J Mol Evol (1995) 40:155–172

JOURNAL OF MOLECULAR EVOLUTION



# Molecular Evolution of the Nicotinic Acetylcholine Receptor: An Example of Multigene Family in Excitable Cells

**Nicolas Le Novère, Jean-Pierre Changeux**

Laboratoire de Neurobiologie Moléculaire, 25, rue du Dr Roux, 75015 Paris, France

Received: 4 June 1994 / Accepted: 22 August 1994

**Abstract.** An extensive phylogenetic analysis of the nicotinic-acetylcholine-receptor subunit gene family has been performed by cladistic and phenetic methods. The conserved parts of amino acid sequences have been analyzed by CLUSTAL V and PHYLIP software. The structure of the genes was also taken in consideration. The results show that a first gene duplication may have occurred before the appearance of Bilateria. Three subfamilies then appeared: I—the neuronal α-bungarotoxin binding-site subunits (α7, α8); III—the neuronal nicotinic subunits (α2–α6, β2–β4), which also contain the muscle acetylcholine-binding subunit (α1); and IV—the muscle non-α subunits (β1, γ, δ, ϵ). The Insecta subunits (subfamily II) could be orthologous to family III and IV. Several tissular switches of expression from neuron to muscle and the converse can be inferred from the extant expression of subunits and the reconstructed trees. The diversification of the neuronal nicotinic subfamily begins in the stem lineage of chordates, the last duplications occurring shortly before the onset of the mammalian lineage. Such evolution parallels the increase in complexity of the cholinergic systems.

**Key words:** Nicotinic receptor — Ligand-gated channel — Multigene family — Gene phylogeny

*Correspondence to:* N. Le Novère

*Abbreviations:* α-bungarotoxin (α-Bgt), acetylcholine (ACh), maximum of parsimony (MP), million years ago (MYA), neighbor-joining (NJ), nicotinic acetylcholine receptor (nAChR)

## Introduction

Acetylcholine (ACh) has long been recognized as a neurotransmitter active in Bilateria nervous system and muscle. Two distinct categories of receptors are engaged in the biological effects of ACh: the muscarinic and nicotinic receptors. Muscarinic receptors belong to the superfamily of G-protein-coupled receptors; they consist of single integral proteins with seven transmembrane segments and interact, on their cytoplasmic face, with heterotrimeric G-proteins. Nicotinic receptors (nAChR) belong to the superfamily of ligand-gated ion channels; they are hetero-oligomers composed of five subunits, each with four transmembrane domains (Devillers-Thiéry et al. 1993; Galzi and Changeux 1994). ACh binding causes an ionic channel, most often cationic, to open, resulting in a rapid change in the electrical, and secondarily metabolic, state of the target cell (Greenberg et al. 1986; Bertrand et al. 1993).

The nAChR of striated muscle is the best-characterized member of the ligand gated-ion-channel superfamily (Changeux 1990; Karlin 1993): it is a heteropentamer (with the stoichiometry $\alpha_2\beta\gamma\delta$). According to current models (Bertrand et al. 1993), the ion channel forms along the axis of pseudosymmetry perpendicular to the cell membrane. The subunits share a similar hydropathic profile with four short hydrophobic domains (MI–MIV) and two long hydrophilic domains. The largest, relatively conserved, hydrophilic domain is located at the amino-terminal side of the subunit polypeptide, and the other, highly variable, joins hydrophobic do-

mains MIII and MIV. The amino-terminal hydrophilic domain carries the ACh binding site (Devillers-Thiéry et al. 1993) and faces the synaptic cleft, whereas the other hydrophilic domain is exposed to the cytoplasm.

Molecular cloning and sequencing studies have revealed the existence of a family of genes, expressed in neurons, which code for nAChR subunits **homologous** (see Appendix 1 for definitions of boldface terms) to those of muscle nAChR (Sargent 1993). In the jawed vertebrate nervous systems, several subunits (named α2–α8) have been identified which share with the muscle-type receptor α1 subunit the pair of cysteines shown to contribute to the ACh binding site (Wada et al. 1988; Shoepfer et al. 1990; Cockcroft et al. 1992; Karlin 1993). Other homologous chains, lacking the cysteine pair, have been characterized and named non-α or β2–β4. As for muscle nAChR, the functional neuronal nAChR is an heteropentamer made up by the assembly of α and β subunits, with a putative stoichiometry in vitro of $\alpha_2\beta_3$ (Anand et al. 1991; Cooper et al. 1991). The recent evidence that α5 is coprecipitated with another α and β subunit in some neuronal nAChRs indicates that more than two different subunits may assemble together to form a receptor molecule (Conroy et al. 1992; Vernallis et al. 1993). In contrast, in reconstituted systems, the α7 or α8 subunits can form functional homo-oligomers (Couturier et al. 1990; Revah et al. 1991; Anand et al. 1993). The autoradiographic studies in the brain revealed that $^3$H-nicotine labels receptors formed by subunits α2–α6 and β2–β4 but not receptors formed by subunits α7 and α8 (Clarke et al. 1985), which are labeled by α-bungarotoxin (α-Bgt).

The combinatorial diversity resulting from the assembly of the multiple neuronal subunits results in a wide spectrum of structurally and functionally distinct nAChRs with different pharmacological specificities and ion-channel properties (Role 1992). Such differences have been directly demonstrated in *Xenopus* oocytes and mouse fibroblasts after heterologous expression (Luetje and Patrick 1991; Whiting et al. 1991). Furthermore, multiple functionally distinct types of nAChRs have been detected in different brain areas and subcellular compartments (Mulle et al. 1991).

The nAChR is present in the whole phylum of Bilateria, from nematodes to humans (Gerschenfeld 1973; Darlison et al. 1993; Fleming et al. 1993; Leech and Sattelle 1993). Several nAChR neuronal subunits have been cloned in *Drosophila,* locust, and nematode (Gundelfinger 1992; Fleming et al. 1993). In the insect nervous system, ACh is the major excitatory neurotransmitter, in contrast to vertebrates, where glutamate predominates. At the neuromuscular junction, glutamate is the excitatory transmitter in arthropods, whereas it is ACh in vertebrates. Since some lines of evidence suggest that nAChR is also responsible for neuromuscular transmission in nematodes, molluscs, and annelids (Gerschenfeld 1973; Segerberg and Stretton 1993), it is of interest to assess whether the original form of nAChR appeared in muscle or in neurons.

The neuronal nAChRs provide a good example of a multigene family differentially expressed in the nervous system. Its evolution deserves comparison with the increase in complexity of the vertebrate nervous system and, in particular, cholinergic systems. Some partial trees have already been constructed, but without comparative methods and without statistical support (Brehm et al. 1991; Cockcroft et al. 1992). Here we provide a molecular phylogenetic study of the whole family of nAChR genes.

## Materials and Methods

The programs used in this work were run on a Sun computer in a UNIX environment. The sequences were loaded from Genbank and EMBL databases (Table 1) with Sequence Analysis Software Package 7.1 from the Genetic Computer Group.

*Alignments of the Sequences.* Alignments were performed using CLUSTAL V software (Higgins and Sharp 1988). This program compares the sequences in pairs according to Wilbur and Lipman (1983) (gap penalty = 3) and builds a preliminary tree by an unweighted pair-group method of arithmetic averages (UPGMA) (Sneath and Sokal 1973). Then the program aligns all sequences in order of decreased similarity according to Feng and Doolittle (1987) (fixed and floated gap penalty = 10). The use of different values of gap cost changed neither topology nor the ratio of branch lengths but did result in a homothetic transformation of the trees. The similarities have been determined by the Dayhoff PAM 250 matrix. The protein sequences were aligned after the following modifications:

Deletions of the signal peptide (corresponding to tα1 aa 1–27), the small nonconserved part in amino-terminal part (corresponding to tα1 R188), the highly variable cytoplasmic region (corresponding to tα1 aa 356–393), and the carboxy-terminal part (corresponding to tα1 aa 452–461). The alignment obtained with 48 sequences shows 394 sites with 357 **informative sites** (Appendix 2).

To determine the branching of the nematode sequence onachr (Fig. 4) which amino-terminal part is not known, a further deletion (corresponding to the 38 amino-terminal aa of cα8 from the Appendix 2 alignment) was performed on 12 sequences. The alignment of the 13 sequences shows 351 sites with 263 informative sites (Appendix 3).

*Sequence Analyses.* Inferences on gene evolution were obtained with the PHYLIP 3.5c software of Felsenstein (1993).

The cladistic method was the maximum of parsimony (MP) (Fitch 1971, program PROTPARS). The mouse 5-HT3 subunit and the rat glycine α3 subunit were used as outgroups. The use of the rat GABA α1 subunit instead of the glycine α3 subunit did not change the results (data not shown). The phenetic method was neighbor-joining (NJ) (Saitou and Nei 1987, program NEIGHBOR). The distance matrix was provided by the Dayhoff PAM matrix (Dayhoff 1979, program PROTDIST). The statistical test used to determine the strength of the trees was bootstrap resampling (Felsenstein 1985) with the SEQBOOT (seed: 5) and CONSENSE programs.

*Analyses of Gene Structure.* The mixed-parsimony algorithm with the Wagner method (Eck and Dayhoff 1966, program MIX) and the

**Table 1.** Genes used in this study. Abbreviations are those used in the text and trees. The first letters represent the species, followed by the name of the subunit[a]

| Gene | Species | Acc. no. | Ref. |
|---|---|---|---|
| bα1 | *Bos taurus* | X02509 | Noda et al. *Nature* 305:818 (1983) |
| bβ1 | *Bos taurus* | X00962 | Tanabe et al. *Eur J Biochem* 144:11 (1984) |
| bδ | *Bos taurus* | X02473 | Kubo et al. *Eur J Biochem* 149:5 (1985) |
| bε | *Bos taurus* | X02597 | Takai et al. *Nature* 315:761 (1985) |
| bγ | *Bos taurus* | M28307 | Takai et al. *Eur J Biochem* 143:109 (1984) |
| cα2 | *Gallus domesticus* | M07339-44 | Nef et al. *EMBO J* 7:595 (1988) |
| cα3 | *Gallus gallus* | M37336 | Couturier et al. *JBC* 265:17560 (1990) |
| cα4 | *Gallus domesticus* | X07348-53,99 | Nef et al. *EMBO J* 7:595 (1988) |
| cα5 | *Gallus gallus* | J05642 | Couturier et al. *JBC* 265:17560 (1990) |
| cα7 | *Gallus gallus* | X68586 | Couturier et al. *Neuron* 5:847 (1990) |
| cα8 | *Gallus gallus* | X52296 | Schoepfer et al. *Neuron* 5:35 (1990) |
| cβ2 | *Gallus domesticus* | X53092 | Schoepfer et al. *Neuron* 1:241 (1988) |
| cβ4 | *Gallus gallus* | J05643 | Couturier et al. *JBC* 265:17560 (1990) |
| cδ | *Gallus gallus* | K02903 | Nef et al. *PNAS* 81:7975 (1984) |
| cγ | *Gallus gallus* | K02904 | Nef et al. *PNAS* 81:7975 (1984) |
| dαLi | *Drosophila melanogaster* | X07194 | Bossy et al. *EMBO J* 7:611 (1988) |
| dα2 | *Drosophila melanogaster* | X53583 | Sawruk et al. *EMBO J* 9:2671 (1990) |
| dβ2 | *Drosophila melanogaster* | X55676 | Sawruk et al. *FEBS Lett* 273:177 (1990) |
| dnachr | *Drosophila melanogaster* | X04016 | Hermans-Borgmeyer et al. *EMBO J* 5:1503 (1986) |
| gfα3 | *Carassius auratus* | X54051 | Hieber et al. *NAR* 18:5293 (1990) |
| gfβ2 | *Carassius auratus* | X54052 | Hieber et al. *NAR* 18:5307 (1990) |
| gfnα2 | *Carassius auratus* | X14786 | Cauley et al. *J Cell Biol* 108:637 (1989) |
| gfnα3 | *Carassius auratus* | M29529 | Cauley et al. *J Neurosci* 10:670 (1990) |
| hα1 | *Homo sapiens* | Y00762 | Schoepfer et al. *FEBS Lett* 226:235 (1988) |
| hα3 | *Homo sapiens* | M37981 | Mihovilovic et al. *J Exp Neurol* 111:175 (1991) |
| hα5 | *Homo sapiens* | M83712 | Chini et al. *PNAS* 89:1572 (1992) |
| hα7 | *Homo sapiens* | X70297 | Peng et al. *Mol Pharmacol* 45:546 (1994) |
| hβ1 | *Homo sapiens* | X14830 | Beeson et al. *NAR* 17:4391 (1989) |
| hβ2 | *Homo sapiens* | X53179 | Anand et al. *NAR* 18:4272 (1990) |
| hβ3 | *Homo sapiens* | | Willoughby et al. *Neurosci Lett* 155:136 (1993) |
| hβ4 | *Homo sapiens* | X68275 | Tarroni et al. *FEBS Lett* 312:66 (1992) |
| hδ | *Homo sapiens* | X55019 | Luther et al. *J Neurosci* 9:1082 (1989) |
| hε | *Homo sapiens* | X66403 | Beeson et al. unpublished |
| mser | *Mus musculus* | M74425 | Maricq et al. *Science* 254:432 (1991) |
| nα1 | *Naja naja* | M26388 | Neumann et al. *PNAS* 86:7255 (1989) |
| onachr | *Onchocerca volvulus* | L20465 | Ajuh and Egwang unpublished (1993) |
| rα2 | *Rattus norvegicus* | L10077 | Wada et al. *Science* 240:330 (1988) |
| rα3 | *Rattus norvegicus* | X03440 | Boulter et al. *Nature* 319:368 (1986) |
| rα4 | *Rattus norvegicus* | M15681-82 | Goldman et al. *Cell* 48:965 (1987) |
| rα5 | *Rattus norvegicus* | J05231 | Boulter et al. *J Biol Chem* 265:4472 (1990) |
| rα6 | *Rattus norvegicus* | L08227 | Boulter unpublished (1988) |
| rα7 | *Rattus norvegicus* | M85273 | Seguele et al. *J Neurosci* 13:596 (1993) |
| rβ2 | *Rattus norvegicus* | — | Deneris et al. *Neuron* 1:45 (1988) |
| rβ3 | *Rattus norvegicus* | J04636 | Deneris et al. *J Biol Chem* 264:6268 (1989) |
| rβ4 | *Rattus norvegicus* | J05232, M89971, M33951-3, M89989 | Boulter et al. *J Biol Chem* 265:4472 (1990) |
| rδ | *Rattus norvegicus* | X74835 | Witzemann et al. *Eur J Biochem* 194:437 (1990) |
| rε | *Rattus norvegicus* | X13252 | Criado et al. *NAR* 16:10920 (1988) |
| rγ | *Rattus norvegicus* | X74834 | Witzemann et al. *Eur J Biochem* 194:437 (1990) |
| rglyα3 | *Rattus norvegicus* | M55250 | Kuhse et al. *J Biol Chem* 265:22317 (1990) |
| sαL1 | *Schistocerca gregaria* | X55439 | Marshall et al. *EMBO J* 9:4391 (1991) |
| tα1 | *Torpedo californica* | X00963 | Noda et al. *Nature* 299:793 (1982) |
| tβ1 | *Torpedo californica* | J00964 | Noda et al. *Nature* 301:251 (1983) |
| tδ | *Torpedo californica* | J00965 | Noda et al. *Nature* 301:251 (1983) |
| tγ | *Torpedo californica* | J00966 | Ballivet et al. *PNAS* 79:4466 (1982) |
| xα1a | *Xenopus laevis* | X17244 | Hartman et al. *Nature* 343:372 (1990) |
| xα1b | *Xenopus laevis* | X07067 | Baldwin et al. *J Cell Biol* 106:469 (1988) |
| xb1 | *Xenopus laevis* | U04618 | Kullberg et al. *Rec Chan* (1994) in press |
| xδ | *Xenopus laevis* | X07069 | Baldwin et al. *J Cell Biol* 106:469 (1988) |
| xγ | *Xenopus laevis* | X07068 | Baldwin et al. *J Cell Biol* 106:469 (1988) |

[a] The six sign codes are the accession numbers of the Genbank-Embl databases. Rat α4 is the isoform 4-2

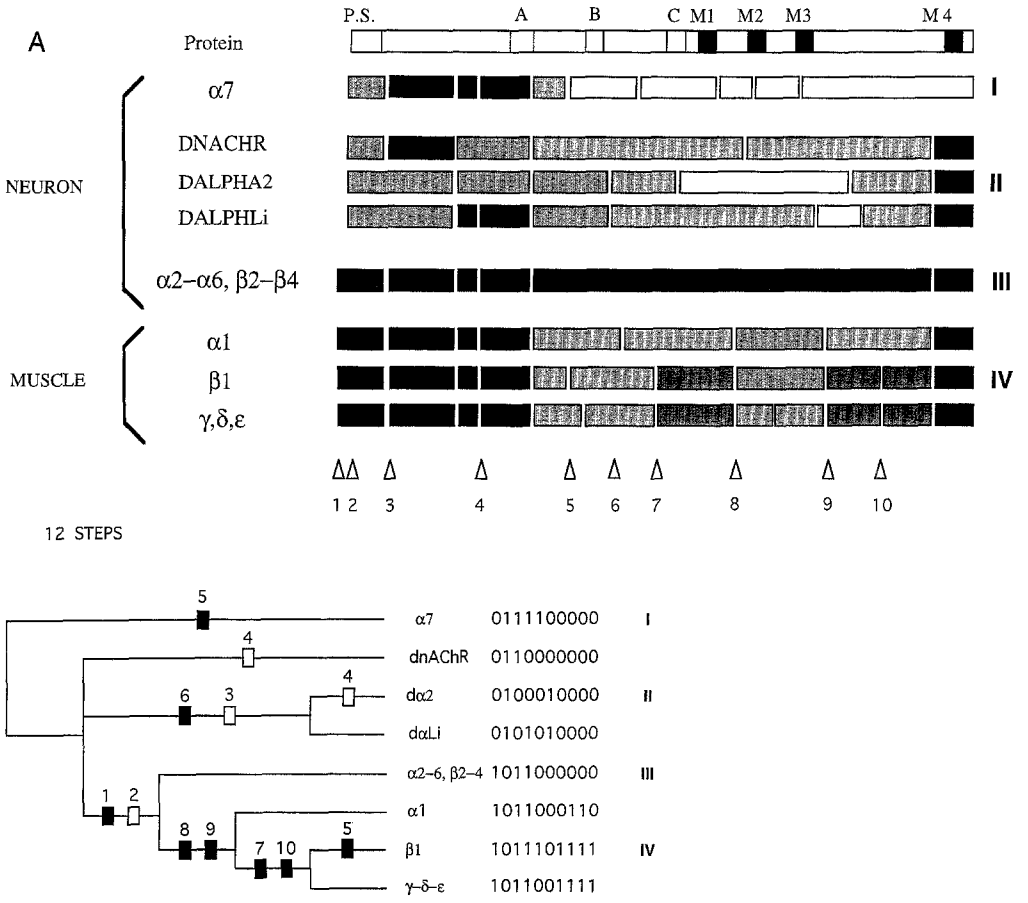


**Fig. 1.** **A** Structure of the subunit genes. Only the exons at least partially coding are represented. The gray level grossly reflects the conservation of the exon through the family (i.e., the presence of an exonic frontier at this place in different subunits, but not the sequence similitude between exons). A, B, C: binding site loops. *M1, M2, M3, M4:* transmembrane segments. The *arrowheads* mark the informative limits (in a cladistic acception). Adapted from Jonas et al. (1990) with the help of Alain Bessis. **B** Cladogram constructed with the exonic structure of the genes from an MP analysis which gave three equivalent trees. The informative limit of each gene is coded at the right of its name (*0:* absence; *1:* presence). *Open box:* loss of a limit; *filled box:* gain of a limit. The dendogram is arbitrarily rooted. The branch lengths make no sense.

compatibility method (Le Quesne 1969; Estabrook et al. 1976, program CLIQUE) were used to analyze the genomic structure.

*Construction of Figures.* The majority-rule consensus trees were constructed by the program DRAWGRAM. The results of CONSENSE analysis after the bootstrap resamplings are written in ovals on the node considered.

The determination of approximated time divergence between subunits (Fig. 5) is based on the NJ analysis of the Appendix 2 protein alignment. The external branches of the resulting tree showed an approximate molecular clock for each group. We are then able to determine the dates of the last duplications. However, the rates of evolution vary greatly between the subgroups, and the precocious duplications can't be calculated in this way. To determine the date of divergence between two subunits, we averaged the branch lengths and the evolution rate between all the **orthologs.** The estimated rate of evolution is obtained by dividing the branch length by the duration. The dates used are: *Torpedo*/osteichthyans 450 MYA, goldfish/Tetrapoda 405 MYA, *Xenopus*/Amniota 365 MYA, chicken/mammals 310 MYA, mouse/human 110 MYA (Benton 1990). For instance, each external branch inside the δ, ϵ, γ group provides an estimated rate of evolution of between $3.2 \times 10^{-10}$ and $7.3 \times 10^{-10}$ substitution per site and per year ($M = 5.3$ $\sigma = 1.2$), which is not far from a molecular clock.

## Results

### *Evolution of the Gene Structure*

The number of exons identified in the gene nAChR subunits varies largely in the family although some common features can be recognized (Fig. 1A). Four subfamilies can be identified on the basis of the genomic structure. These are: I—the *neuronal α-bungarotoxin-binding-site subunit* subfamily; II—*the Arthropoda neuronal subunit* subfamily; III—the *vertebrate neuronal nicotinic subunit* subfamily; and IV—the *muscle subunit* subfamily.

The genes of subfamily IV possess 11 or 12 exons, of which ten are conserved. In subfamily III, the main part of the coding sequence is distributed within a single exon. The structure of α1 and α7 genes differs from the two "holotypes" (III or IV). However, it is difficult to determine if the structure of these two genes is mainly **plesiomorph** or contains **autapomorphies.** In order to

extract information from genomic structure, we made a cladistic analysis of the frontiers between introns and exons. Only the ten informative sites were considered (Fig. 1). A parsimony analysis of two-state character (i.e., presence or absence of the frontier) gave three 12-step cladograms (summarized in Fig. 1B with an arbitrary root corresponding to sequence analyses; see below). A compatibility analysis gave the same results (although automatically rooted at a different point, i.e., between muscle and neuronal genes). One explanation is that the genic structure of type I subunits is mainly made of autapomorphies, whereas that of α1 is plesiomorphic in subfamily IV. On the basis of gene-structure analysis, this latter subunit would be a sister group of all other subunits of subfamily IV. However, sequence analyses (see below) make α1 a sister group of subfamily III. A translocation (e.g., between exons 4 and 5) could mask the real onset of the α1 subunit. Further studies of sequence homology between exons in **paralogs** will help to clarify this issue.

*Sequences Analysis*

The sequence analysis revealed the existence of the same four subfamilies of nAChR subunits as did analysis of the gene structure (Figs. 2 and 3). We obtained successive divergences of subfamilies I, then II and, at last III and IV. The position of subfamily II as a sister group of subfamily IV is weakly supported by NJ (Fig. 3) analysis and not by MP analysis (Fig. 2). These subunits appeared **polyphyletic** with MP analysis and **monophyletic** with NJ analysis. However, their position was supported by very weak bootstrap score. The β2 and β4 subunits were branched with subfamily IV with a weak bootstrap score. This position may be an artifact resulting from the precocious appearance and the weak divergence of these two subunits.

In subfamily IV, the β1 subunit diverged first followed by the δ. A subsequent duplication resulted in the γ and ϵ subunits. This latter duplication seems to have occurred shortly before the divergence of *Torpedo* subunits (i.e., before the divergence of the elasmobranch lineage).

In subfamily III, several groupings were present in more than 99% of trials: (β2, β4), (β3, α5), (α2, α4), (α3, α6). A first duplication gave β2 and β4. The position of β2 and β4 was unstable, jumping between subfamily III and subfamily IV according to the sequences sampled. However, if we consider the gene structure, the neuronal localization, and the pharmacological characteristics of these subunits, β2 and β4 have to be placed in subfamily III. Then we observe the separation of β3 and α5. At last, a monophyletic group formed by the two pairs (α4, α2) and (α3, α6) is present in MP and NJ analyses. The position of α1 does not match the gene-structure analysis. This strange position of α1 inside the

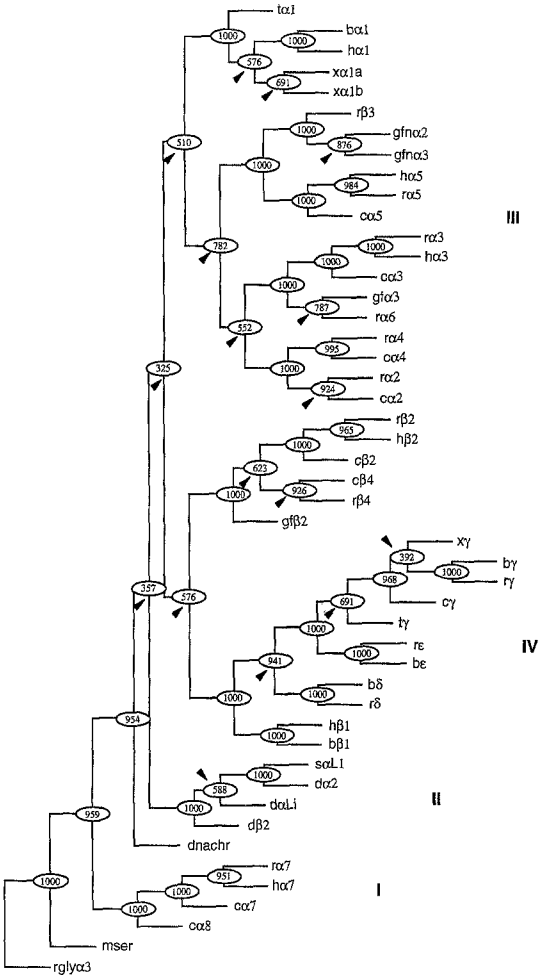


**Fig. 2.** Bootstrap majority-rule consensus tree obtained from 1000 MP replicates (SEQBOOT, PROTPARS and CONSENSE programs) with the alignment shown in Appendix 2. The nodes indicated by an *arrowhead* are uncertain.

neuronal subgroup is, however, weakly supported by bootstrap scores. The last clear duplications arose in the lineage of teleosteans which possess two homologs of β3 (appeared about 280 MYA), and in Tetrapoda, which have two homologs of the goldfish β2, β2, and β4 subunits. Moreover, goldfish α3 is not clearly homolog to tetrapod α3 (NJ Fig. 3) or α6 (MP Fig. 2), which could then appear only after the divergence of teleosteans.

*Evolution of the Stoichiometry*

As the nAChR is an oligomer formed by subunits coded by paralogs, it is reasonable to assume that the primitive receptor resulted from the assembly of just one subunit. Thus, the ability for a subunit to form functional homo-oligomers could reflect a plesiomorphic (''primitive'') mode of functioning. α-Bgt-sensitive homo-oligomers from *Locusta migratoria* (Breer et al. 1985) have been purified and reconstituted in vitro (Hanke and Breer 1986). The dαL1 subunit of *Schistocerca gregaria* forms

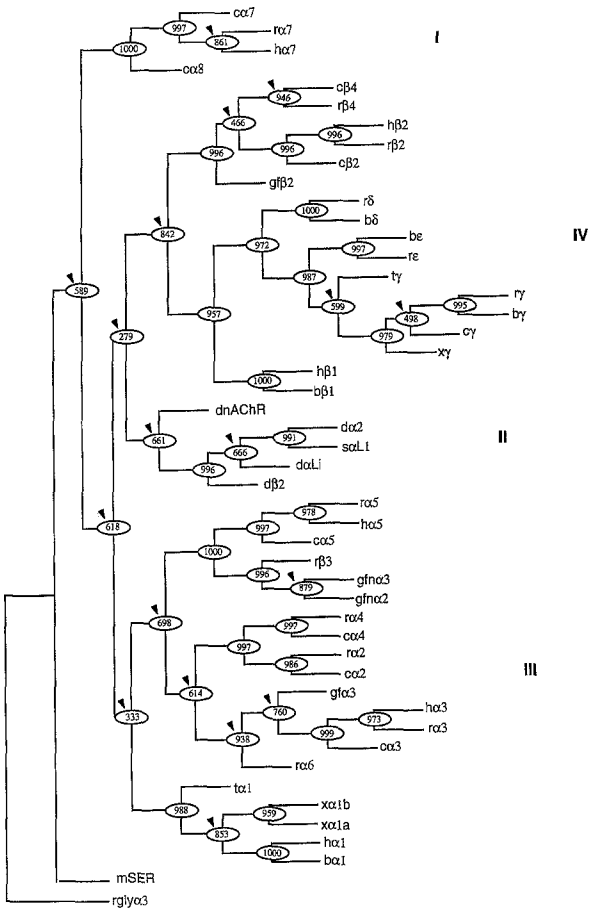


**Fig. 3.** Bootstrap majority-rule consensus tree obtained from 1000 NJ replicates (SEQBOOT, PROTDIST, NEIGHBOR, and CONSENSE programs) with the alignment shown in Appendix 2. The nodes indicated by an *arrowhead* are uncertain.

functional homo-oligomeric channels (Marshall et al. 1990) blocked by α-Bgt in vitro. dα2 (also called SAD for *second alpha subunit*), the putative *Drosophila* homolog of locust sαL1, forms functional receptors alone in *Xenopus* oocytes (Sawruk et al. 1990) though these receptors display an atypical pharmacology. In the same way, α7 from chicken is able to form homo-oligomers in *Xenopus* oocytes (Couturier et al. 1990; Revah et al. 1991; Anand et al. 1993). In contrast, vertebrate nAChR subunits from subfamilies III and IV, expressed in *Xenopus* oocyte, cannot form functional homo-oligomeric channels.

### *Pharmacological Argument for Monophyly of Neuronal Nicotinic Subfamily*

Although a small number of mutations sometimes suffice to dramatically change the properties of a receptor (e.g., Galzi et al. 1992), the pharmacological properties of the families of ligand-gated ion channels seem to diverge slowly. *Ascaris* muscle (Walker et al. 1992), *Aplysia* (Ono and Salvaterra 1981), αL1 (insect class 2, Marshall et al. 1990), dnAChR (insect class 1, Schloss et al. 1988), chicken α7 (Couturier et al. 1990; Anand et al. 1993), and vertebrate striated muscle (Lee and Chang 1966; Changeux et al. 1970) receptors are α-Bgt sensitive. Although α1 belongs to subfamily III, the functional α-Bgt sites of the vertebrate muscle receptor are formed partially by the subunits of subfamily IV. Moreover, if α1 is placed as a sister group of all other subunits of subfamily III (as indicated by the gene structure), the loss of α-Bgt sensitivity in the neuronal nicotinic subfamily is a **synapomorphy.** In addition, *Ascaris* muscle receptor (Walker et al. 1992), *Aplysia* neuronal receptors (Ono and Salvaterra 1981), and the receptor formed by sαL1 of *Schistocerca* (Marshall et al. 1990) are sensitive to strychnine, an antagonist of the glycine receptor, and to bicuculline, an antagonist of the $GABA_A$ receptor. α7 is also sensitive to strychnine (Anand et al. 1993). The members of a multigene family can then share pharmacological properties, even after a long divergence (probably more than 1,000 MYA here). Receptors of subfamily III do not seem to be blocked by these antagonists (Clément Léna, personal communication). Overall, the evidence from pharmacological studies further supports the notion of the monophyly of the subfamily III.

## Discussion

The analyses presented in this paper lead to the reconstruction of a global history of nAChR evolution. Although several nodes have not been perfectly resolved, the major relationships between subunits were clarified. Except for α1 and (β2, β4) all the analyses performed were congruent.

### *Hypothetically Missing Genes*

Subfamily I diverged before the split insects/vertebrates, and this subfamily could be present in insects. (The cloned insect subunits are orthologous to the subfamily III and IV.)

Neuromuscular transmission via nAChRs is known to occur in nematodes (Gerschenfeld 1973; Walker et al. 1992), annelids, molluscs (Gerschenfeld 1973), and vertebrates but not in insects and crustaceans. The chemical excitation of muscle in the Bilateria nonvertebrates/nonarthropods has to be mediated by subunits which do not belong to the subfamily IV (Fig. 2). Thus, neuromuscular transmission in vertebrates is not **homologous** to that occurring in other phyla.

The ϵ subunit seems to be present in the whole Gnathostomata phylum. This is consistent with the reported presence of an ϵ subunit in *Xenopus,* yet this subunit has not been cloned in chicken.

The bootstrap confirms that α7 and α8 diverged prior

to the separation of Sauropsida and Theropsida. Thus, an α8 subunit may be present in mammals.

### *Reconstructed History of the nAChR-Subunit Gene Family*

Based on present and previous results, the history of the nAChR-subunit gene family can be reconstructed as follows (Fig. 5):

We can plausibly assume (still without proof) that in the primitive metazoans (e.g., coelenterates) nAChR was made of a single subunit able to form homo-oligomers. The coelenterates have no true muscle cells but already have multipolar neurons, of ectodermal origin. This first nAChR presumably had a neuronal localization. With the appearance of a third embryonic sheet, the nAChR acquired a novel role in neuromuscular transmission. However, if the nematodes and the molluscs have a muscle nAChR, it is not homologous to the vertebrate subfamily IV. Indeed, this latter plausibly appeared after the differentiation of Deuterostomata. The subunit cloned in *Onchocerca* does not possess the third loop of the ACh binding site (Devillers-Thiéry et al. 1993) and might be a non-α subunit. The NJ analysis of the alignment shown in appendix 3 (Fig. 4) determined this subunit to be an extra group of three *Drosophila* subunits containing two α and one non-α subunits. The idea of a precocious emergence of the insect subunits is supported by the ability of these subunits to form homo-oligomers in vitro. Assuming that alphaL1 of *Schistocerca* and alpha2 of *Drosophila* are orthologs, the duplication between them and dαLi is older than the divergence of Orthoptera and Diptera—i.e., older than 300 MYA (Labandeira and Sepkoski 1993). In Deuterostomata, several duplications occurred to give extant subfamily IV, which was complete in vertebrate phylum before the appearance of chondrichthyes (450 MYA) and extant subfamily III, one of the paralogs being expressed in muscle.

Several tissular switches of expression from neuron to muscle or from muscle to neuron can be hypothetized (Fig. 5). Between the divergence of subfamily II and the split subfamily III/subfamily IV (in the chordate lineage), one switch of expression might have given a muscle receptor, possibly homopentameric. After the first duplication between the ancestor of subfamily IV and α1 (a duplication which is responsible of the heteromeric muscle receptor), a further duplication from α1 provided a new gene, which expression became neuronal. The evolution of the promoters (and of the transcription regulators) may thus have played a role as important as gene duplication in the diversification of the nAChR family.

The neuronal non-α subunit group is likely to be polyphyletic, whereas the neuronal α subunits (the "binding subunits") would form a monophyletic group. α5, which lacks some important aromatic amino acids in the third

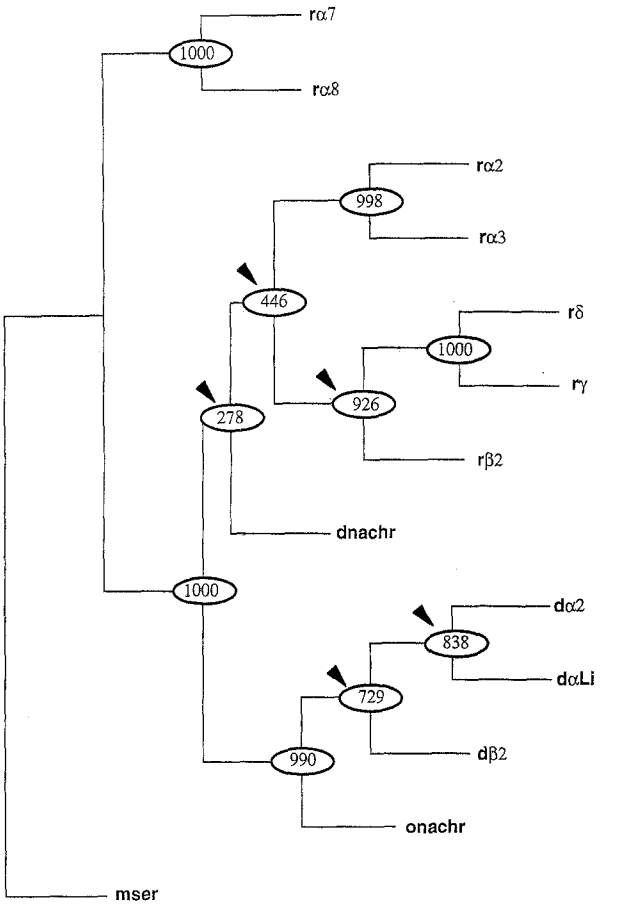


**Fig. 4.** Bootstrap majority-rule consensus tree obtained from 1,000 NJ replicates with the alignment shown in Appendix 3, presenting the possible emergence of the nematode subunit. The nodes indicated by an *arrowhead* are uncertain.

loop of the ACh binding site, cannot form functional receptors in vitro with any β subunit (Boulter et al. 1990) but is coprecipitated from endogeneous material with other α subunits (Conroy et al. 1992; Vernallis et al. 1993); it thus may represent a new type of "structural" subunit and should therefore be given another name. α5 and β3 could be called γ2 and γ3. Then the three types of subfamily III subunits could form monophyletic groups—the tribes α, β, and γ.

### *Growth of the Neuronal Nicotinic Subfamily and Increase in Complexity of the Cholinergic System*

The multiple duplications in subfamily III parallel the progressive increased complexity of the chordate nervous system—in particular, of the cholinergic system. At the beginning of the evolution of this phylum, one subunit was plausibly present in the nervous system, resulting from the duplication of α1 (in addition to the ancestor of α7 and α8). The diversity of the group increased during the first 400 MY, until the appearance of Tetrapoda. The whole evolution of the subfamily occurred in

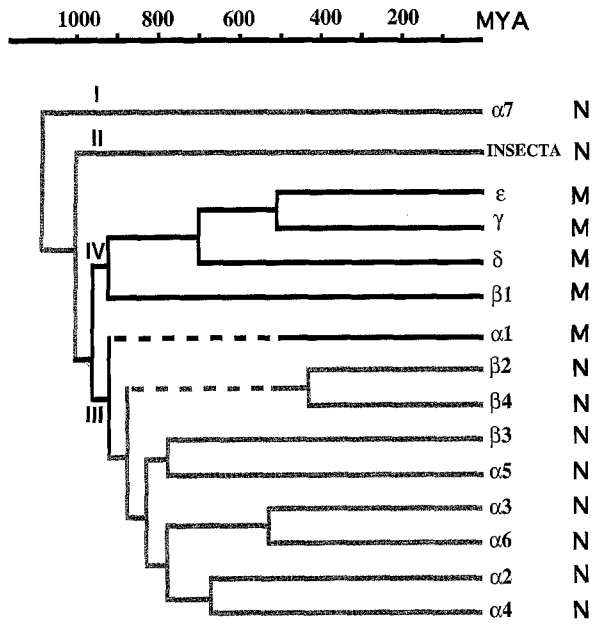


**Fig. 5.** Summary tree, integrating the results of the whole study. The dates of the last divergences have been calculated from the protein alignment of Appendix 2 (cf. Materials and Methods). α7/α8: 380 MYA; ε/γ: 508 MYA; ε,γ/δ: 711 MYA; ε,γ,δ/β1: 926 MYA; α3/α6: 529 MYA; α2/α4: 669 MYA; α5/β3: 770 MYA. The ages of the precocious divergences have been approximately inferred from the divergence of nematodes (1,000 MYA: Vanfleteren et al. 1994). *M:* muscle subunit; *N:* neuronal subunit. The *gray branches* represent subunits putatively expressed in neurons. The *black branches* represent subunits possibly expressed in muscle.

the first half of Deuterostomata history (from 600–800 MYA to about 300 MYA). In the first prechordate fossils, we find only two ganglia, the peripheral and the cerebroid ganglia. *Branchiostoma* has only one pseudovesicle in the head. Spinal chord and cholinergic peripheral nervous system were present early in the vertebrate lineage (although the complete autonomous system was reached only in mammals). Lamprey already has five vesicles but the main development of the brain and particularly of the forebrain occurred in Gnathostomata.

In situ hybridization (Deneris et al. 1989; Wada et al. 1989, 1990; Zoli et al. 1995) as well as immunohistochemical (Britto et al. 1992; Hill et al. 1993) studies have shown that, in rat and chicken brain, α4 and β2 mRNA distribution is diffuse, whereas α2, α3, α5, α6, β3, and β4 are mainly restricted to a few major cholinergic or cholinoceptive pathways, which, however, also express α4 and β2. β2 has diverged early in the neuronal subfamily history (Figs. 2, 3, and Results). α4 and β2 could represent a "fossil" expression, which was present in most areas of the ancestral brain. When a duplication occurred, one of the paralogs kept the specific role of the "father gene," whereas the other paralog had to acquire a new role. This role can be defined by a new domain of expression (like the switches muscle/neuron developed before) or by a modified function. The hypothesis that α4 maintained its previous role while another paralog acquired a new role is supported by some evidence on the α2 subunit. In chick brain, the α2 subunit is restricted to the lateral spiriform nucleus (Daubas et al. 1990) but, in rat brain it is restricted to the interpeduncular nucleus (Wada et al. 1989)—a nonhomolog structure. (The interpeduncular nucleus also exists in the chick brain.) Moreover, there is no homolog of the lateral spiriform nucleus in rat brain, a fact that points to the genesis of this structure after the divergence of the bird lineage. It is attractive to suppose that a gene duplication occurred a short time before the branching of Theropsida and Sauropsida (i.e., before 310 MYA, Benton 1990). This time would have been too short to define the specificity of α2 (in contrast to α4, which maintained its ancient role). Then two independent specificities of expression took place in the two phyla. Accordingly, a transgene with the avian α2 gene (including the promoter) is expressed throughout the rat brain, mostly in cholinergic structures (motor nuclei and basal telencephalon) (Daubas et al. 1993). (Nevertheless, this distribution corresponds neither to the distribution of endogenous α2 nor to that of α4.)

In the same way, the duplications α3/α6 and β2/β4 occurred a little before, or a little after, the split between the teleost and the tetrapod lineages. In the rat brain, α3 and β4 are mainly expressed in the medial habenula, a cholinergic and cholinoceptive structure. However, in a teleost fish (*Phoxinus phoxinus*), an immunocytochemical study did not find any cholinergic cell and found only a small number of cholinergic fibers in the habenula (Ekström 1987). If these characteristics are plesiomorph, there could be again a correlation between gene duplications and a further change of function. α4 and β2 could thus have kept the ancestor role of the neuronal nAChR, whereas other paralogs could have found new functional specificities in the evolving cholinergic systems.

We have shown, on the basis of cladograms and phenograms, that the first duplications in the nAChR occurred before the divergence of nematodes. Several nAChR subfamilies were identified. There is congruence between sequence and gene-structure analyses, and the three subfamilies present in vertebrates correspond closely to the functional subgroups (described from anatomical, pharmacological, and structural considerations). Two phenomena seem to have generated the wealth of the family. First, several switches of expression seem to have occurred from neuron to muscle and the opposite. Second, multiple gene duplications gave the extant number of paralogs. The neuronal nicotinic-subunit subfamily (type III subunits) appeared at the beginning of the chordate phylum and grew until the separation of Sauropsida and Theropsida lineages. This diversification, both quantitative and functional, paralleled the increase in complexity of the cholinergic systems. A link between an increased combinatorial complexity of subunit combinations and a larger plasticity in the functioning of these pathways is plausible.

*Acknowledgments.* We thank Clément Léna, Pascal Tassy, and Fredj Tékaia for helpful advice. We also thank Richard Miles, Michele Zoli, and the anonymous referee for critical reading of the manuscript. This work was supported by the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale (contract No. 872004), the Collège de France, the Direction des Recherches et Etudes Techniques (contract No. 90/142), and the EEC (contract no. 94-1060).

## References

Anand R, Conroy WG, Schoepfer R, Whiting P, Lindstrom J (1991) Neuronal nicotinic acetylcholine receptors expressed in Xenopus oocytes have a pentameric quaternary structure. J Biol Chem 266: 11192–11198

Anand R, Peng X, Lindstrom J (1993) Homomeric and native alpha7 acetylcholine receptors exhibit remarkably similar but non-identical pharmacological properties, suggesting that the native receptor is a heteromeric protein complex. FEBS Lett 327:241–246

Benton MJ (1990) Phylogeny of the major tetrapod groups: morphological data and divergence dates. J Mol Evol 30:409–424

Bertrand D, Galzi JL, Devillers-Thiéry A, Bertrand D, Changeux JP (1993) Stratification of the channel domain in neurotransmitter receptors. Curr Opin Cell Biol 5:688–693

Boulter J, O'Shea-Greenfield A, Duvoisin R, Connolly JG, Wada E, Jensen A, Gardner PD, Ballivet M, Deneris ES, McKinnon D, Heinemann Patrick J (1990) α3, α5, and β4: three members of the rat neuronal nicotinic acetylcholine receptor-related gene family form a gene cluster. J Biol Chem 265:4472–4482

Breer H, Kleene R, Hinz G (1985) Molecular forms and subunit structure of the acetylcholine receptor in the central nervous system of insects. J Neurosci 5:3386–3392

Brehm P, Okamura Y, Mandel G (1991) Ion channel evolution. Semin Neurosci 3:355–367

Britto LRG, Keyser KT, Lindstrom JM, Karten HJ (1992) Immunohistochemical localization of nicotinic acetylcholine receptor subunits in the mesencephalon and diencephalon of the chick (*Gallus gallus*). J Comp Neurol 317:325–340

Changeux JP, Kasai M, Lee CY (1970) The use of snake venom toxin to characterize the cholinergic receptor protein. Proc Natl Acad Sci USA 67:1241–1247

Changeux JP (1990) Functional architecture and dynamics of the nicotinic acetylcholine receptor: an allosteric ligand-gated ion channel. Fidia Res Found Neurosci Award Lectures 4:21–168

Clarke PBS, Schwartz RD, Paul SM, Pert CB, Pert A (1985) Nicotinic binding in rat brain: autoradiographic comparison of [$^3$H]acetylcholine, [$^3$H]nicotine and [$^{125}$I]α-bungarotoxin. J Neurosci 5:1307–1315

Cockcroft VB, Osguthorpe DJ, Barnard EA, Friday AE, Lunt GG (1992) Ligand-gated channels. Homology and diversity. Mol Neurobiol 4:129–169

Conroy G, Vernallis AB, Berg DK (1992) The α5 gene product assembles with multiple acetylcholine receptor subunits to form distinctive receptor subtypes in brain. Neuron 9:679–691

Cooper E, Couturier S, Ballivet M (1991) Pentameric structure and subunit stoichiometry of a neuronal acetylcholine receptor. Nature 350:235–238

Couturier S, Bertrand D, Matter JM, Hernandez MC, Bertrand S, Millar N, Valera S, Barkas T, Ballivet M (1990) A neuronal nicotinic acetylcholine receptor subunit (a7) is developmentally regulated and forms a homo-oligomeric channel blocked by a-BTX. Neuron 5:847–856

Darlison MG, Hutton ML, Harvey RJ (1993) Molluscan ligand-gated ion-channel receptors. In: Pichon Y (ed) EXS 63, comparative molecular neurobiology. Birkhäuser, Basel, pp 48–64

Daubas P, Devillers-Thiéry A, Geoffroy B, Martinez S, Bessis A, Changeux JP (1990) Differential expression of the neuronal acetylcholine receptor α2 subunit gene during chick brain development. Neuron 5:49–60

Daubas P, Salmon AM, Zoli M, Geoffroy B, Devillers-Thiéry A, Bessis A, Médevielle F, Changeux JP (1993) Chicken neuronal acetylcholine receptor α2-subunit gene exhibits neuron-specific expression in the brain and spinal cord of transgenic mice. Proc Natl Acad Sci USA 90:2237–2241

Dayhoff MO (1979) Atlas of protein sequence and structure, vol 5, supplement 3, 1978. National Biomedical Research Foundation, Washington DC

Deneris ES, Boulter J, Swanson LW, Patrick J, Heinemann S (1989) β3: a new member of nicotinic acetylcholine receptor gene family is expressed in brain. J Biol Chem 264:6268–6272

Devillers-Thiéry A, Galzi JL, Eiselé JL, Bertrand S, Bertrand D, Changeux JP (1993) Functional architecture of the nicotinic acetylcholine receptor: a prototype of ligand-gated ion channels. J Memb Biol 136:97–112

Eck RV, Dayhoff MO (1966) Atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, MD

Ekström P (1987) Distribution of choline acetyltransferase-immunoreactive neurons in the brain of a cyprinid teleost (*Phoxinus phoxinus* L). J Comp Neurol 256:494–515

Estabrook GF, Johnson CS Jr, McMorris FR (1976) A mathematical foundation for the analysis of character compatibility. Math Biosci 23:181–187

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Felsenstein J (1993) PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Fitch WW (1971) Toward defining the course of evolution: minimum change for a specified tree topology. Syst Zool 20:406–416

Flemming JT, Tornoe C, Riina HA, Coadwell J, Lewis JA, Sattelle DB (1993) Acetylcholine receptor molecules of the nematode *Caenorhabditis elegans*. In: Pichon Y (ed) EXS 63, comparative molecular neurobiology. Birkhäuser, Basel, pp 65–80

Galzi JL, Devillers-Thiéry A, Hussy N, Bertrand S, Changeux JP, Bertrand D (1992) Mutations in the channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic. Nature 359:500–505

Galzi JL, Changeux JP (1994) Ligand-gated ion channel as unconventional allosteric proteins. Curr Opin Struct Biol 4:554–565

Gerschenfeld HM (1973) Chemical transmission in invertebrate central nervous systems and neuromuscular junctions. Physiol Rev 53:1–119

Greenberg ME, Ziff EB, Greene LA (1986) Stimulation of neuronal acetylcholine receptors induces rapid gene transcription. Science 234:80–83

Gundelfinger ED (1992) How complex is the nicotinic receptor system of insects? TINS 15:206–211

Hanke W, Breer H (1986) Channel properties of an insect neuronal acetylcholine receptor protein reconstituted in planar lipid bilayers. Nature 321:171–174

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237–244

Hill JA, Zoli M, Bourgeois JP, Changeux JP (1993) Immunocytochemical localization of a neuronal nicotinic receptor: The β2 subunit. J Neurosci 13:1551–1568

Jonas P, Baumann A, Merz B, Gundelfinger ED (1990) Structure and developmental expression of the Da2 gene encoding a novel nicotinic acetylcholine receptor protein of *Drosophila melanogaster*. FEBS Lett 269:264–268

Karlin A (1993) Structure of nicotinic acetylcholine receptors. Curr Opin Neurobiol 3:299–309

Labandeira CC, Sepkoski JJ Jr (1993) Insect diversity in the fossil record. Science 261:310–315

Lake JA (1990) Origin of the metazoa. Proc Natl Acad Sci USA 87: 763–766

Lee CY, Chang CC (1966) Modes of actions of purified toxins from elapid venoms on neuro-muscular transmission. Mem Inst Butantan Sao Paulo 33:555–572

Leech CA, Sattelle DB (1993) Acetylcholine receptor/channel of insects. In: Pichon Y (ed) EXS 63, Comparative molecular neurobiology. Birkhäuser, Basel, pp 81–97

Le Quesne WJ (1969) A method of selection of characters in numerical taxonomy. Syst Zool 18:201–205

Luetje CW, Patrick J (1991) Both α- and β-subunits contribute to the agonist sensitivity of the neuronal nicotinic acetylcholine receptor. J Neurosci 11:837–845

Marshall J, Buckingham SD, Shingai R, Lunt GG, Goosey MW, Darlison MG, Satelle DB, Barnard EA (1990) Sequence and functional expression of a single α subunit of an insect nicotinic receptor. EMBO J 9:4391–4398

Mulle C, Vidal C, Benoît P, Changeux JP (1991) Existence of different subtypes of nicotinic acetylcholine receptors in the rat habenulo-interpeduncular system. J Neurosci 11:2588–2597

Ono JK, Salvaterra PM (1981) Snake alpha-toxin effects on cholinergic and noncholinergic responses of Aplysia californica neurons. J Neurosci 1:259–270

Revah F, Bertrand D, Galzi JL, Devillers-Thiéry A, Mulle C, Hussy N, Bertrand S, Ballivet M, Changeux JP (1991) Mutations in the channel domain alter desensitization of a neuronal nicotinic receptor. Nature 353:846–849

Role LW (1992) Diversity in primary structure and function of neuronal nicotinic acetylcholine receptor channels. Curr Opin Neurobiol 2:254–262

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sargent PB (1993) The diversity of neuronal nicotinic acetylcholine receptors. Annu Rev Neurosci 16:403–443

Sawruk E, Schloss P, Betz H, Schmitt B (1990) Heterogeneity of *Drosophila* nicotinic receptors: SAD, a novel developmentally regulated α-subunit. EMBO J 9:2671

Schloss P, Hermans-Borgmeyer I, Betz H, Gundelfinger ED (1988) Neuronal acetylcholine receptor in *Drosophila:* the ARD protein is a component of a high affinity α-bungarotoxin binding complex. EMBO J 7:2889–2894

Schoepfer R, Conroy WG, Whiting P, Gore M, Lindstrom J (1990) Brain α-bungarotoxin binding protein cDNAs and MAbs reveal subtypes of this branch of the ligand-gated ion channel gene superfamily. Neuron 5:35–48

Segerberg MA, Stretton AOW (1993) Actions of cholinergic drugs in the nematode *Ascaris suum.* Complex pharmacology of muscle and motorneurons. J Gen Physiol 101:271–296

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Vanfleteren JR, Van de Peer Y, Blaxter ML, Tweedie SAR, Trotman C, Lu L, Van Hauwaert ML, Moens L (1994) Molecular genealogy of some nematode taxa as based on cytochrome c and globin amino acid sequences. Mol Phylogen Evol 3:92–101

Vernallis AB, Conroy WG, Berg DK (1993) Neurons assemble acetylcholine receptors with as many as three kinds of subunits while maintaining subunit segregation among receptor subtypes. Neuron 10:451–464

Wada K, Ballivet M, Boulter J, Connolly J, Wada E, Deneris ES, Swanson LW, Heinemann S, Patrick J (1988) Functional expression of a new pharmacological subtype of brain nicotinic acetylcholine receptor. Science 240:330–334

Wada E, Wada K, Boulter J, Deneris E, Heinemann S, Patrick J, Swanson LW (1989) Distribution of α2, α3, α4 and β2 neuronal nicotinic receptor subunit mRNAs in the central nervous system: a hybridation histochemical study in the rat. J Comp Neurol 284: 314–335

Wada E, McKinnon D, Heinemann S, Patrick J, Swanson LW (1990) The distribution of mRNA encoded by a new member of the neuronal nicotinic acetylcholine receptor gene family (α5) in the rat central nervous system. Brain Res 526:45–53

Walker RJ, Colquhoun L, Holden-Dye L (1992) Pharmacological profiles of the GABA and acetylcholine receptors from the nematode, *Ascaris suum.* Acta Biol Hung 43:59–68

Whiting P, Schoepfer R, Lindstrom J, Priestley T (1991) Structural and pharmacological characterization of the major brain nicotinic acetylcholine receptor subtype stably expressed in mouse fibroblast. Mol Pharmacol 40:463–472

Wilbur WJ, Lipman DJ (1983) Rapid similarity searches of nucleic acid and protein data banks. Proc Natl Acad Sci USA 80:726–730

Zoli M, Le Novère N, Hill JA, Changeux JP (1995) Developmental regulation of nicotinic receptor subunit mRNAs in the rat central and peripheral nervous system. J Neurosci (in press)

## Appendix 1: Glossary of the Cladistic Terms Present in the Text

**Homology:** Two similar characters are homologous when they come from a common ancestor. There is homology between two genes if they arose by duplication. Characters are homologous if (1) they look like each other, (2) they do not coexist in the same organism (the arms and the wings of angels), and (3) if they provide the same phylogenies as other homologous characters.

**Ortholog:** Two orthologs are homologs that have arisen by speciation (the "same" gene in two different species).

**Paralog:** Two paralogs are homologs that have arisen by duplication in the same organism.

**Monophyly:** A monophyletic group is formed by one ancestor and all its descendants.

**Polyphyly:** A polyphyletic group is formed by different subgroups, which do not share any common ancestor belonging to the group.

A **plesiomorphy** is an ancestor character. It is not useful in reconstructing phylogenies.

An **autapomorphy** is a derived character, present in one descendant only. It is not useful in reconstructing phylogenies

A **synapomorphy** is a derived character, shared by a monophyletic group of several descendants. The synapomorphies are the only useful characters with which to infer phylogenies

**Informative site:** A character is informative if it exists under at least two states present twice. A character presenting everytime the same state as well as a character different in every compared object are not informative

```
bα1     ETRLVAKLFE D--YNSVVRP VEDHRQAVEV TVGLQLIQLI NVDEVNQIVT
bβ1     EGRLREKLFS --GYDSTVRP AREVGDRVWV SIGLTLAQLI SLNEKDEEMS
bδ      EERLIRHLFE EKAYNKELRP AAHKESV-EI SLALTLSNLI SLKEVEETLT
bε      ELRLYHYLFD --TYDPGRRP VQEPEDTVTI SLKVTLTNLI SLNEKEETLT
bγ      EERLLGDLMQ --GYNPHLRP AEHDSDVVNV SLKLTLTNLI SLNEREEALT
cα2     EDRLFKHLFT G--YNRWSRP VPNTSDVVIV KFGLSIAQLI DVDEKNQMMT
cα3     EHRLYAALFK N--YNQFVRP VKNASDPVII QFEVSMSQLV KVDEVNQIME
cα4     EERLLKKLFS G--YNKWSRP VANISDVVLV RFGLSIAQLI DVDEKNQMMT
cα5     EDRLFKHLFE D--YQRWVRP VEHLNDTIKI KFGLAISQLV DVDEKNQLMT
cα7     QRKLYKELLK N--YNPLERP VANDSQPLTV YFTLSLMQIM DVDEKNQVLT
cα8     QRRLYRDLLR N--YNRLERP VMNDSQPIVV ELQLSLLQII DVDEKNQVLI
cβ2     EERLVEYLLD PTRYNKLIRP ATNGSQLVTV QLMVSLAQLI SVHEREQIMT
cβ4     EEKLMNHLLS PDRYNKLIRP AVNSSQLVSI ELQVSLAQLI SVNEREQIMT
cγ      EEKLLQDLMT --NYNRHLRP ALRGDQVIDV TLKLTLTNLI SLNEREETLT
dα2     AKRLYDDLLS N--YNRLIRP VSNNTDTVLV KLGLRLSQLI DLNLKDQILT
dαLi    AKRLYDDLLS N--YNRLIRP VGNNSDRLTV KMGLRLSQLI DVNLKNQIMT
dβ2     TKRLYDDLLS N--YNRLIRP VVNNTETLTV WLGLKLSQLI EVNLKNQVMT
dnachr  EERLVRDLFR G--YNKLIRP VQNMTQKVGV RFGLAFVQLI NVNEKNQVMK
gfα3    EDRLFRRLFR R--YNQFIRP VENVSDPVTV EFEVSISQLV KVDEVNQIME
gfβ2    LRS--DFLLG PERYNKLIRP AVNKSQQVTI GIKVSLAQLI SVNEREQIMT
gfnα2   EDALLRELFQ G--YQRWVRP VQHANHSVKV RFGLKISQLV DVDEKNQLMT
gfnα3   EDTLLRNLFR G--YQKWVRP ILHANDTITV RFGLKISQLV DVDEKNHLMT
hα1     ETRLVAKLFK D--YSSVVRP VEDHRQVVEV TVGLQLIQLI NVDEVNQIVT
hα3     EHRLFERLFE D--YNEIIRP VANVSDPVII HFEVSMSQLV KVDEVNQIME
hα5     EDSLLKDLFQ D--YERWVRP VEHLNDKIKI KFGLAISQLV DVDEKNQLMT
hα7     QRKLYKELVK N--YNPLERP VANDSQPLTV YFSLSLLQIM DVDEKNQVLT
hβ1     EGRLREKLFS --GYDSSVRP AREVGDRVRV SVGLILAQLI SLNEKDEEMS
hβ2     EERLVEHLLD PSRYNKLIRP ATNGSELVTV QLMVSLAQLI SVHEREQIMT
mser    LLRLSDHLLA --NYKKGVRP VRDWRKPTTV SIDVIMYAIL NVDEKNQVLT
rα2     EDRLFKHLFG G--YNRWARP VPNTSDVVIV RFGLSIAQLI DVDEKNQMMT
rα3     EHRLFQYLFE D--YNEIIRP VANVSHPVII QFEVSMSQLV KVDEVNQIME
rα4     EERLLKRLFS G--YNKWSRP VGNISDVVLV RFGLSIAQLI DVDEKNQMMT
rα5     EDSLFRDLFE D--YERWVRP VEHLSDKIKI KFGLAISQLV DVDEKNQLMT
rα6     EEQLFHTLFA H--YNRFIRP VENVSDPVTV HFELAITQLA NVDEVNQIME
rα7     QRRLYKELVK N--YNPLERP VANDSQPLTV YFSLSLLQIM DVDEKNQVLT
rβ2     EERLVEHLLD PSRYNKLIRP ATNGSELVTV QLMVSLAQLI SVHEREQIMT
rβ3     EDALLRHLFQ G--YQKWVRP VLNSSDIIKV YFGLKISQLV DVDEKNQLMT
rβ4     EEKLMDDLLN KTRYNNLIRP ATSSSQLISI RLELSLSQLI SVNEREQIMT
rδ      EQRLIQHLFE EKGYNKELRP VARKEDIVDV ALSLTLSNLI SLKEVEETLT
rε      ELSLYHHLFD --NYDPECRP VRRPEDTVTI TLKVTLTNLI SLNEKEETLT
rγ      EERLLADLMR --NYDPHLRP AERDSDVVNV SLKLTLTNLI SLNEREEALT
rglyα3  SDFLDKLMGR TSGYDARIRP -NFKGPPVNV TCNIFINSFG SIAETTMDYR
sαl1    AKRLYDDLLS N--YNRLIRP VSNNTDTVLV KLGLRLSQLI DLNLKDQILT
tα1     ETRLVANLLE N--YNKVIRP VEHHTHFVDI TVGLQLIQLI SVDEVNQIVE
tγ      EGRLIEKLLG --DYDKRIIP AKTLDHIIDV TLKLTLTNLI SLNEKEEALT
xα1a    ESRLINDLFK S--YNKVVRP VKAFKDKVVV TVGLQLIQLI NVNEVNQIVT
xα1b    ETRLIGDLFA N--YNKVVRP VETYKDQVVV TVGLQLIQLI NVDEVNQIVS
xγ      EERLLNDLMK --NYNKNLRP VEKDGDIISV SIKLTLTNLI SLNEKEEALT

bα1     TNVRLKQQWV DYNLKWNPDD YGGVKKIHIP SEKIWRPDLV LYNNADGDFA
bβ1     TKVYLDLEWT DYRLSWDPEE HEGIDSLRIS AESVWLPDVV LLNNNDGNFD
bδ      TNVWIEQGWT DSRLQWDAED FGNISVLRLP ADMVWLPEIV LENNNDGSFQ
bε      TSVWIGIDWQ DYRLNYSKGD FGGVETLRVP SELVWLPEIV LENNIDGQFG
bγ      TNVWIEMQWC DYRLRWDPRD YGGLWVLRVP STMVWRPDIV LENNVDGVFE
cα2     TNVWLKQEWS DYKLRWNPED FDNVTSIRVP SEMIWIPDIV LYNNADGEFA
cα3     TNLWLKHIWN DYKLRWNPVD YGGAEFIRVP SGQIWKPDIV LYNNAVGDFQ
cα4     TNVWVKQEWH DYKLRWDPQE YENVTSIRIP SELIWRPDIV LYNNADGDFA
cα5     TNVWLKQEWI HVKLRWNPED YAGITSIRVP SDSIWIPDIV LYDNADGRFE
cα7     TNIWLQMYWT DHYLQWNVSE YPGVKNVRFP DGLIWKPDIL LYNSADERFD
cα8     TNAWLQMYWV DIYLSWDQYE YPGVQNLRFP SDQIWVPDIL LYNSADERFD
cβ2     TNVWLTQEWE DYRLTWKPED FDNMKKVRLP SKHIWLPDVV LYNNADGMYE
cβ4     TNVWLNQEWI DYRLAWKPSD YEGINMLRIP AKHIWLPDIV LYNNADGTYE
cγ      TNVWIEMQWS DYRLRWDPDK YDDIQQLRVP SAMVWLPDIV LENNIDGTFE
dα2     TNVWLEHEWQ DHKFKWDPSE YGGVTELYVP SEHIWLPDIV LYNNADGEYV
dαLi    TNVWVEQEWN DYKLKWNPDD YGGVDTLHVP SEHIWHPDIV LYNNADGNYE
dβ2     TNLWVKQRWF DYKLRWDPEE YGGVEQLYVP SEHIWVPDIV LYNNWDGNYE
dnachr  SNVWLRLVWY DYQLQWDEAD YGGIGVLRLP PDKVWKPDIV LFNNADGNYE
gfα3    TNLWLRHIWN DYKLKWLPAE FDGIEFIRVP SNKIWRPDIV LYNNAVGDFL
```

**Appendix 2:** Alignment of 48 nAChR Subunits Made by the CLUSTAL V Software—Total Sites: 394; Informative Sites: 357

| | | | | |
|---|---|---|---|---|
| gfβ2 | TNVWLTQEWT | DYRLVWDPNE | YEGIKKLRIP | SQHIWLPDIV LYNNADGVYE |
| gfnα2 | TNVWLWQEWL | DYKLRWNPEN | YGGITSIRVP | SESIWLPDIV LYENADGRFE |
| gfnα3 | TNVWLWQEWT | DYKLRWNPED | YGGITSIRVP | SETIWLPDIV LYENADGRFE |
| hα1 | TNVRLKQQWV | DYNLKWNPDD | YGGVKKIHIP | SEKIWRPDVV LYNNADGDFA |
| hα3 | TNLWLKQIWN | DYKLKWNPSD | YGGAEFMRVP | AQKIWKPDIV LYNNAVGDFQ |
| hα5 | TNVWLKQEWI | DVKLRWNPDD | YGGIKVIRVP | SDSSWTPDIV LFDNADGRFE |
| hα7 | TNIWLQMSWT | DHYLQWNVSE | YPGVKTVRFP | DGQIWKPDIL LYNSADERFD |
| hβ1 | TKVYLDLEWT | DYRLSWDPAE | HEGIDSLRIT | AESVWLPDVV LLNNNDGNFD |
| hβ2 | TNVWLTQEWE | DYRLTWKPEE | FDNMKKVRLP | SKHIWLPDVV LYNNADGMYE |
| mser | TYIWYRQYWT | DEFLQWTPED | FDNVTKLSIP | TDSIWVPDIL INEFVDVG-K |
| rα2 | TNVWLKQEWN | DYNVRWDPAE | FGNVTSLRVP | SEMIWIPDIV LYNNADGEFA |
| rα3 | TNLWLKQIWN | DYKLKWKPSD | YQGVEFMRVP | AEKIWKPDIV LYNNADGDFQ |
| rα4 | TNVWVKQEWH | DYKLRWDPGD | YENVTSIRIP | SELIWRPDIV LYNNADGDFA |
| rα5 | TNVWLKQEWI | DVKLRWNPDD | YGGIKIIRVP | SDSLWIPDIV LFDNADGRFE |
| rα6 | TNLWLRHVWK | DYRLCWDPTE | YDGIETLRVP | ADNIWKPDIV LYNNAVGDFQ |
| rα7 | TNIWLQMSWT | DHYLQWNMSE | YPGVKNVRFP | DGQIWKPDIL LYNSADERFD |
| rβ2 | TNVWLTQEWE | DYRLTWKPQH | FDNMKKVRLP | SKHIWLPDVV LYNNADGMYE |
| rβ3 | TNVWLKQEWT | DQKLRWNPEE | YGGINSIKVP | SESLWLPDIV LFENADGRFE |
| rβ4 | TSIWLKQEWT | DYRLAWNSSC | YEGVNILRIP | AKRVWLPDIV LYNNADGTYE |
| rδ | TNVWIDHAWI | DSRLQWNANE | FGNITVLRLP | SDMVWLPEIV LENNNDGSFQ |
| rε | TSVWIGIEWQ | DYRLNFSKDD | FAGVEILRVP | SEHVWLPEIV LENNIDGQFG |
| rγ | TNVWIEMQWC | DYRLRWDPKD | YEGLWILRVP | STMVWQPDIV LGNNVDGVFE |
| rglyα3 | VNIFLRQKWN | DPRLAYSEYP | DDSLDLDPSM | LDSIWKPDLF FANEKGANFH |
| sαl1 | TNVWLEHEWQ | DHKFRWDPAE | YGGVTELYVP | SEHIWLPDIV LYNNADGEYV |
| tα1 | TNVRLRQQWI | DVRLRWNPAD | YGGIKKIRLP | SDDVWLPDLV LYNNADGDFA |
| tγ | TNVWIEIQWN | DYRLSWNTSE | YEGIDLVRIP | SELLWLPDVV LENNVDGQFE |
| xα1a | TNVRLKQQWE | DVHLKWDPED | YGGIKKVRIP | SSDIWRPDIV LYNNADGDFA |
| xα1b | TNIRLKQQWR | DVNLKWDPAK | YGGVKKIRIP | SSDVWSPDLV LYNNADGDFA |
| xγ | TNVWVEMQWK | DYRLSWDPND | YHGISMMRIP | STSVWLPDVG LENNVDGTFD |

| | | | | |
|---|---|---|---|---|
| bα1 | IVKFTKVLLD | YTGHITWTPP | AIFKSYCEII | VTHFPFDEQN CSMKLGTWTY |
| bβ1 | VALDINVVVS | SDGSMRWQPP | GIYRSSCSIQ | VTYFPFDWQN CTMVFSSYSY |
| bδ | ISYSCNVLIY | PSGSVYWLPP | AIFRSSCPIS | VTYFPFDWQN CSLKFSSLKY |
| bε | VAYEANVLVS | EGGYLSWLPP | AIYRSTCAVE | VTYFPFDWQN CSLVFRSQTY |
| bγ | VALYCNVLVS | PDGCVYWLPP | AIFRSSCPVS | VTFFPFDWQN CSLIFQSQTY |
| cα2 | VTHMTKAHLF | SNGKVKWVPP | AIYKSSCSID | VTYFPFDQQN CKMKFGSWTY |
| cα3 | VDDKTKALLK | YTGDVTWIPP | AIFKSSCKID | VTYFPFDYQN CTMKFGSWSY |
| cα4 | VTHLTKAHLF | YDGRIKWMPP | AIYKSSCSID | VTFFPFDQQN CKMKFGSWTY |
| cα5 | GT-STKTVVK | YDGTIAWTPP | VNYKSSCTID | VTFFPFDLQN CSMKFGSWTY |
| cα7 | ATFHTNVLVN | SSGHCQYLPP | GIFKSSCYID | VRWFPFDVQK CNLKFGSWTY |
| cα8 | ATFHTNVLVN | YSGSCQYIPP | GILKSTCYID | VRWFPFDVQK CDLKFGSWTH |
| cβ2 | VSFYSNAVIS | YDGSIFWLPP | AIYKSACKIE | VKHFPFDQQN CTMKFRSWTY |
| cβ4 | VSLYTNAIVQ | NNGSIRWLPP | AIYKSACKIE | VKHFPFDQQN CTLKFRSWTY |
| cγ | ITLYTNVLVY | PDGSIYWLPP | AIYRSSCSIH | VTYFPFDWQN CTMVFQSQTY |
| dα2 | VTTMTKAILH | YTGKVVWTPP | AIFKSSCEID | VRYFPFDQQT CFMKFGSWTY |
| dαLi | VTIMTKAILH | HTGKVVWKPP | AIYKSFCEID | VEYFPFDEQT CFMKFGSWTY |
| dβ2 | VTLMTKATLK | YTGEVFWEPP | AIYKSSCEMN | VEYFPYDEQI CFMKFGSWTY |
| dnachr | VRYKSNVLIY | PTGEVLWVPP | AIYQSSCTID | VTYFPFDQQT CIMKFGSWTF |
| gfα3 | VEDKTKALLK | YDGTITWVPP | AIFKSSCPMD | ITYFPFDYQN CSMKFGSWTY |
| gfβ2 | VSFYCNAVVS | NTGDIFWLPP | AIYKSACAIE | VRNFPFDQQN CTLKFRSWTY |
| gfnα2 | GSLMTKAIVR | YNGMITWTPP | ASYKSACTMD | VTFFPFDRQN CSMKFGSWTY |
| gfnα3 | GSLMTKAIVR | FNGTIMWTPP | ASYKSSCTMD | VTFFPFDRQN CSMKFGSWTY |
| hα1 | IVKFTKVLLQ | YTGHITWTPP | AIFKSYCEII | VTHFPFDEQN CSMKLGTWTY |
| hα3 | VDDKTKALLK | YTGEVTWIPP | AIFKSSCKID | VTYFPFDYQN CTMKFGSWSY |
| hα5 | GT-STKTVIR | YNGTVTWTPP | ANYKSSCTID | VTFFPFDLQN CSMKFGSWTY |
| hα7 | ATFHTNVLVN | SSGHCQYLPP | GIFKSSCYID | VRWFPFDVQH CKLKFGSWSY |
| hβ1 | VALDISVVVS | SDGSVRWQPP | GIYRSSCSIQ | VTYFPFDWQN CTMVFSSYSY |
| hβ2 | VSFYSNAVVS | YDGSIFWLPP | AIYKSACKIE | VKHFPFDQQN CTMKFRSWTY |
| mser | SPNIPYVYVH | HRGEVQNYKP | LQLVTACSLD | IYNFPFDVQN CSLTFTSWLH |
| rα2 | VTHMTKAHLF | FTGTVHWVPP | AIYKSSCSID | VTFFPFDQQN CKMKFGSWTY |
| rα3 | VDDKTKALLK | YTGEVTWIPP | AIFKSSCKID | VTYFPFDYQN CTMKFGSWSY |
| rα4 | VTHLTKAHLF | YDGRVQWTPP | AIYKSSCSID | VTFFPFDQQN CTMKFGSWTY |
| rα5 | GA-STKTVVR | YNGTVTWTQP | ANYKSSCTID | VTFFPFDLQN CSMKFGSWTY |
| rα6 | VEGKTKALLK | YDGVITWTPP | AIFKSSCPMD | ITFFPFDHQN CSLKFGSWTY |
| rα7 | ATFHTNVLVN | ASGHCQYLPP | GIFKSSCYID | VRWFPFDVQQ CKLKFGSWSY |
| rβ2 | VSFYSNAVVS | YDGSIFWLPP | AIYKSACKIE | VKHFPFDQQN CTMKFRSWTY |
| rβ3 | GSLMTKAIVK | SSGTVSWTPP | ASYKSSCTMD | VTFFPFDRQN CSMKFGSWTY |
| rβ4 | VSVYTNVIVR | SNGSIQWLPP | AIYKSACKIE | VKHFPFDQQN CTLKFRSWTY |
| rδ | ISYACNVLVS | DSGHVTWLPP | AIFRSSCPIS | VTYFPFDWQN CSLKFSSLKY |

**Appendix 2:** Continued

```
rε     VAYDCNVLVY EGGSVSWLPP AIYRSTCAVE VTYFPFDWQN CSLIFRSQTY
rγ     VALYCNVLVS PDGCIYWLPP AIFRSSCSIS VTYFPFDWQN CSLVFQSQTY
rglyα3 EVTTDNKLLR INGNVLYSIR LTLTLSCPMD LKNFPMDVQT CIMQLESFGY
sαl1   VTTMTKAVLH HTGKVVWTPP AIFKSSCEID VRYFPFDQQT CFMKFGSWTY
tα1    IVHMTKLLLD YTGKIMWTPP AIFKSYCEII VTHFPFDQQN CTMKLGIWTY
tγ     VAYYANVLVY NDGSMYWLPP AIYRSTCPIA VTYFPFDWQN CSLVFRSQTY
xα1a   IVQETKVLLD YTGKIIWLPP AIFKSYCEMI VTYFPFDLQN CSMKLGTWTY
xα1b   ISKDTKILLE YTGKITWTPP AIFKSYCEII VTYFPFDQQN CSMKFGTWTY
xγ     IALYTNTLVS SDGSMYWLPP AIYRSSCPVV VTYFPFDWQN CSIVFQSQTY

bα1    DGSVVVINPE SDPDLSNFME SGEWVIKESR GWKHWVFYAC CPST--PYLD
bβ1    DSSEVSLQTG LSIHEGTFIE NGQWEIIHKP SRLIQPSVDP RGGGEGRREE
bδ     TTKEITLSLK QAIDPEGFTE NGEWEIVHRP ARVNVDPS-V PLDSPNR-QD
bε     NAEEVEFVFA VDIDTEAYTE NGEWAIDFCP G-VIRRHDGD SAGGPGE-TD
bγ     STNEINLQLS QEIDPEAFTE NGEWAIRHRP AKMLLDEAA- PAEEAGH-QK
cα2    DKAKIDLENM EHVDLKDYWE SGEWAIINAI GRYNSKKYDC CTEI--Y-PD
cα3    DKAKIDLVLI GSMNLKDYWE SGEWAIIKAP GYKHDIKYNC CEEI--Y-TD
cα4    DKAKIDLVSM HSVDQLDYWE SGEWVIINAV GNYNSKKYEC CTEI--Y-PD
cα5    DGSQVDIILE DYVDKRDFFD NGEWEIVTAT GSKGNRTDGC CW----Y-PF
cα7    GGWSLDLQMQ -EADISGYIS NGEWDLVGIP GKRTESFYEC CKEP--Y-PD
cα8    SGWLIDLQML -EADISNYIS NGEWDLVGVP GKRNELYYEC CKEP--Y-PD
cβ2    DRTEIDLVLK SEASLDDFTP SGEWDIVALP GRRNENPDD- --ST--Y-VD
cβ4    DHTEIDMVLK TSASMDDFTP SGEWDIVALP GRRTENPLD- --PN--Y-VD
cγ     SANEINLLLT VEIDPEAFTE NGEWAIKHRP ARKIINSGRF TPDDIQY-QQ
dα2    DGDQIDLKHI SGIDLREYYP SVEWDILGVP AERHEKYYPC CAEP--Y-PD
dαLi   DGYMVDLRHL K-IDLQDYYI SVEWDIMRVP AVRNEKFYSC CEEP--Y-LD
dβ2    NGAQVDLKHL D-IDLTEFYL SVEWDILEVP ATKNEEYYPD TLEP--F-SD
dnachr NGDQVSLA-L YNVDLSDYWK SGTWDIIEVP AY-LNVYEGD SNHP--TETD
gfα3   DKAKIDLVLI GSVNLKDFWE SGEWEIIDAP GYKHDIKYNC CEEI--Y-PD
gfβ2   DRTELDLVLT SDASRDDYTP SGEWDIVSLP GRKNEDPND- --LT--Y-LD
gfnα2  DGNMVKLVLI NQVDRSDFFD NGEWEILSAT GVKGSRQDSH LS----Y-PY
gfnα3  DGTMVDLTLL DAVDRKDFFD NGEWEILNAT GQRGSRRDGI YS----Y-PY
hα1    DGSVVAINPE SDPDLSNFME SGEWVIKESR GWKHSVTYSC CPDT--PYLD
hα3    DKAKIDLVLI GSMNLKDYWE SGEWAIIKAP GYKHDIKYNC CEEI--Y-PD
hα5    DGSQVDIILE DQVDKRDFFD NGEWEIVSAT GSKGNRTDSC CW----Y-PY
hα7    GGWSLDLQMQ -EADISGYIP NGEWDLVGIP GKRSERFYEC CKEP--Y-PD
hβ1    DSSEVTLQTG LGIHEGTFIE NGQWENIHKP SRLIQPPGDP RGGREGQRQE
hβ2    DRTEIDLVLK SEASLDDFTP SGEWDIVALP GRRNENPDD- --ST--Y-VD
mser   TIQDINITLW RRSDKSIFIN QGEWELLEV- ---FPQFKEF SIDISNSYAE
rα2    DKAKIDLEQM ERVDLKDYWE SGEWAIINAT GTYNSKKYDC CAEI--Y-PD
rα3    DKAKIDLVLI GSMNLKDYWE SGEWAIIKAP GYKHEIKYNC CEEI--Y-QD
rα4    DKAKIDLVSI HSVDQLDFWE SGEWVIVDAV GTYNTRKYEC CAEI--Y-PD
rα5    DGSQVDIILE DQVDRTDFFD NGEWEIMSAM GSKGNRTDSC CW----Y-PY
rα6    DKAEIDLLII GSVDMNDFWE NSEWEIVDAS GYKHDIKYNC CEEI--Y-TD
rα7    GGWSLDLQMQ -EADISSYIP NGEWDLMGIP GKRNEKFYEC CKEP--Y-PR
rβ2    DRTEIDLVLK SDASLDDFTP SGEWDIIALP GRRNENPDD- --ST--Y-VD
rβ3    DGTMVDLILI NEVDRKDFFD NGEWEILNAK GMKGNRREGF YS----Y-PF
rβ4    DHTEIDMVLK SPAIMDDFTP SGEWDIVALP GRRTVNPQD- --PS--Y-VD
rδ     TAKEIRLSLK QEIDPEGFTE NGEWEIVHRA AKVNVDPS-V PMDSTNH-QD
rε     NAEEVELIFA VDIDTAAFTE NGEWAIDYCP G-MIRHYEGG STEDPGE-TD
rγ     STSEINLQLS QEIDPEAFTE NGEWAIRHRP AKMLLDPVT- PAEEAGH-QK
rglyα3 TMNDLIFEWQ DEAPVQQFLL KEEKDLRYCT KHYNTGKFTC ----------
sαl1   DGDQIDLKHI N-IDLREYYP SVEWDILGVP AERHEKYYPC CAEP--Y-PD
tα1    DGTKVSISPE SDPDLSTFME SGEWVMKDYR GWKHWVYYTC CPDT--PYLD
tγ     NAHEVNLQLS AEIDPEDFTE NGEWTIRHRP AKKNYNW-QL TKDDTDF-QE
xα1a   DGTLVVINPE NDPDLSNFME SGEWYMKDYR CWKHWVYYDC CPET--PYLD
xα1b   DGSLLVINPE RDPDLSNFMA SGEWMMKDYR CWKHWVYYTC CPDK--PYLD
xγ     SANEIELLLT VDIDPEAFTE NGEWAIKHMP AKRIINH-RL PRDDVNY-QQ

bα1    ITYHFVMQRL PLYFIVNVII PCLLFSFLTG LVFYLPTDSG -EKMTLSISV
bβ1    VTFYLIIRRK PLFYLVNVIA PCILITLLAI FVFYLPPDAG -EKMGLSIFA
bδ     VTFYLIIRRK PLFYVINILV PCVLISFMIN LVFYLPADCG -EKTSMAISV
bε     VIYSLIIRRK PLFYVINIIV PCVLISGLVL LAYFLPAQAG GQKCTVSINV
bγ     VVFYLLIQRK PLFYVINIIA PCVLISSVAI LIYFLPAKAG GQKCTVAINV
cα2    ITFYFVIRRL PLFYTINLII PCLLISCLTV LVFYLPSDCG -EKITLCISV
cα3    ITYSLYIRRL PLFYTINMII PCLLISFLTV LVFYLPSDCG -EKVTLCISV
cα4    ITYSFIIRRL PLFYTINLII PCLLISCLTV LVFYLPSECG -EKITLCISV
cα5    VTYSFIIRRL PLFYTLFLII PCIGLSFLTV LVFYLPSNEA -EKISLCTSV
cα7    ITFTVTMRRR TLYYGLNLLI PCVLISALAL LVFLLPADSG -EKISLGITV
```

**Appendix 2:** Continued

| | | | | | |
|---|---|---|---|---|---|
| cα8 | VTYTITMRRR | TLYYGLNLLI | PCVLISGLAL | LVFLLPADSG | -EKISLGITV |
| cβ2 | ITYDFIIRRK | PLFYTINLII | PCILITSLAI | LVFYLPSDCG | -EKMTLCISV |
| cβ4 | VTYDFIIKRK | PLFYTINLII | PCVLITSLAI | LVFYLPSDCG | -EKMTLCISV |
| cγ | VIFYLIIQRK | PLFYIINIIV | PCVLISSMAV | LVYFLPAKAG | GQKCTVSINV |
| dα2 | IFFNITLRRK | TLFYTVNLII | PCVGISYLSV | LVFYLPADSG | -EKIALCISI |
| dαLi | IVFNLTLRRK | TLFYTVNLII | PCVGISFLSV | LVFYLPSDSG | -EKISLCISI |
| dβ2 | ITFKLTMRRK | TLFYTVNLIV | PCVALTFLTV | LVFYLPSDSG | -EKVTLCISI |
| dnachr | ITFYIIIRRK | TLFYTVNLIL | PTVLISFLCV | LVFYLPAEAG | -EKVTLGISI |
| gfα3 | ITYSFYIRRL | PLFYTINLII | PCLLISFLTI | LVFYLPSDCG | -EKVTLCISV |
| gfβ2 | ITYDFVIKRK | PLFYTINLII | PCVLITSLAI | LVFYLPSDCG | -EKVTLCMSV |
| gfnα2 | ITYSFILKRL | PLFYTLFLII | PCLGLSFLTV | LVFYLPSDEG | -EKVSLSTSV |
| gfnα3 | VTYSFILKRL | PLFYTLFLII | PCLGLSFLTV | LVFYLPSDEG | -EKLLLSTSV |
| hα1 | ITYHFVMQRL | PLYFIVNVII | PCLLFSFLTG | LVFYLPTDSG | -EKMTLSISV |
| hα3 | ITYSLYIRRL | PLFYTINLII | PCLLISFLTV | LVFYLPSDCG | -EKVTLCISV |
| hα5 | VTYSFVIKRL | PLFYTLFLII | PCIGLSFLTV | LVFYLPSNEG | -EKICLCTSV |
| hα7 | VTFTVTMRRR | TLYYGLNLLI | PCVLISALAL | LVFLLPADSG | -EKISLGITV |
| hβ1 | VIFYLIIRRK | PLFYLVNVIA | PCILITLLAI | FVFYLPPDAG | -EKMGLSIFA |
| hβ2 | ITYDFIIRRK | PLFYTINLII | PCVLITSLAI | LVFYLPSDCG | -EKMTLCISV |
| mser | MKFYVIIRRR | PLFYAVSLLL | PSIFLMVVDI | VGFCLPPDSG | -ERVSFKITL |
| rα2 | VTYYFVIRRL | PLFYTINLII | PCLLISCLTV | LVFYLPSECG | -EKITLCISV |
| rα3 | ITYSLYIRRL | PLFYTINLII | PCLLISFLTV | LVFYLPSDCG | -EKVTLCISV |
| rα4 | ITYAFIIRRL | PLFYTINLII | PCLLISCLTV | LVFYLPSECG | -EKVTLCISV |
| rα5 | ITYSFVIKRL | PLFYTLFLII | PCIGLSFLTV | VVFYLPSNEG | -EKISLCTSV |
| rα6 | ITYSFYIRRL | PMFYTINLII | PCLFISFLTV | LVFYLPSDCG | -EKVTLCISV |
| rα7 | CTYTVTMRRT | TLYYGLNLLI | PCVLISALAL | LVFLLPADSG | -EKISLGITV |
| rβ2 | ITYDFIIRRK | PLFYTINLII | PCVLITSLAI | LVFYLPSDCG | -EKMTLCISV |
| rβ3 | VTYSFVLRRL | PLFYTLFLII | PCLGLSFLTV | LVFYLPSDEG | -EKLSLSTSV |
| rβ4 | VTYDFIIKRN | ALFYTINLII | PCVLITSLAI | LVFYLPSDCG | -EKMTLCISV |
| rδ | VTFYLIIRRK | PLFYIINILV | PCVLISFMIN | LVFYLPGDCG | -EKTSVAISV |
| rε | VIYTLIIRRK | PLFYVINIIV | PCVLISGLVL | LAYFLPAQAG | GQKCTVSINV |
| rγ | VVFYLLIQRK | PLFYVINIIV | PCVLISSVAI | LIYFLPAKAG | GQKCTVATNV |
| rglyα3 | IEVRFHLERQ | MGYYLIQMYI | PSLLIVILSW | VSFWINMDAA | PARVALGITT |
| sα1l | IFFNITLRRK | TLFYTVNLIV | PCVGISYLSV | LVFYLPADSG | -EKIALCISI |
| tα1 | ITYHFIMQRI | PLYFVVNVII | PCLLFSFLTG | LVFYLPTDSG | -EKMTLSISV |
| tγ | IIFFLIIQRK | PLFYIINIIA | PCVLISSLVV | LVYFLPAQAG | GQKCTLSISV |
| xα1a | ITYHFLLQRL | PLYFIVNVVI | PCLLFSFLTG | LVFYLPTDSG | -EKITLSVSV |
| xα1b | ITYHFVLQRL | PLYFIVNVII | PCLLFSFLTG | LVFYLPTDSG | -EKMTLSISV |
| xγ | IVFYLIIQRK | PLFYIINIIV | PCVLISFVSI | LVYFLPAKAG | GQKCTVSINI |

| | | | | | |
|---|---|---|---|---|---|
| bα1 | LLSLTVFLLV | IVELIPSTSS | AVPLIGKYML | FTMVFVIASI | IITVIVINTH |
| bβ1 | LLTLTVFLLL | LADKVPETSL | SVPIIIKYLM | FTMVLVTFSV | ILSVVVLNLH |
| bδ | LLAQSVFLLL | ISKRLPATSM | AIPLIGKFLL | FGMVLVTMVV | VICVIVLNIH |
| bε | LLAQTVFLFL | IAQKTPETSL | SVPLLGRYLI | FVMVVATLIV | MNCVIVLNVS |
| bγ | LLAQTVFLFL | VAKKVPETSQ | AVPLISKYLT | FLLVVTILIV | VNAVVVLNVS |
| cα2 | LLSLTVFLLL | ITEIIPSTSL | VIPLIGEYLL | FTMIFVTLSI | IITVFVLNVH |
| cα3 | LLSLTVFLLV | ITETIPSTSL | VIPLIGEYLL | FTMIFVTLSI | IITVFVLNVH |
| cα4 | LLSLTVFLLL | ITEIIPSTSL | VIPLIGEYLL | FTMIFVTLSI | IITVFVLNVH |
| cα5 | LVSLTVFLLV | IEEIIPSSSK | VIPLIGEYLV | FTMIFVTLSI | VITVFAINIH |
| cα7 | LLSLTVFMLL | VAEIMPATSD | SVPLIAQYFA | STMIIVGLSV | VVTVIVLQYH |
| cα8 | LLSLTVFMLL | VAEIMPATSD | SVPLIAQYFA | SIMVIVGLSV | VVTVLVLQFH |
| cβ2 | LLALTVFLLL | ISKIVPPTSL | DVPLVGKYLM | FTMVLVTFSI | VTSVCVLNVH |
| cβ4 | LLALTVFLLL | ISKIVPPTSL | DVPLIGKYLM | FTMVLVTFSI | VTSVCVLNVH |
| cγ | LLAQTVFLFL | IAQKVPETSQ | AVPLIGKYLT | FLMVVTVVIV | VNAVIVLNVS |
| dα2 | LLSQTMFFLL | ISEIIPSTSL | ALPLLGKYLL | FTMLLVGLSV | VITIIILNIH |
| dαLi | LLSLTVFFLL | LAEIIPPTSL | TVPLLGKYLL | FTMMLVTLSV | VVTIAVLNVN |
| dβ2 | LVSLTVFFLL | LAEIIPPTSL | AVPLLGKYLL | FTMILVSLSV | WTTVCVLNIH |
| dnachr | LLSLVVFLLL | VSKILPPTSL | VLPLIAKYLL | FTFIMNTVSI | LVTVIIINWN |
| gfα3 | LLSLTVFLLV | ITETIPSTSL | VIPLIGEYLL | FTMIFVTLSI | VITVFVLNVH |
| gfβ2 | LLALTVFLLL | ISKIVPPTSL | AVPLIGKYLM | FTMVLVTFSI | VTSVCVLNVH |
| gfnα2 | LVSLTVFLLV | IEEIIPSSSK | VIPLIGEYLL | FIMIFVTLSI | IVTIFVINVH |
| gfnα3 | LVSLTVFLLV | IEEIIPSSSK | VIPLIGEYLL | FIMIFVTFSI | IVTLFVINVH |
| hα1 | LLSLTVFLLV | IVELIPSTSS | AVPLIGKYML | FTMVFVIASI | IITVIVINTH |
| hα3 | LLSLTVFLLV | ITETIPSTSL | VIPLIGEYLL | FTMIFVTLSI | VITVFVLNVH |
| hα5 | LVSLTVFLLV | IEEIIPSSSK | VIPLIGEYLV | FTMIFVTLSI | MVTVFAINIH |
| hα7 | LLSLTVFMLL | VAEIMPATSD | SVPLIAQYFA | STMIIVGLSV | VVTVIVLQYH |
| hβ1 | LLTLTVFLLL | LADKVPETSL | SVPIIIKYLM | FTMVLVTFSV | ILSVVVLNLH |
| hβ2 | LLALTVFLLL | ISKIVPPTSL | DVPLVGKYLM | FTMVLVTFSI | VTSVCVLNVH |
| mser | LLGYSVFLII | VSDTLPAT-I | GTPLIGVYFV | VCMALLVISL | AETIFIVRLV |
| rα2 | LLSLTVFLLL | ITEIIPSTSL | VIPLIGEYLL | FTMIFVTLSI | VITVFVLNVH |

**Appendix 2:** Continued

```
rα3     LLSLTVFLLV ITETIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH
rα4     LLSLTVFLLL ITEIIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH
rα5     LVSLTVFLLV IEEIIPSSSK VIPLIGEYLV FTMIFVTLSI MVTVFAINIH
rα6     LLSLTVFLLV ITETIPSTSL VIPLVGEYLL FTMIFVTLSI VVTVFVLNIH
rα7     LLSLTVFMLL VAEIMPATSD SVPLIAQYLP STMIIVGLSV VVTVIVLRYH
rβ2     LLALTVFLLL ISKIVPPTSL DVPLVGKYLM FTMVLVTFSI VTSVCVLNVH
rβ3     LVSLTVFLLV IEEIIPSSSK VIPLIGEYLL FIMIFVTLSI IVTVFVINVH
rβ4     LLALTFFLLL ISKIVPPTSL DIPLIGKYLL FTMVLVTFSI VTTVCVLNVH
rδ      LLAQSVFLLL ISKRLPATSM AIPLVGKFLL FGMVLVTMVV VICVIVLNIH
rε      LLAQTVFLFL IAQKIPETSL SVPLLGRYLI FVMVVATLIV MNCVIVLNVS
rγ      LLAQTVFLFL VAKKVPETSQ AVPLISKYLT FLMVVTILIV VNSVVVLNVS
rglyα3 VLTMTTQSSG SRASLPKVSY -VKAIDIWMA VCLLFV-FSA LLEYAAVNFV
sαl1    LLSQTMFFLL ISEIIPSTSL ALPLLGKYLL FTMVLVGLSV VITIMVLNVH
tα1     LLSLTVFLLV IVELIPSTSS AVPLIGKYML FTMIFVISSI IITVVVINTH
tγ      LLAQTIFLFL IAQKVPETSL NVPLIGKYLI FVMFVSMLIV MNCVIVLNVS
xα1a    LLSLVVFLLV IVELIPSTSS AVPLIGKYML FTMVFVIASI VITVIVINTH
xα1b    LLSLTVFLLV IVELIPSTSS AVPLIGKYML FTMVFVIASI IITVIVINTH
xγ      LLAQTVFLFL VAQKIPETST SVPLIVKYLT FLMVVTITIV ANAVIVLNIS

bα1     HRSPST-HVM PEWVRKVFID TIPNIMFFST MKRP--PEVK SAIEGIKYIA
bβ1     HRSPHT-HQM PLWVRQIFIH KLPLYL---G LKRPLPPELR EVVSSISYIA
bδ      FRTPST-HVL SEPVKKLFLE TLPEIL---H MSRPLFSELK PAVDGANFIV
bε      LRTPTT-HAM SPRLRYVLLE LLPQLL---G SGAPAAPEIR CCVDAVNFVA
bγ      LRSPHT-HSM ARGVRKVFLR LLPQLL---R MHVRAAPAIQ ACVEACNLIA
cα2     HRSPST-HTM PHWVRSFFLG FIPRWLF--- MKRPLSPSIL RALEGVQYIA
cα3     YRTPKT-HTM PVWVRTIFLN LLPRIMF--- MTRPLSPEMR DAIESVKYIA
cα4     HRSPRT-HTM PDWVRRVFLD IVPRLLF--- MKRPMSPALK LAVEGVHYIA
cα5     HRSSSTHNAM APWVRKIFLH KLPKLLC--- MRSH----LE AALDSIRYIT
cα7     HHDPDG-GKM PKWTRVILLN WCAWFL---R MKRPGDPDLA KILEEVRYIA
cα8     HHDPQA-GKM PRWVRVILLN WCAWFL---R MKKPTIPVIV KILEEVQFIA
cβ2     HRSPTT-HTM PPWVRTLFLR KLPALLF--- MKQ---PGLE EAVEGVRFIA
cβ4     HRSPST-HTM PPWVKLVFLE RLPAYLF--- MKR---PEVQ EAIDGVSFIA
cγ      LRTPNT-HSM SQRVRQVWLH LLPRYL---G MHMPASPEIR ACVEACNHIA
dα2     YRKPST-HKM RPWIRSFFIK RLPKLL---- LMR---VELE KAIHNVMFIQ
dαLi    FRSPVT-HRM APWVQRLFIQ ILPKLLC--- IER---PEME KTIEGSRFIA
dβ2     FRSPST-HNM SRLVRKLFLH FMPKLMM--- MRR---TEVL QALRAVRFIA
dnachr  FRGPRT-HRM PMYIRSIFLH YLPAFLF--- MKR---PEAS KATEAVEFIA
gfα3    YRTPMT-HTM PSWVRTVFLR ALPRVML--- MRRPVSPEIK QAIESVKYIA
gfβ2    HRSPST-HYM PEWVKCVFLH KLPAFLL--- MRR---PDVD EAIDGVRFIA
gfnα2   HRSSATYHPM SPWVRSLFLQ RLPHLLC--- MRGNLINLLE QATNSVRYIS
gfnα3   HRSSATYHPM APWVKSLFLQ RLPRLLC--- MRGHWIALLE KATHSVHYIS
hα1     HRSPST-HVM PNWVRKVFID TIPNIMFFST MKRP--PEVK SAIEGIKYIA
hα3     YRTPTT-HTM PSWVKTVFLN LLPRVMF--- MTRPLSPEIK EAIQSVKYIA
hα5     HRSSSTHNAM APLVRKIFLH TLPKLLS--- MRSH----LE AALDSIRYIT
hα7     HHDPDG-GKM PKWTRVILLN WCAWFL---R MKRPGDPDLA KILEEVRYIA
hβ1     HRSPHT-HQM PLWVRQIFIH KLPLYL---R LKRPLLPELR EVVSSISYIA
hβ2     HRSPTT-HTM APWVKVVFLE KLPALLF--- MQQ---PGLR EAVDGVRFIA
mser    HKQ-DLQRPV PDWLRHLVLD RIAWILC--- LPREASLAVR GLLQELSSIR
rα2     HRSPST-HNM PNWVRVALLG RVPRWLM--- MNRPLSPQIQ KALEGVHYIA
rα3     YRTPTT-HTM PTWVKAVFLN LLPRVMF--- MTRPLSPEIK EAIQSVKYIA
rα4     HRSPRT-HTM PAWVRRVFLD IVPRLLF--- MKRP-SPALT RAVEGVQYIA
rα5     HRSSSTHNAM APWVRKIFLH KLPKLLC--- MRSH----LE AALDCIRYIT
rα6     YRTPAT-HTM PKWVKTMFLQ VFPSILM--- MRRPHPPDVE DVIDSVQFIA
rα7     HHDPDG-GKM PKWTRIILLN WCAWFL---R MKRPGDPDLA KILEEVRYMP
rβ2     HRSPTT-HTM APWVKVVFLE KLPTLLF--- LQQ---PGLR EAVDGVRFIA
rβ3     HRSSSTYHPM APWVKRLFLQ RLPRWLC--- MKDPLVAFLE KASESIRYIS
rβ4     HRSPST-HTM ASWVKECFLH KLPTFLF--- MKR---PDLQ EALEGVSFIA
rδ      FRTPST-HVL SEGVKKFFLE TLPKLL---H MSRPLFNEMK PAVDGANFIV
rε      LRTPTT-HAT SPRLRQILLE LLPRLL---G LSPPAAPEVR CCVDAVNFVA
rγ      LRSPHT-HSM ARGVRKVFLR LLPQLL---R MHVHASPAIQ ACVDACNLMA
rglyα3 SRQHKELRKR KNKTEAFALE KFYRFLSFTA YG--MGPCLQ AK-DGVVPKG
sαl1    YRKPST-HKM APWVRKVFIR RLPKLL---- LMR---VELE KAIHNVLFIQ
tα1     HRSPST-HTM PQWVRKIFID TIPNVMFFST MKRA--PDVK SAIEGVKYIA
tγ      LRTPNT-HSL SEKIKHLFLG FLPKYL---G MQLEFAPEIK SCVEACNFIA
xα1a    HRSPST-HIM PQWLKKIFIE TIPRVMFFST MKRP--PDVK SAIEGAKYVA
xα1b    HRSPST-HTM PPWVRKIFIE TIPNIMFFST MKRP--PDVK SAIEGIKYIA
xγ      LRTPNT-HSM SSTVRELCLR TVPRLL---R MHLRAAPEIR TCVEACCHIA
```

**Appendix 2:** Continued

| | |
|---|---|
| bα1 | LLSLTVFLLV IVELIPSTSS AVPLIGKYML FTMVFVIASI IITVIVINTH |
| bβ1 | LLTLTVFLLL LADKVPETSL SVPIIIKYLM FTMVLVTFSV ILSVVVLNLH |
| bδ | LLAQSVFLLL ISKRLPATSM AIPLIGKFLL FGMVLVTMVV VICVIVLNIH |
| bε | LLAQTVFLFL IAQKTPETSL SVPLLGRYLI FVMVVATLIV MNCVIVLNVS |
| bγ | LLAQTVFLFL VAKKVPETSQ AVPLISKYLT FLLVVTILIV VNAVVVLNVS |
| cα2 | LLSLTVFLLL ITEIIPSTSL VIPLIGEYLL FTMIFVTLSI IITVFVLNVH |
| cα3 | LLSLTVFLLV ITETIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH |
| cα4 | LLSLTVFLLL ITEIIPSTSL VIPLIGEYLL FTMIFVTLSI IITVFVLNVH |
| cα5 | LVSLTVFLLV IEEIIPSSSK VIPLIGEYLV FTMIFVTLSI VITVFAINIH |
| cα7 | LLSLTVFMLL VAEIMPATSD SVPLIAQYFA STMIIVGLSV VVTVIVLQYH |
| cα8 | LLSLTVFMLL VAEIMPATSD SVPLIAQYFA SIMVIVGLSV VVTVLVLQFH |
| cβ2 | LLALTVFLLL ISKIVPPTSL DVPLVGKYLM FTMVLVTFSI VTSVCVLNVH |
| cβ4 | LLALTVFLLL ISKIVPPTSL DVPLIGKYLM FTMVLVTFSI VTSVCVLNVH |
| cγ | LLAQTVFLFL IAQKVPETSQ AVPLIGKYLT FLMVVTVVIV VNAVIVLNVS |
| dα2 | LLSQTMFFLL ISEIIPSTSL ALPLLGKYLL FTMLLVGLSV VITIIILNIH |
| dαLi | LLSLTVFFLL LAEIIPPTSL TVPLLGKYLL FTMMLVTLSV VVTIAVLNVN |
| dβ2 | LVSLTVFFLL LAEIIPPTSL AVPLLGKYLL FTMILVSLSV WTTVCVLNIH |
| dnachr | LLSLVVFLLL VSKILPPTSL VLPLIAKYLL FTFIMNTVSI LVTVIIINWN |
| gfα3 | LLSLTVFLLV ITETIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH |
| gfβ2 | LLALTVFLLL ISKIVPPTSL AVPLIGKYLM FTMVLVTFSI VTSVCVLNVH |
| gfnα2 | LVSLTVFLLV IEEIIPSSSK VIPLIGEYLL FIMIFVTLSI IVTIFVINVH |
| gfnα3 | LVSLTVFLLV IEEIIPSSSK VIPLIGEYLL FIMIFVTFSI IVTLFVINVH |
| hα1 | LLSLTVFLLV IVELIPSTSS AVPLIGKYML FTMVFVIASI IITVIVINTH |
| hα3 | LLSLTVFLLV ITETIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH |
| hα5 | LVSLTVFLLV IEEIIPSSSK VIPLIGEYLV FTMIFVTLSI MVTVFAINIH |
| hα7 | LLSLTVFMLL VAEIMPATSD SVPLIAQYFA STMIIVGLSV VVTVIVLQYH |
| hβ1 | LLTLTVFLLL LADKVPETSL SVPIIIKYLM FTMVLVTFSV ILSVVVLNLH |
| hβ2 | LLALTVFLLL ISKIVPPTSL DVPLVGKYLM FTMVLVTFSI VTSVCVLNVH |
| mser | LLGYSVFLII VSDTLPAT-I GTPLIGVYFV VCMALLVISL AETIFIVRLV |
| rα2 | LLSLTVFLLL ITEIIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH |
| rα3 | LLSLTVFLLV ITETIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH |
| rα4 | LLSLTVFLLL ITEIIPSTSL VIPLIGEYLL FTMIFVTLSI VITVFVLNVH |
| rα5 | LVSLTVFLLV IEEIIPSSSK VIPLIGEYLV FTMIFVTLSI MVTVFAINIH |
| rα6 | LLSLTVFLLV ITETIPSTSL VIPLVGEYLL FTMIFVTLSI VVTVFVLNIH |
| rα7 | LLSLTVFMLL VAEIMPATSD SVPLIAQYLP STMIIVGLSV VVTVIVLRYH |
| rβ2 | LLALTVFLLL ISKIVPPTSL DVPLVGKYLM FTMVLVTFSI VTSVCVLNVH |
| rβ3 | LVSLTVFLLV IEEIIPSSSK VIPLIGEYLL FIMIFVTLSI IVTVFVINVH |
| rβ4 | LLALTFFLLL ISKIVPPTSL DIPLIGKYLL FTMVLVTFSI VTTVCVLNVH |
| rδ | LLAQSVFLLL ISKRLPATSM AIPLVGKFLL FGMVLVTMVV VICVIVLNIH |
| rε | LLAQTVFLFL IAQKIPETSL SVPLLGRYLI FVMVVATLIV MNCVIVLNVS |
| rγ | LLAQTVFLFL VAKKVPETSQ AVPLISKYLT FLMVVTILIV VNSVVVLNVS |
| rglyα3 | VLTMTTQSSG SRASLPKVSY -VKAIDIWMA VCLLFV-FSA LLEYAAVNFV |
| sα11 | LLSQTMFFLL ISEIIPSTSL ALPLLGKYLL FTMVLVGLSV VITIMVLNVH |
| tα1 | LLSLTVFLLV IVELIPSTSS AVPLIGKYML FTMIFVISSI IITVVVINTH |
| tγ | LLAQTIFLFL IAQKVPETSL NVPLIGKYLI FVMFVSMLIV MNCVIVLNVS |
| xα1a | LLSLVVFLLV IVELIPSTSS AVPLIGKYML FTMVFVIASI VITVIVINTH |
| xα1b | LLSLTVFLLV IVELIPSTSS AVPLIGKYML FTMVFVIASI IITVIVINTH |
| xγ | LLAQTVFLFL VAQKIPETST SVPLIVKYLT FLMVVTITIV ANAVIVLNIS |

**Appendix 2:** Continued

```
cα8     IIDVDEKNQV LITNAWLQMY WVDIYLSWDQ YEYPGVQNLR FPSDQIWVPD
dα2     LIDLNLKDQI LTTNVWLEHE WQDHKFKWDP SEYGGVTELY VPSEHIWLPD
dαLi    LIDVNLKNQI MTTNVWVEQE WNDYKLKWNP DDYGGVDTLH VPSEHIWHPD
dβ2     LIEVNLKNQV MTTNLWVKQR WFDYKLRWDP EEYGGVEQLY VPSEHIWVPD
dnachr  LINVNEKNQV MKSNVWLRLV WYDYQLQWDE ADYGGIGVLR LPPDKVWKPD
mser    ILNVDEKNQV LTTYIWYRQY WTDEFLQWTP EDFDNVTKLS IPTDSIWVPD
onachr  IIDVHEIDQI MTCSVWLKQV WIDKKLSWNP EIYGGVSVLY VPYEMVWVPD
rα2     LIDVDEKNQM MTTNVWLKQE WNDYNVRWDP AEFGNVTSLR VPSEMIWIPD
rα3     LVKVDEVNQI METNLWLKQI WNDYKLKWKP SDYQGVEFMR VPAEKIWKPD
rα7     IMDVDEKNQV LTTNIWLQMS WTDHYLQWNM SEYPGVKNVR FPDGQIWKPD
rβ2     LISVHEREQI MTTNVWLTQE WEDYRLTWKP QHFDNMKKVR LPSKHIWLPD
rδ      LISLKEVEET LTTNVWIDHA WIDSRLQWNA NEFGNITVLR LPSDMVWLPE
rγ      LISLNEREEA LTTNVWIEMQ WCDYRLRWDP KDYEGLWILR VPSTMVWQPD

cα8     ILLYNSADER FDATFHTNVL VNYSGSCQYI PPGILKSTCY IDVRWFPFDV
dα2     IVLYNNADGE YVVTTMTKAI LHYTGKVVWT PPAIFKSSCE IDVRYFPFDQ
dαLi    IVLYNNADGN YEVTIMTKAI LHHTGKVVWK PPAIYKSFCE IDVEYFPFDE
dβ2     IVLYNNWDGN YEVTLMTKAT LKYTGEVFWE PPAIYKSSCE MNVEYFPYDE
dnachr  IVLFNNADGN YEVRYKSNVL IYPTGEVLWV PPAIYQSSCT IDVTYFPFDQ
mser    ILINEFVDVG -KSPNIPYVY VHHRGEVQNY KPLQLVTACS LDIYNFPFDV
onachr  IVLYNTVDSN YNITISTKAT LRYDGQVTWD SPAIFKTLCQ IDVRWFPFDE
rα2     IVLYNNADGE FAVTHMTKAH LFFTGTVHWV PPAIYKSSCS IDVTFFPFDQ
rα3     IVLYNNADGD FQVDDKTKAL LKYTGEVTWI PPAIFKSSCK IDVTYFPFDY
rα7     ILLYNSADER FDATFHTNVL VNASGHCQYL PPGIFKSSCY IDVRWFPFDV
rβ2     VVLYNNADGM YEVSFYSNAV VSYDGSIFWL PPAIYKSACK IEVKHFPFDQ
rδ      IVLENNNDGS FQISYACNVL VSDSGHVTWL PPAIFRSSCP ISVTYFPFDW
rγ      IVLGNNVDGV FEVALYCNVL VSPDGCIYWL PPAIFRSSCS ISVTYFPFDW

cα8     QKCDLKFGSW THSGWLIDLQ MLEA-DISNY ISNGEWDLVG VPGKRNELYY
dα2     QTCFMKFGSW TYDGDQIDLK HISGIDLREY YPSVEWDILG VPAERHEKYY
dαLi    QTCFMKFGSW TYDGYMVDLR HLK-IDLQDY YISVEWDIMR VPAVRNEKFY
dβ2     QICFMKFGSW TYNGAQVDLK HL-DIDLTEF YLSVEWDILE VPATKNEEYY
dnachr  QTCIMKFGSW TFNGDQVSLA LYN-VDLSDY WKSGTWDIIE VPAYLNV--Y
mser    QNCSLTFTSW LHTIQDINIT LWRRSDKSIF INQGEWELLE VFPQFKEFSI
onachr  QNCHFKFGSW TYTEDLLNLE LLDSIDLSDY YPSVEWDIMS RIAKRRTKNY
rα2     QNCKMKFGSW TYDKAKIDLE QMERVDLKDY WESGEWAIIN ATGTYNSKKY
rα3     QNCTMKFGSW SYDKAKIDLV LIGSMNLKDY WESGEWAIIK APGYKHEIKY
rα7     QQCKLKFGSW SYGGWSLDLQ MQEA-DISSY IPNGEWDLMG IPGKRNEKFY
rβ2     QNCTMKFRSW TYDRTEIDLV LKSDASLDDF TPSGEWDIIA LPGRRNENPD
rδ      QNCSLKFSSL KYTAKEIRLS LKQEIDPEGF TENGEWEIVH RAAKVNVDP-
rγ      QNCSLVFQSQ TYSTSEINLQ LSQEIDPEAF TENGEWAIRH RPAKMLLDP-

cα8     E--CCKEPYP DVTYTITMRR RTLYYGLNLL IPCVLISGLA LLVFLLPADS
dα2     --PCCAEPYP DIFFNITLRR KTLFYTVNLI IPCVGISYLS VLVFYLPADS
dαLi    --SCCEEPYL DIVFNLTLRR KTLFYTVNLI IPCVGISFLS VLVFYLPSDS
dβ2     --PDTLEPFS DITFKLTMRR KTLFYTVNLI VPCVALTFLT VLVFYLPSDS
dnachr  EGDSNHPTET DITFYIIIRR KTLFYTVNLI LPTVLISFLC VLVFYLPAEA
mser    DIS---NSYA EMKFYVIIRR RPLFYAVSLL LPSIFLMVVD IVGFCLPPDS
onachr  LTSFSDEAFI DIIFYLELRR KPLFYTVNLV FPCVGISFLT IVAFYLPPRS
rα2     --DCCAEIYP DVTYYFVIRR LPLFYTINLI IPCLLISCLT VLVFYLPSEC
rα3     --NCCEEIYQ DITYSLYIRR LPLFYTINLI IPCLLISFLT VLVFYLPSDC
rα7     E--CCKEPYP RCTYTVTMRR TTLYYGLNLL IPCVLISALA LLVFLLPADS
rβ2     --DS---TYV DITYDFIIRR KPLFYTINLI IPCVLITSLA ILVFYLPSDC
rδ      SVPMDSTNHQ DVTFYLIIRR KPLFYIINIL VPCVLISFMI NLVFYLPGDC
rγ      VTPAEEAGHQ KVVFYLLIQR KPLFYVINII VPCVLISSVA ILIYFLPAKA

cα8     G-EKISLGIT VLLSLTVFML LVAEIMPATS DSVPLIAQYF ASIMVIVGLS
dα2     G-EKIALCIS ILLSQTMFFL LISEIIPSTS LALPLLGKYL LFTMLLVGLS
dαLi    G-EKISLCIS ILLSLTVFFL LISEIIPSTS LALPLLGKYL LFTMLLVGLS
dβ2     G-EKVTLCIS ILVSLTVFFL LLAEIIPPTS LTVPLLGKYL LFTMMLVTLS
dnachr  G-EKVTLGIS ILLSLVVFLL LVSKILPPTS LVLPLIAKYL LFTFIMNTVS
mser    G-ERVSFKIT LLLGYSVFLI IVSDTLPAT- IGTPLIGVYF VVCMALLVIS
onachr  G-EKVTLCIL ILVALTVFYL LLKDIIPATS IALPLFGKYL LFTMIMVSLS
rα2     G-EKITLCIS VLLSLTVFLL LITEIIPSTS LVIPLIGEYL LFTMIFVTLS
rα3     G-EKVTLCIS VLLSLTVFLL VITETIPSTS LVIPLIGEYL LFTMIFVTLS
rα7     G-EKISLGIT VLLSLTVFML LVAEIMPATS DSVPLIAQYL PSTMIIVGLS
rβ2     G-EKMTLCIS VLLALTVFLL LISKIVPPTS LDVPLVGKYL MFTMVLVTFS
rδ      G-EKTSVAIS VLLAQSVFLL LISKRLPATS MAIPLVGKFL LFGMVLVTMV
rγ      GGQKCTVATN VLLAQTVFLF LVAKKVPETS QAVPLISKYL TFLMVVTILI
```

**Appendix 3:** Alignment of 13 nAChR Subunits Made by the CLUSTAL V Software—Total Sites: 351; Informative Sites: 263

| | | | | | |
|---|---|---|---|---|---|
| cα8 | VVVTVLVLQF | HHHDPQAGKM | PRWVRVILLN | WCAWFLRMKK | PTIPVIVKIL |
| dα2 | VVITIIILNI | HYRKPSTHKM | RPWIRSFFIK | RLPKLL-LMR | V---ELEKAI |
| dαLi | VVVTIAVLNV | NFRSPVTHRM | APWVQRLFIQ | ILPKLLCIER | P---EMEKTI |
| dβ2 | VWTTVCVLNI | HFRSPSTHNM | SRLVRKLFLH | FMPKLMMMRR | T---EVLQAL |
| dnachr | ILVTVIIINW | NFRGPRTHRM | PMYIRSIFLH | YLPAFLFMKR | ---PEASKAT |
| mser | LAETIFIVRL | VHKQDLQRPV | PDWLRHLVLD | RIAWILCLPR | EASLAVRGLL |
| onachr | VLVTVISLNL | HFRRPSTHRM | PIWVKWLFLR | ILPKILFMRR | ----HVIKAF |
| rα2 | IVITVFVLNV | HHRSPSTHNM | PNWVRVALLG | RVPRWLMMNR | PLSPQIQKAL |
| rα3 | IVITVFVLNV | HYRTPTTHTM | PTWVKAVFLN | LLPRVMFMTR | PLSPEIKEAI |
| rα7 | VVVTVIVLRY | HHHDPDGGKM | PKWTRIILLN | WCAWFLRMKR | PGDPDLAKIL |
| rβ2 | IVTSVCVLNV | HHRSPTTHTM | APWVKVVFLE | KLPTLLFLQQ | ---PGLREAV |
| rδ | VVICVIVLNI | HFRTPSTHVL | SEGVKKFFLE | TLPKLLHMSR | PLFNEMKPAV |
| rγ | VVNSVVVLNV | SLRSPHTHSM | ARGVRKVFLR | LLPQLLRMHV | HASPAIQACV |

| | | | | | | |
|---|---|---|---|---|---|---|
| cα8 | EEVQFIAMRF | RKQDEGEEIC | SEWKFAAAVI | DRLCLVAFTL | FAIICTFTIL | M |
| dα2 | HNVMFIQHHM | QRQDEFNAED | QDWGFVAMVM | DRLFLWLFMI | ASLVGT-FVI | L |
| dαLi | EGSRFIAQHV | KNKDKFESVE | EDWKYVAMVL | DRMFLWIFAI | ACVVGTALII | L |
| dβ2 | RAVRFIAQHI | KDADKDNEIV | EDWKFVSMVL | DRFFLWLFTL | SCVFGTLAII | C |
| dnachr | EAVEFIAEHL | RNEDLYIQTR | EDWKYVAMVI | DRLQLYIFFI | VTTAGTVGIL | M |
| mser | QELSSIRHFL | EKRDEMREVA | RDWLRVGYVL | DRLLFRIYLL | AVLAYSITLV | T |
| onachr | ENVCFIAQLL | KKKDREAMID | EDWKFVARVL | DRLFLLLFSI | ACFLGTILIL | F |
| rα2 | EGVHYIADRL | RSEDADSSVK | EDWKYVAMVV | DRIFLWLFII | VCFLGTIGLF | L |
| rα3 | QSVKYIAENM | KAQNVAKEIQ | DDWKYVAMVI | DRIFLWVFIL | VCILGTAGLF | L |
| rα7 | EEVRYMPTAY | RCQDESEVIC | SEWKFAACVV | DRLCLMAFSV | FTIICTIGIL | M |
| rβ2 | DGVRFIADHM | RSEDDDQSVR | EDWKYVAMVI | DRLFLWIFVF | VCVFGTVGMF | L |
| rδ | DGANFIVNHM | RDQNSYNEEK | DNWNQVARTV | DRLCLFVVTP | VMVVGTAWIF | L |
| rγ | DACNLMARAR | HQQSHFDSGN | EEWLLVGRVL | DRVCFLAMLS | LFICGTAGIF | L |

**Appendix 3:** Continued