Volume 235, number 1,2, 40–46 FEB 06089 August 1988

Characterization of an invertebrate nicotinic acetylcholine receptor gene: the *ard* gene of *Drosophila melanogaster*

Erich Sawruk, Irm Hermans-Borgmeyer[+], Heinrich Betz and Eckart D. Gundelfinger

ZMBH, Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

Received 1 June 1988

The *ard* gene encodes a neuronal nicotinic acetylcholine receptor (AChR) protein from *Drosophila* (ARD protein). Cytogenetically this gene maps at position 64B/C on the left arm of the 3rd chromosome. Five introns interrupt the protein coding region of the gene, and one is found upstream of the translation start site. The *ard* gene thus contains less introns than vertebrate muscle AChR genes, but, with one exception, the positions of the resident introns are precisely conserved. Implications for the evolution of AChR genes are discussed.

Neuronal acetylcholine receptor; Gene structure; Evolution; (*Drosophila melanogaster*)

1. INTRODUCTION

Ligand-gated ion channels, including both excitatory, e.g. glutamate and acetylcholine receptors (AChRs), as well as inhibitory, e.g. glycine and γ-aminobutyric acid receptors, are crucial for signal transmission at many chemical synapses. The subunits of these oligomeric receptor proteins are encoded by a large super-family of genes [1,2]. In higher vertebrates, five different genes encode polypeptides of the two developmental isoforms of muscle AChR [3,4], and at least four additional ones neuronal AChR subunits [5–7]. Considerable complexity also appears to exist in the case of mammalian inhibitory neurotransmitter receptor proteins (Becker, C.M. and Schofield, P., personal communications).

In insects, acetylcholine is not used as a transmitter at neuromuscular synapses [8], and at least one insect neuronal AChR has been shown to be a homo-oligomeric protein [9]. Thus, fewer and phylogenetically more primitive AChR polypeptides can be assumed to occur in invertebrates. In *Drosophila melanogaster*, the primary structures of two different AChR polypeptides have been deduced from encoding DNAs: the α-like subunit (ALS), which resembles the ligand binding subunit of vertebrate AChRs [10], and the ARD protein, a putative non-ligand binding subunit [11–13]. The ARD protein is part of a membrane protein which binds α-bungarotoxin, a potent antagonist of vertebrate muscle AChR (Schloß et al., submitted). The corresponding gene is expressed in nervous tissue at all stages of *Drosophila* development ([13]; Hermans-Borgmeyer et al., submitted). Here we have analyzed structural features of the *ard* gene. Our data suggest a high degree of conservation of the organization of AChR genes during phylogeny.

Correspondence address: E. Sawruk, ZMBH, Universität Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

[+] *Present address:* Molecular Neurobiology Laboratory, Salk Institute, San Diego, CA 92138, USA

Abbreviation: AChR, acetylcholine receptor

2. MATERIALS AND METHODS

2.1. *Materials*

A genomic DNA library prepared from *D. melanogaster*, Oregon R, in phage EMBL4 was kindly provided by V. Pirrotta. The cDNA clone 4D8 encoding the γ-subunit of *Torpedo*

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

californica AChR [14] was obtained from M. Ballivet and S. Heinemann. *Drosophila* cDNA libraries, prepared from head-specific poly(A^+) RNA of adult flies, were obtained from P. Salvaterra and H. Steller.

2.2. *Screening of the genomic library and characterization of clones*

Genomic clones were isolated by low stringency hybridization with the 4D8 cDNA probe as described previously and verified by cross-hybridization with the ARD1 cDNA clone [11].

The three recombinant phages with the longest DNA inserts were mapped with various restriction enzymes [15]. Appropriate fragments of the gARD5 phage carrying all *Eco*RI fragments which hybridized in genomic Southern blots to the *Hin*dIII/*Eco*RI fragment of the ARD1/32 cDNA clone [12] were subcloned into plasmid pUC 18 and M13 phages mp 18 and mp 19 [16]. DNA sequences were obtained using the dideoxy chain termination method [17]. Computer-assisted sequence analysis was performed using the BSA program library of the DKFZ (Heidelberg).

2.3. *Southern blot analysis*

Genomic DNA of *D. melanogaster* was digested with different restriction enzymes, separated on 1% agarose gels, and transferred to nitrocellulose filters (Schleicher and Schuell) or Nylon membranes (Gene screen, NEN) according to Herrmann et al. [18]. DNA probes were labelled to specific activities of 2×10^7–3×10^8 cpm/μg using a nick translation kit (Amersham-Buchler). Low stringency hybridization with the 4D8 cDNA probe was performed according to McGinnis et al. [19]. Stringent hybridization conditions with the *Hin*dIII/*Eco*RI restriction fragment of the ARD1/32 cDNA were according to Church and Gilbert [20].

3. RESULTS AND DISCUSSION

3.1. *Isolation of genomic DNA clones encoding the ARD protein*

Southern blots of *Drosophila* genomic DNA digested with *Eco*RI restriction endonuclease were hybridized under conditions of low stringency with a cDNA fragment encoding the γ-subunit of the *Torpedo* electroplax AChR [14]. As shown in fig.1A, two hybridizing bands of ≈3 kb and ≈7 kb were detected. Under more stringent conditions, an *ard* cDNA probe hybridized preferentially to the 3 kb *Eco*RI fragment of genomic DNA (fig.1B). In addition, three other *Eco*RI fragments of about 1 kb, 7.5 kb and 8 kb were detected, two of which (1 kb and 7.5 kb) also encode short regions of the *ard* mRNA (cf. fig.2). The 8 kb band probably represents a cross-hybridizing fragment of a related gene. Also, the hybridization pattern of *ard* cDNA to genomic DNA cleaved with other restriction enzymes can only be explained if one assumes cross-hybridization with related sequences even under the relatively stringent conditions used in this experiment (cf. figs 1B and 2).

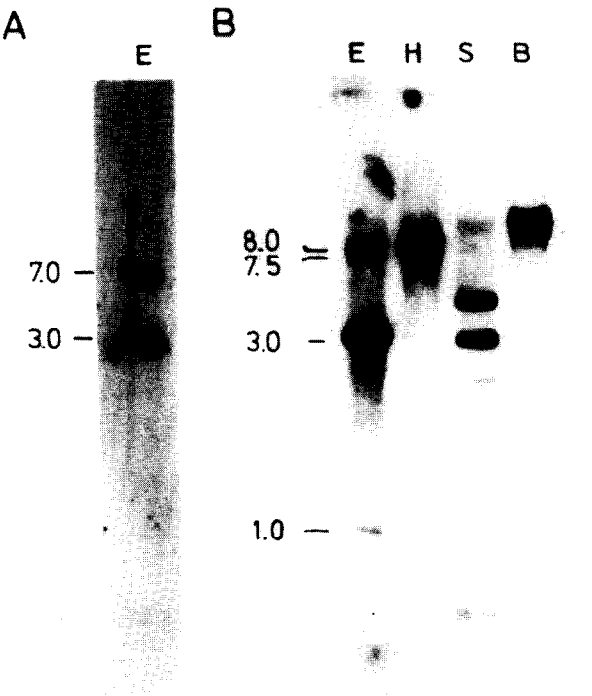


Fig.1. Southern blot analysis of *Drosophila* genomic DNA hybridized to *Torpedo* AChR γ-subunit cDNA (A) and the *Hin*dIII/*Eco*RI subclone of ARD1/32 cDNA (B). 5 μg (A) or 3 μg (B) of genomic DNA were digested with restriction endonucleases *Eco*RI (E), *Hin*dIII (H), *Sst*I (S), or *Bam*HI (B). For details see section 2. Lambda wild type DNA cleaved with *Eco*RI/*Hin*dIII was used to size the hybridizing fragments.

The *Torpedo* AChR γ-subunit probe was employed to screen a *Drosophila* genomic DNA library under stringency conditions where about 55% sequence homology should be recognized [19]. Of the six recombinant phages isolated, all hybridized to *ard* cDNA under stringent conditions. The gARD5 clone which contained three of the *Eco*RI fragments detected in the genomic Southern blot was analyzed in detail.

3.2. *Structure and cytogenetic localization of the* ard *gene*

The exon-intron organization of the *ard* gene and a comparison to several cDNA clones analyzed are presented in fig.2, and the determined nucleotide sequence in fig.3. The entire gene covers >7 kb and has at least six introns. The 5′-end of the *ard* gene and its promoter region have not been

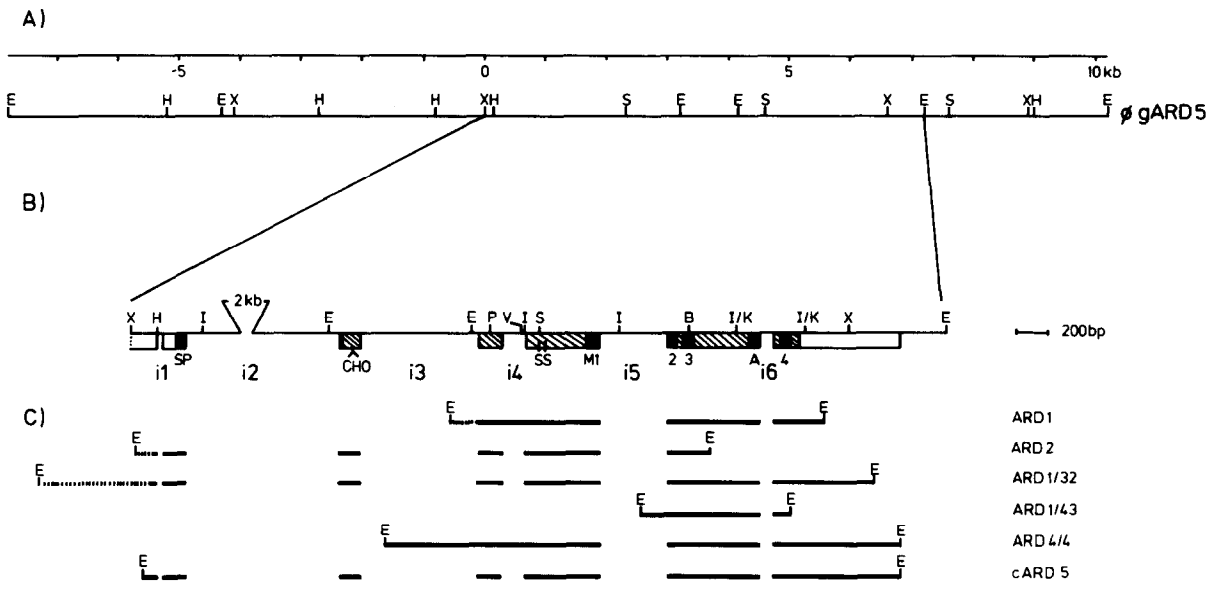


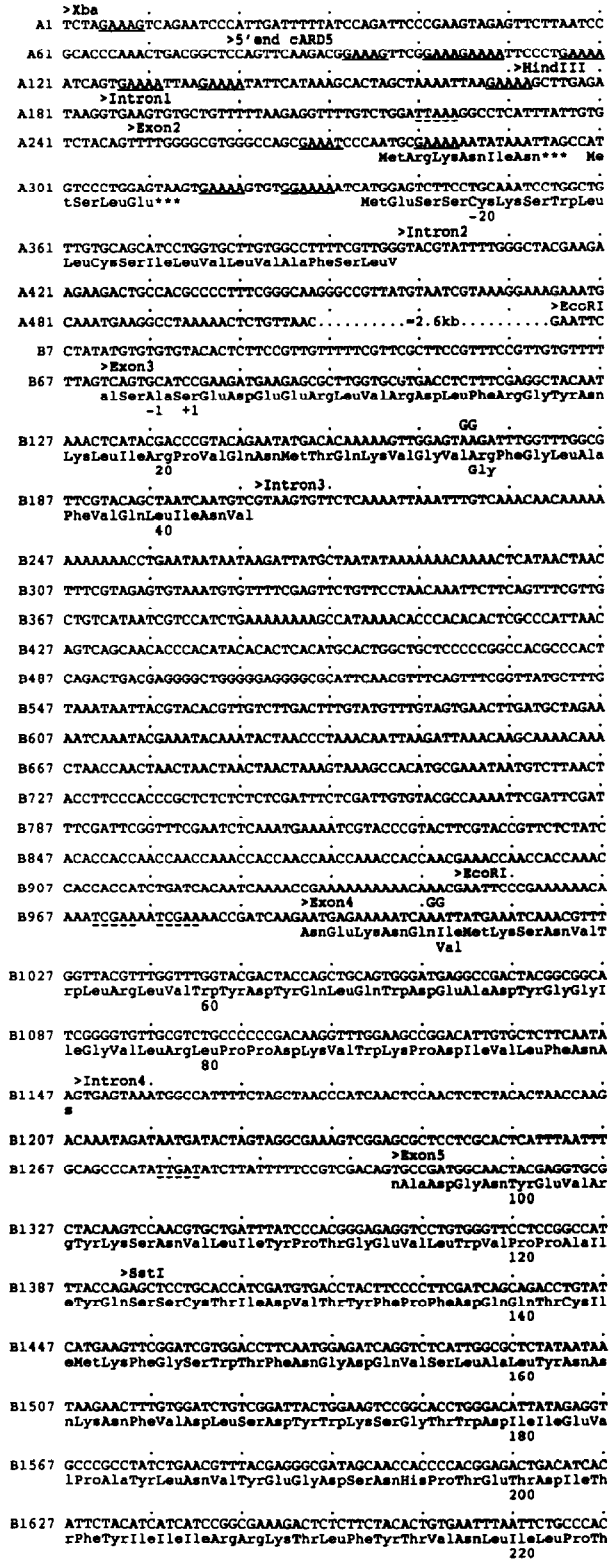
Fig.2. Structure of the *ard* gene. (A) Restriction map of phage gARD5. (B) Enlarged physical map of the *ard* gene region analyzed. Restriction sites are B, *Bst*EII; E, *Eco*RI; H, *Hin*dIII; I, *Hin*dII; K, *Hpa*I; P, *Pst*I; S, *Sst*I; V, *Eco*RV; X, *Xba*I. Exons are indicated by boxes: open, untranslated sequences; hatched, protein coding region; filled, characteristic regions [28] such as signal peptide (SP), membrane spanning regions (M1, 2, 3, 4) and amphipathic α-helix (A). CHO indicates a potential *N*-glycosylation site. SS indicates two cysteines conserved in all ligand-gated ion channel proteins [1,2]. Introns are numbered i1–i6. The *Xba*I restriction site located 79 nucleotides upstream of the 5′-end of the cDNA clone with the longest 5′-extension (cARD5) was arbitrarily defined as zero. (C) Alignment of some of the cDNA clones analyzed. Solid lines: complementary DNAs. *Eco*RI sites resulting from cDNA cloning are indicated. Most cDNA clones were isolated from a head-specific cDNA library provided by P. Salvaterra; the cARD5 cDNA was derived from a head-specific cDNA library obtained from H. Steller and G. Rubin. Note that three of our cDNA clones (ARD1, ARD2, ARD1/32) have 5′-terminal nucleotide sequences, which are not detected in other cDNAs nor in the genomic DNA (dotted lines).

defined; additional introns may exist upstream of the 5′-end of cARD5, the cDNA clone with the longest 5′-extension (figs 2C,3). From the size of the *ard* mRNA (3.2 kb; [11]), we estimate that about 700 nucleotides of the mature transcript remain to be mapped at the genomic level.

Intron i1, only 63 bp long, is located in the 5′-non-translated region 86 nucleotides upstream of the translation start site of the ARD protein. The five other introns disrupt the protein coding region while the 3′-untranslated sequences of gene and mRNA are colinear (figs 2C,3). All introns are flanked by proper splice donor and acceptor sites and, except in intron i2, sequences which resemble branchpoint consensus sequences are detected at least once within 50 nucleotides upstream of 3′-splice sites [21].

Between intron i1 and the translation start of the ARD protein there are two short open reading frames (ORFs) of 18 and 12 nucleotides which potentially code for a hexa- and a tetrapeptide, respectively. All three start codons share the

Fig.3. Partial nucleotide and deduced amino acid sequence of the *ard* gene. Restriction nuclease cleavage sites given in fig.2A, the 5′-end of the cARD5 clone, 5′-borders of exons and introns, and the 3′-end of the *ard* mRNA are indicated above the nucleotide sequence. The deduced amino acid sequence of the short 5′-ORFs and the ARD protein are given below the nucleotide sequence. Nucleotides are numbered in three blocks (A,B,C) on the right, amino acids below the sequence; the signal peptide is indicated by negative numbers. The GAAAR motifs occurring in the 5′-untranslated region and the three AATAAA potential polyadenylation signals of the 3′-untranslated region are underlined by solid lines, and splice branchpoint consensus sequences by broken lines. The 3′-terminal octamer repeats are marked by a solid line above the sequence. Microheterogeneities are given above the nucleotide and below the amino acid sequences, respectively.



nucleotides −4 and −1 with consensus flanking sequences detected for *Drosophila* translation start sites (C/AAAA/CAUG; [22]). However, only the actual initiation codon of the ARD precursor is preceded by an A in position −3, which is the most favoured nucleotide in this position among all vertebrate translation start consensus sequences. Similar minicistrons upstream of the protein coding regions have been found in a variety of eukaryotic mRNAs [23]. These also include other AChR genes of both nicotinic and muscarinic type [10,24,25] and the structural gene for *Drosophila* acetylcholinesterase, where as much as six short ORFs have been detected [26]. In the yeast *GCN4* gene, upstream ORFs have been shown to be involved in translational control of the *GCN4* transcript [27].

The leader sequence of the *ard* mRNA contains twelve motifs with the consensus sequence GAAAR irregularly distributed on both sites of intron i1. Within the ARD protein coding region, ten nucleotide exchanges were detected among the various cDNAs and the genomic DNA (fig.3). Two of these microheterogeneities result in variations in the amino acid sequence of the ARD protein (Gly or Arg in position 32, Ile or Val in position 49) while the other eight substitutions occur in silent third codon positions. A C⟶T transition in nucleotide C617 affects a *Hpa*I restriction site, which is present in the ARD1, but not in the other cDNAs nor the genomic clone gARD5. Interestingly, six of the microheterogeneities occur in the region encoding the presumptive cytoplasmic domain of the ARD protein, the least conserved segment in all AChR proteins [28]. We attribute the observed sequence variations to the occurrence of various alleles of the *ard* gene rather than to the existence of different highly conserved genes in the haploid genome.

Within the 663 bp 3′-untranslated region, three potential AATAAA polyadenylation signals are found. The most 3′-proximal signal starts 28 nucleotides upstream of the site where cDNAs and genomic DNA diverge (nucleotide C1679 in fig.3; cDNA clone ARD4/4 ends here with a stretch of seven additional As). Neither Northern blot analysis [11] nor the analysis of 3′-ends of the various cDNAs provide any evidence that the two other AATAAA sequences are used as polyadenylation signals in vivo. Starting 72 nucleotides upstream of the 3′-end of the mRNA four YCCGRAGA repeats occur (fig.3). Their proximity to the 3′-end of the mRNA suggests that these octamer repeats may be implicated in the choice of the proper polyadenylation signal and/or 3′-end formation of the mRNA.

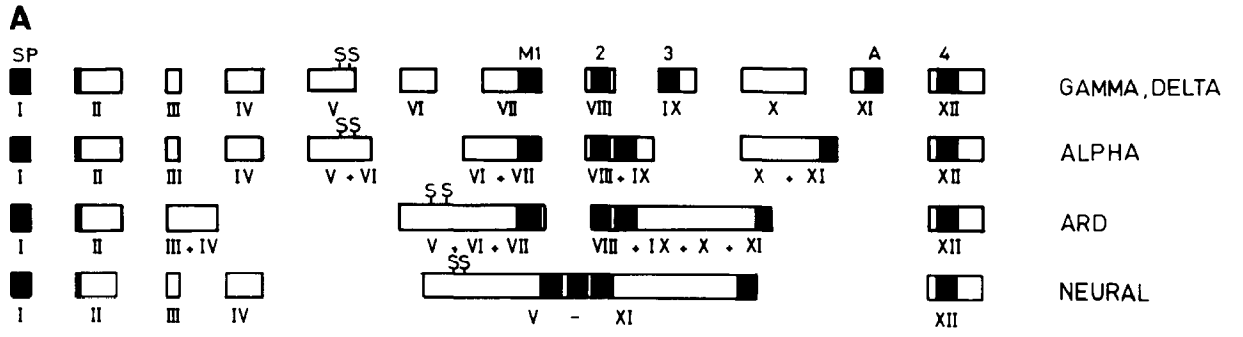
The cDNA clones ARD1, ARD1/43 and ARD4/4 represent partially spliced transcripts of the *ard* gene carrying sequences of introns i3, i4 and i5, respectively (fig.2C). Both the occurrence of translation stop codons and/or lengths of insertions which disrupt the reading frame of the ARD protein suggest that the corresponding transcript may not be translated.

To determine the cytogenetic localization of the *ard* gene, in situ hybridization to polytene salivary gland chromosomes with the biotin-labelled 3 kb *Eco*RI restriction fragment of clone gARD5 was performed. Hybridization was detected only to region 64B/C of the left arm of the third chromosome (fig.4).

3.3. *Evolution of AChR genes*

A comparison of the intron positions of the *ard* gene with those of the previously analyzed genes encoding the human α- [29] and γ-subunits [30], and the chicken α- [7], γ- and δ-subunits of muscle AChR [31], allows several conclusions on the evolution of these receptor genes. Within the protein coding region analyzed here, the *ard* gene contains fewer intervening sequences than the genes of the muscle AChR subunits (fig.5). There, eleven

Fig.4. In situ hybridization of the biotin-labelled 3 kb *Eco*RI fragment of gARD5 to salivary gland polytene chromosomes of *D. melanogaster* (Canton S). Hybridization was carried out essentially as described in [34]. The arrowhead indicates the site of hybridization.



B

Exon VII/VIII

ARD	237	ProAlaGluAlaGly - GluLysValThr	245
human α^1	236	ProThrAspSerGly - GluLysMetThr	244
human	244	ProAlaLysAlaGlyGlyGlnLysCysThr	253
chick	245	ProAlaLysAlaGlyGlyGlnLysCysThr	254
chick δ	252	ProAlaAspSerGly - GluLysMetThr	260
chick $\alpha 2^2$	242	ProSerAspCysGly - GluLysIleThr	250

Fig.5. Comparison of AChR genes. (A) Schematic presentation of exon-intron organization of the protein coding regions of the α-, γ- and δ-subunit genes of the vertebrate muscle AChR [7,29–31], the *ard* gene and the vertebrate neuronal AChR genes [7]. Conserved structural features of the encoded proteins are indicated: SP, signal peptide; M1, 2, 3, 4, membrane spanning regions; A, amphipathic α-helix; SS, conserved cysteines. Exons are numbered in Roman numerals. (B) Location of splice positions (▼) of the intron separating M1 and M2 on the aligned amino acid sequences. Numbers give the positions of the amino acid residues within the mature proteins. [1] chick α is identical to human α; [2] chick $\alpha 2$ is shown as a representative for all chick neuronal genes which have no intron in this position [7].

introns are found in the γ- and δ-subunit genes, and eight within the α-gene. However, the positions of most of the resident introns are conserved, i.e. the introns separating exons I and II, II and III, IV and V, XI and XII. The intron between exons VII and VIII (according to the nomenclature of fig.5) represents the only exception. This intron, which separates the first and second membrane spanning regions of the AChR proteins, is shifted by 8–14 nucleotides downstream as compared to the respective intron of the different muscle receptor genes.

Two possible mechanisms have been proposed to explain variability of intron positions: intron sliding and insertion of new introns at different times in evolution (discussed in [32,33]). While the introns separating exons VII and VIII of the muscle genes and the *ard* gene do not fall in the same phase of the reading frame, the sequences between the various positions are, however, highly conserved among species (fig.5B). Therefore a plausible explanation for the different intron positions may be that they were acquired independently and placed within a 'permissive' region between the two transmembrane domains. An evolutionary progenitor gene would then be expected to have no interruption between regions encoding the first and second membrane spanning segment. Indeed, the vertebrate neuronal AChR-related genes analyzed so far (fig.5) and the *Drosophila als* gene [10] do not possess an intron at this position. The differential intron placement implies that muscle and neuronal AChR genes may have different progenitors which have diverged prior to the evolution of AChR subunit heterogeneity. This hypothesis is also supported by amino acid sequence comparisons. The ARD protein exhibits a higher amino acid sequence homology to vertebrate

neuronal AChR proteins $\alpha 2$, $\alpha 3$, $\alpha 4$ and non-α [5–7] than to muscle receptor subunits.

NOTE ADDED IN PROOF

The nucleotide sequence of the *ard* gene was submitted to the EMBL sequence data library under the accession numbers X07956, X07957 and X07958.

Acknowledgements: We are grateful to R. Paro for help with chromosome mapping, V. Pirrotta, P. Salvaterra, H. Steller, and G. Rubin for providing libraries, C. Udri and H. Krischke for expert technical assistance, and I. Baro for help with the preparation of the figures. E.S. is a recipient of a Boehringer Ingelheim Fonds predoctoral fellowship. This work was supported by grants of the Deutsche Forschungsgemeinschaft (Gu 230/1-1, SFB 317), the Bundesministerium für Forschung und Technologie (BCT 365/7) and the Fonds der Chemischen Industrie.

REFERENCES

[1] 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.

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

[3] Kubo, T., Noda, M., Takai, T., Tanabe, T., Kayano, T., Shimizu, S., Tanaka, K., Takahashi, H., Hirose, T., Inayama, S., Kikuno, R., Miyata, T. and Numa, S. (1985) Eur. J. Biochem. 149, 5–13.

[4] Takai, T., Noda, M., Mishina, M., Shimizu, S., Furutani, Y., Kayano, T., Ikeda, T., Kubo, T., Takahashi, H., Takahashi, T., Kuno, M. and Numa, S. (1985) Nature 315, 761–764.

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

[6] Deneris, E.S., Connolly, J., Boulter, J., Wada, E., Wada, K., Swanson, L.W., Patrick, J. and Heinemann, S. (1988) Neuron 1, 45–54.

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

[8] Gerschenfeld, H.M. (1973) Physiol. Rev. 53, 1–119.

[9] Breer, H., Kleene, R. and Hinz, G. (1985) J. Neurosci. 5, 3386–3392.

[10] Bossy, B., Ballivet, M. and Spierer, P. (1988) EMBO J. 7, 611–618.

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

[12] Gundelfinger, E.D., Hermans-Borgmeyer, I., Zopf, D., Sawruk, E. and Betz, H. (1986) in: Nicotinic Acetylcholine Receptor (Maelicke, A. ed.) NATO ASI Series, vol.H3, pp.437–446, Springer, Berlin.

[13] Wadsworth, S.C., Rosenthal, L.S., Kammermeyer, K.L., Potter, M.B. and Nelson, D.J. (1988) Mol. Cell. Biol. 8, 778–785.

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

[15] Rackwitz, H.-R., Zehetner, G., Frischauf, A.-M. and Lehrach, H. (1984) Gene 30, 195–200.

[16] Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101–106.

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

[18] Herrmann, B., Bucan, M., Mains, P.E., Frischauf, A.-M., Silver, L.M. and Lehrach, H. (1986) Cell 44, 469–476.

[19] McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature 308, 428–433.

[20] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.

[21] Keller, E.B. and Noon, W.A. (1985) Nucleic Acids Res. 13, 4971–4981.

[22] Cavener, D.R. (1987) Nucleic Acids Res. 15, 1353–1361.

[23] Kozak, M. (1986) Cell 47, 481–483.

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

[25] Kubo, T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishina, M., Haga, T., Haga, K., Ichiyama, A., Kangawa, K., Kojima, M., Matsuo, H., Hirose, T. and Numa, S. (1986) Nature 322, 411–416.

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

[27] Mueller, P.P. and Hinnebusch, A.G. (1986) Cell 45, 201–207.

[28] Hucho, F. (1986) Eur. J. Biochem. 158, 211–226.

[29] Noda, M., Furutani, Y., Takahashi, H., Toyosato, M., Tanabe, T., Shimizu, S., Kikyotani, S., Kayano, T., Hirose, T., Inayama, S. and Numa, S. (1983) Nature 305, 818–823.

[20] Shibahara, S., Kubo, T., Perski, H.J., Takahashi, H., Noda, M. and Numa, S. (1985) Eur. J. Biochem. 146, 15–22.

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

[32] Rogers, J. (1985) Nature 315, 458–459.

[33] Gilbert, W., Marchionni, M. and McKnight, G. (1986) Cell 46, 151–154.

[34] Langer-Safer, P.R., Levine, M. and Ward, D.C. (1982) Proc. Natl. Acad. Sci. USA 79, 4381–4385.