortega,J.A. (1987) *Mol. Gen. Genet.*, **207**, 374–384.

Wharton,K.A., Yedvobnick,B., Finnerty,V.G. and Artavanis-Tsakonas,S. (1985) *Cell*, **40**, 55–62.

Whiting,P.J., Schoepfer,R., Swanson,L.W., Simmons,D.M. and Lindstrom, J.M. (1987) *Nature*, **327**, 515–518.

Wilson,P.T., Lentz,T.L. and Hawrot,E. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8790–8794.

Received on November 25, 1987; revised on January 8, 1988