Molecular evolution of the nicotinic acetylcholine receptor: an example of multigene family in excitable cells

Nicolas Le Novère and Jean-Pierre Changeux
Journal of Molecular Evolution (1995) 40 : 155-172

December 5, 1998

Abstract

An extensive phylogenetic analysis of the nicotinic acetylcholine receptor subunit genes family has been performed by cladistic and phenetic methods. The conserved parts of amino acid sequences have been analyzed by CLUSTALV 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 ($\alpha 7$,$\alpha 8$), III-the neuronal nicotinic subunits ($\alpha 2$–$\alpha 6$, $\beta 2$–$\beta 4$) which also contains the muscle acetylcholine binding subunit ($\alpha 1$), and IV-the muscle non-α subunits ($\beta 1$, 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

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 super-family 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 super-family 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; rev. Bertrand et al. 1993).

The nAChR of striated muscle is the best characterized member of the ligand gated ion channel super-family (Changeux 1990; Karlin 1993): it is an hetero-pentamer (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 domains MIII and MIV. The amino-terminal hydrophilic domain, carries the ACh binding site (Devillers-Thiéry 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 definition of boldface terms) to those of muscle nAChR (Sargent 1993). In the jawed vertebrate nervous systems, several subunits (named $\alpha2$–$\alpha8$) have been identified which share with the muscle-type receptor $\alpha1$ 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 cystein pair, have been characterized and named non-α or $\beta2$–$\beta4$. As for muscle nAChR, the functional neuronal nAChR is an hetero-pentamer 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 $\alpha5$ is coprecipitated with another α and β subunit in some neuronal nAChRs indicates that more than 2 different subunits may assemble together to form a receptor molecule (Conroy et al 1992, Vernallis et al. 1993). In contrast, in reconstituted systems, the $\alpha7$ or $\alpha8$ 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 subunit $\alpha2$–$\alpha6$ and $\beta2$–$\beta4$ but not receptors formed by subunit $\alpha7$ and $\alpha8$ (Clarcke et al. 1985), which are labelled 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 (Gershenfeld 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 (Gershenfeld 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 multigene family differentially expressed in the nervous system. Its evolution deserves comparison with the increase in complexity of 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.

Material and Methods

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

Alignments of the Sequences

Alignments were performed using the CLUSTALV 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 similarity have been determined by the Dayhoff PAM 250 matrix. The protein sequences were aligned after the following modifications:

Deletion of the signal peptide (corresponding to tα1 aa 1-27), the small non conserved 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 twelve sequences. the alignment of the 13 sequences show 351 sites with 263 informative sites (Appendix 2).

Sequences 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 the Neighbor-joining (NJ) (Saitou and Nei 1987, program NEIGHBOR). The distance matrix was provided by the Dayhoff PAM matrix (Dayhoff

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. The six sign codes are the accession numbers of the Genbank-Embl databases. The rat a4 is the isoform 4-2.

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)
dARD	*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)
onachr	*Onchocerca volvulus*	L20465	Ajuh and Egwang. Brain Res (1994)
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	Séguéla 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*	J00963	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)
xβ1	*Xenopus laevis*	U04618	Kullberg et al. Rec Chan (1994) in press
xδ	*Xenopus laevis*	X07069	Baldwin et al. J Cell Biol. 106:469 (1988)
xϵ	*Xenopus laevis*	U19612	Murray et al. Neuron (1995)
xγ	*Xenopus laevis*	X07068	Baldwin et al. J Cell Biol. 106:469 (1988)

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 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 beetween subunits (Fig. 5) is based on the NJ analysis of the Appendix 2 protein alignment. The external branches of the resulting tree showed an approximative 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 calculate 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 $3.2 10^{-10}$ and $7.3 10^{-10}$ substitution per site and per year (M=5.3 σ=1.2), which is not far from a molecular clock.

Results

Evolution of the Gene Structure

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

The genes of the subfamily IV possess 11 or 12 exons of which ten are conserved. In the subfamily III, the main part of the coding sequence is distributed within a single exon. The structure of $\alpha 1$ and $\alpha 7$ 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

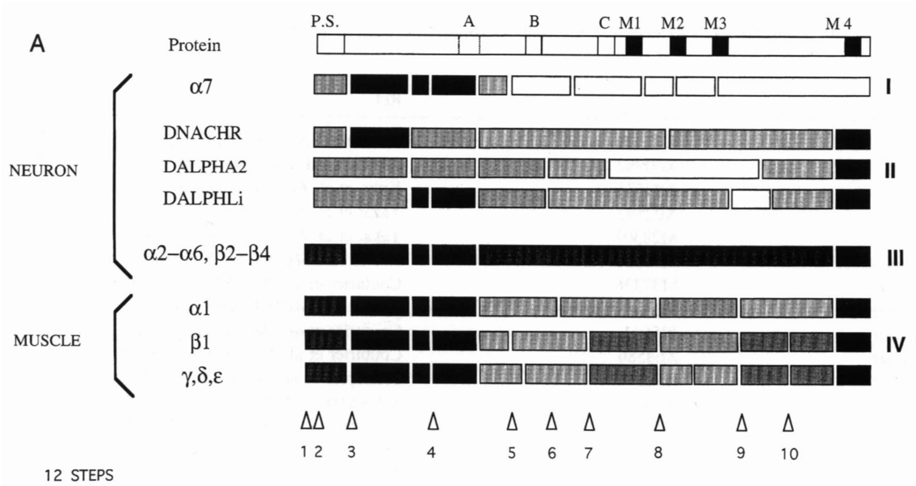

Figure 1: **A**:Structure of the subunit genes. Only the exons at least partialy coding are represented. The gray level grossly reflects the conservation of the exon through the family (i.e. the presence of an exonic frontier at this place in different subunits, but not the sequence similitude between 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.

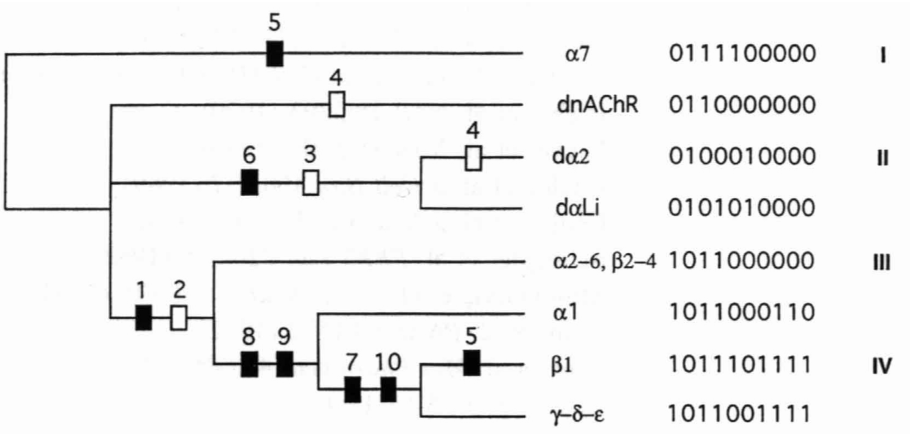

Figure 2: **B**: Cladogram constructed with the exonic structure of the genes from a MP analysis which gave three equivalent trees. The informative limits of each gene is coded at the right of its name (0: absence; 1: presence). open box: loss of a limit; Filled box: gain of a limit. The dendrogram is arbitrarily rooted. The branch lengths make no sense.

or absence of the frontier) gave three 12 step cladograms (summarized in Fig. 1.B 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 can be that the genic structure of type I subunits is mainly made of **autapomorphies**, whereas that of $\alpha1$ is **plesiomorphic** in the 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) place $\alpha1$ as a sister group of subfamily III. A translocation (e.g. between exon 4 and 5) could mask the real onset of the $\alpha1$ 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 $\beta2$ and $\beta4$ subunits were branched with the 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 $\beta1$ 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: $(\beta2, \beta4), (\beta3, \alpha5), (\alpha2, \alpha4), (\alpha3, \alpha6)$. A first duplication gave $\beta2$ and $\beta4$. The position of $\beta2$ and $\beta4$ was unstable, jumping between subfamily III and subfamily IV according to the sequences sampled. However, if we consider the gene structure, the neuronal localisation and the pharmacological characteristics of these subunits, $\beta2$ and $\beta4$ have to be placed in subfamily III. Then, we observe the separation of $\beta3$ and $\alpha5$. At last, a **monophyletic** group formed by the two pairs $(\alpha4, \alpha2)$ and $(\alpha3, \alpha6)$ is present in MP and NJ analyses. The position of $\alpha1$ does not match the gene structure analysis. This strange position of $\alpha1$ inside the neuronal subgroup is, however, weakly supported by bootstrap scores. The last clear duplications arose in the lineage of teleosteans which possess two **homologs** of $\beta3$ (appeared about 280 MYA), and in Tetrapoda which have two **homologs** of the goldfish $\beta2, \beta2$ and $\beta4$ subunits. Moreover, goldfish $\alpha3$ is not clearly **homolog** to tetrapod $\alpha3$ (NJ Fig. 3) or $\alpha6$ (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. Then, the ability for a subunit to form functional homo-oligomers

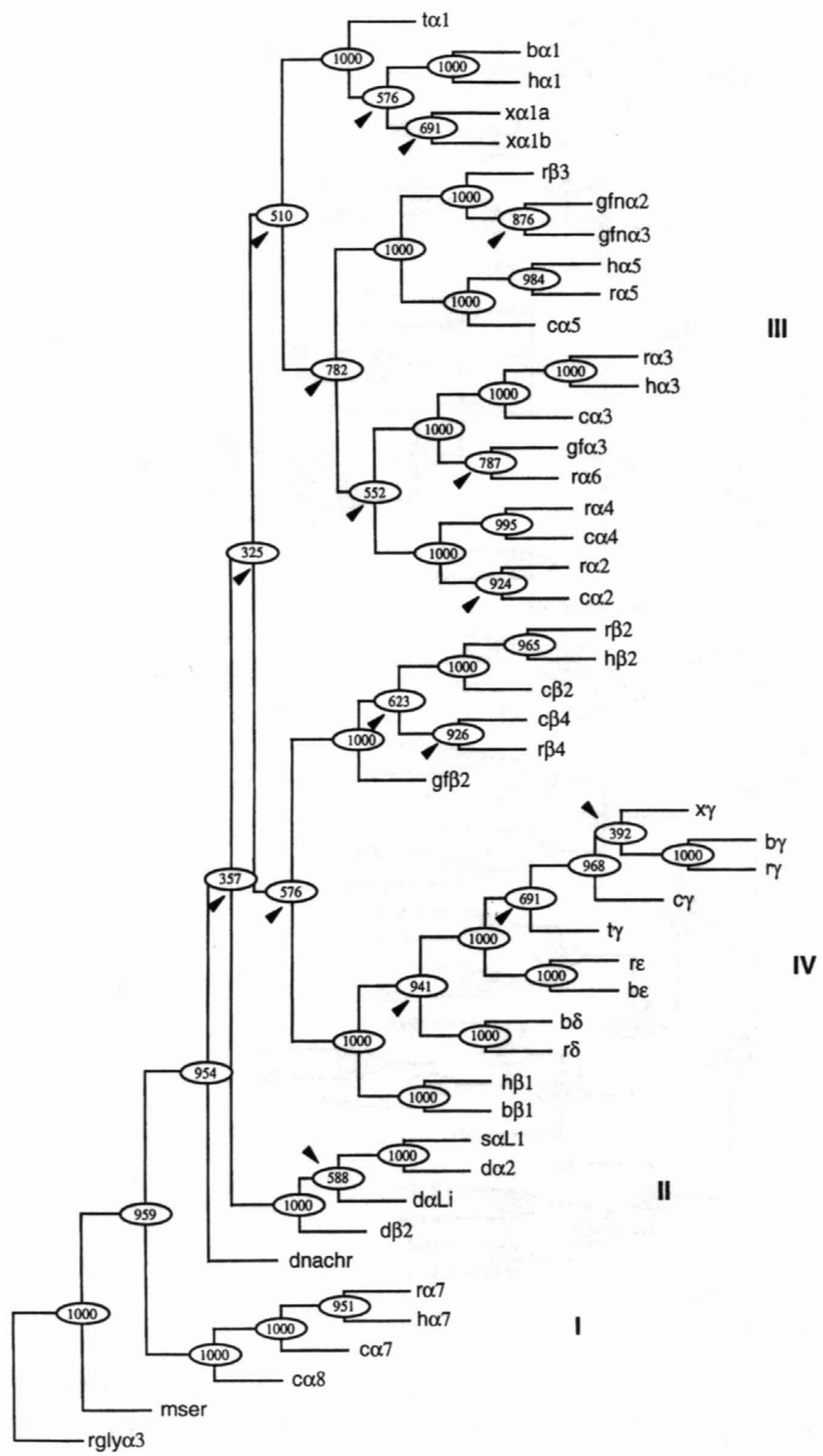

Figure 3: Bootstrap Majority-Rule consensus tree obtained from 150 MP replicates (SEQBOOT, PROTPARS and CONSENSE programs) with the alignment shown in appendix 1. The nodes indicated by an arrowhead are uncertain.

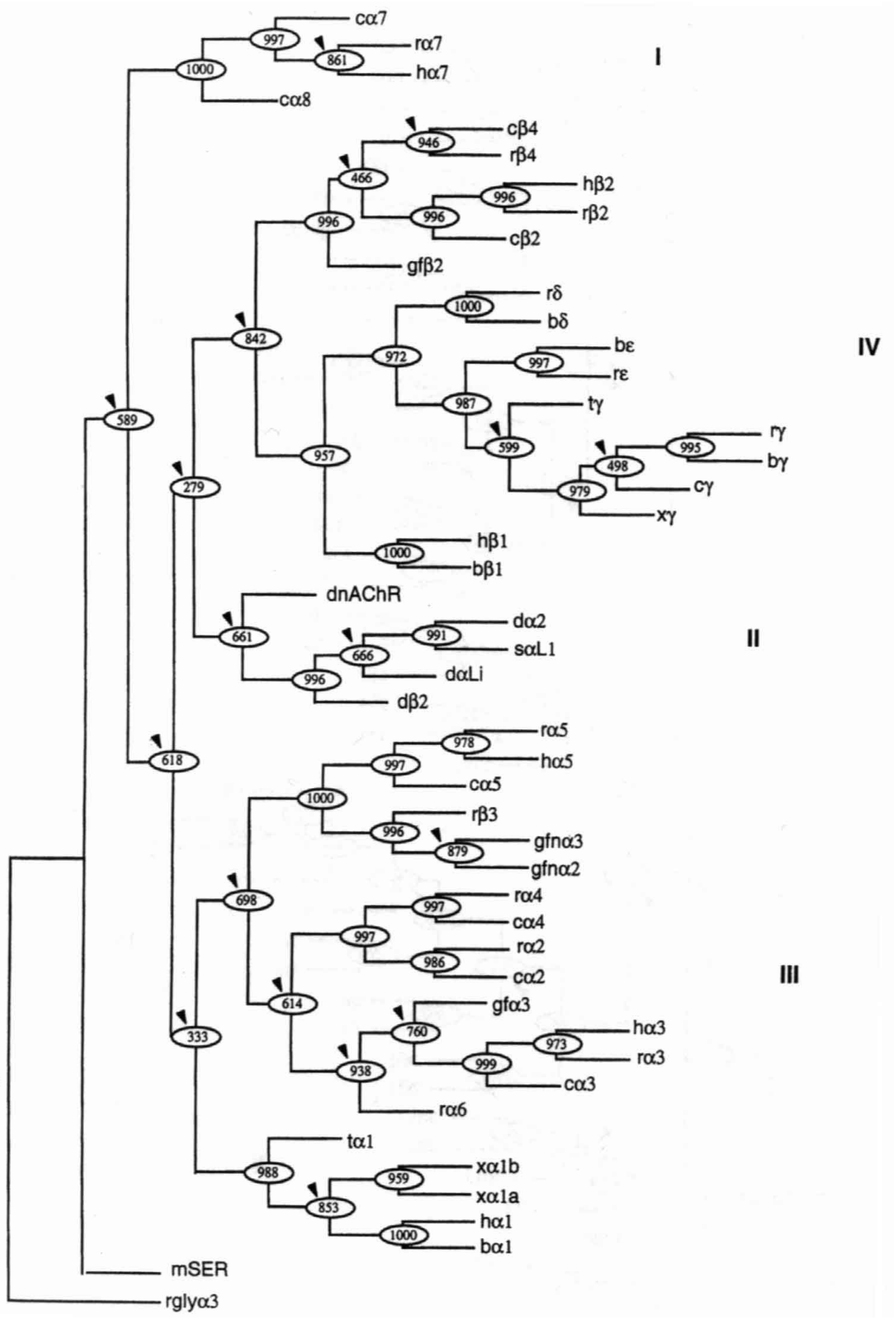

Figure 4: Bootstrap Majority-Rule consensus tree obtained from 300 NJ replicates (SEQBOOT, PROTDIST, NEIGHBOR and CONSENSE programs) with the alignment shown in appendix 1. The nodes indicated by an arrowhead are uncertain.

could reflect a **plesiomorphic** ("primitive") mode of functioning. α-Bgt sensitive homooligomers 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 functional homo-oligomeric channels (Marshall et al. 1990) blocked by a-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 functionnal 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 class2, Marshall et al. 1990), dnAChR (insect class1, 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. Though α1 belongs to the subfamily III, the functional α-Bgt sites of the vertebrate muscle receptor is formed partially by the subunits of subfamily IV. Moreover, if α1 is placed as a sister group of all other subunit 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 1000 MYA here). Receptors of the 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 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 non-vertebrates/non-arthropods 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 occuring in other phyla.

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

The bootstrap confirms that $\alpha7$ and $\alpha8$ diverged prior to the separation of Sauropsida and Theropsida. Thus, an $\alpha8$ subunit may be present in mammals.

Reconstructed History of the nAChR Subunit Gene Family

Based on present and previous results, the history of 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 have already multipolar neurons, of ectodermal origin. This first nAChR presumably had a neuronal localisation. 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), placed this subunit as 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 αL1 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 the extant subfamily IV, which was complete in vertebrate phylum before the appearence of chondrichthyes (450 MYA) and the 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 $\alpha1$, (a duplication which is responsible of the heteromeric muscle receptor), a further duplication from $\alpha1$ provided a new gene, which expression became neuronal. The evolution of the promoters (and of the transcription regulators) may thus have played a

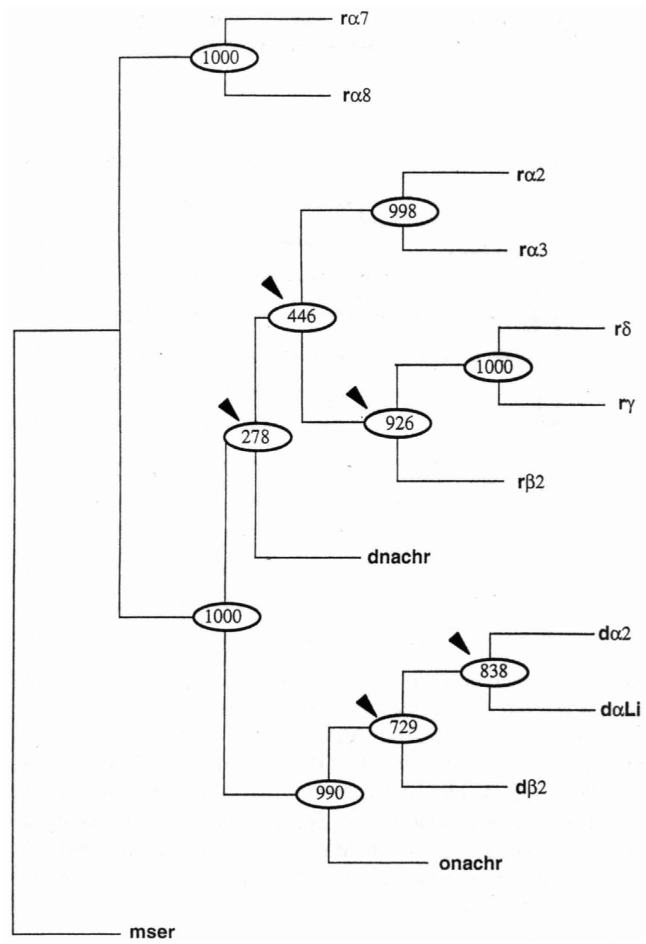

Figure 5: Bootstrap Majority-Rule consensus tree obtained from 1000 NJ replicates with the alignment shown in Appendix 3, presenting the possible emergence of the nematode subunit. The nodes indicated by an arrowhead are uncertain.

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. $\alpha 5$, which lacks some important aromatic amino acids in the third loop of 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), thus may represent a new type of "structural" subunit and should be given another name. $\alpha 5$ and $\beta 3$ could be called $\gamma 2$ and $\gamma 3$. 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 $\alpha 1$ (In addition to the ancestor of $\alpha 7$ and $\alpha 8$). The diversity of the group increased during the first 400 MYA, until the appearance of Tetrapoda. The whole evolution of the subfamily occurred in 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 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. 1994) as well as immunohistochemical (Britto et al. 1992; Hill et al. 1993) studies have shown that, in rat and chicken brain, $\alpha 4$ and $\alpha 2$ mRNA distribution is diffuse, whereas $\alpha 2, \alpha 3, \alpha 5, \alpha 6, \beta 3$ and $\beta 4$ are mainly restricted to a few major cholinergic or cholinoceptive pathways, which, however, also express $\alpha 4$ and $\beta 2$. $\beta 2$ has diverged early in the neuronal subfamily history (Figs. 2, 3 and results). $\alpha 4$ and $\beta 2$ 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** has to acquire a new role. This role can be defined by a new domain of expression (like the switches muscle/neuron developped before) or by a modified function. The hypothesis that $\alpha 4$ maintained its previous role while another **paralog** acquired a new role is supported by some evidence on the $\alpha 2$ subunit. In chick brain, the $\alpha 2$ 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 $\alpha 2$ (in contrast to $\alpha 4$ which maintained its ancient role). Then two independent

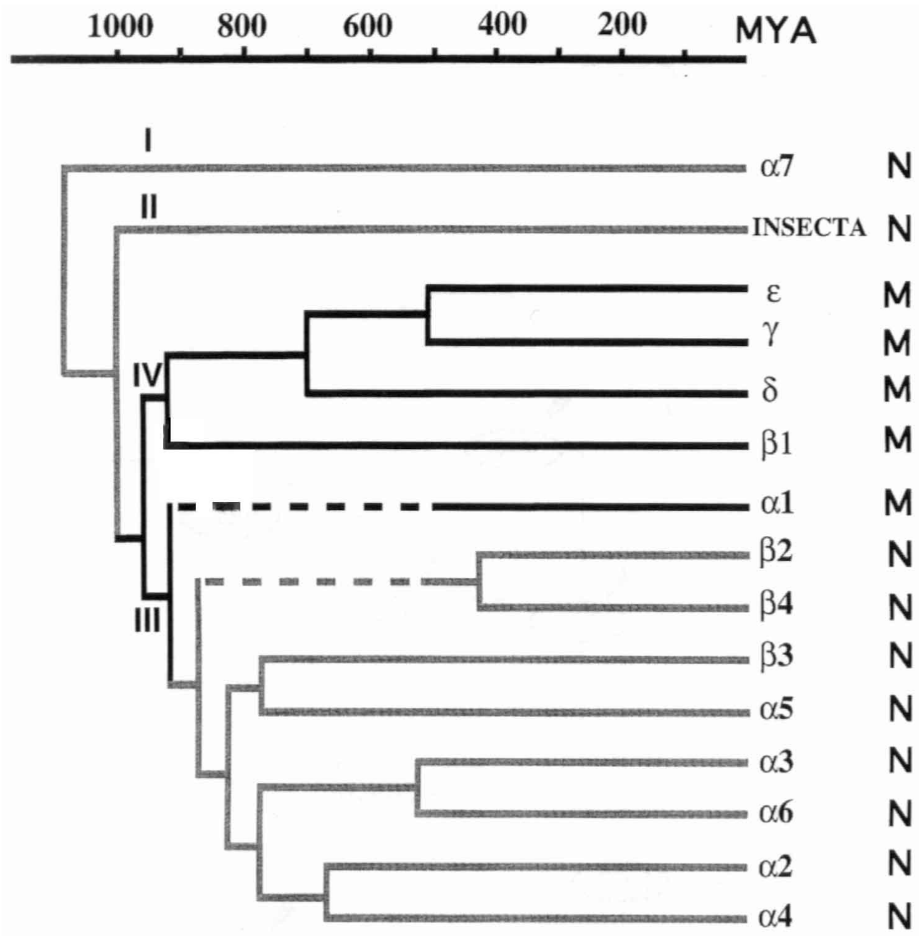

Figure 6: Summary tree, integrating the results of the whole study. The dates of the last divergences have been calculated from the protein alignment of Appendix 2 (cf Materiels and Methods). $\alpha 7/\alpha 8$: 380 MYA; ϵ/γ: 508 MYA; $\epsilon\gamma/\delta$: 711 MYA; $\epsilon, \gamma, \delta/\beta 1$: 926 MYA; $\alpha 3/\alpha 6$: 529 MYA; $\alpha 2/\alpha 4$: 669 MYA; $\alpha 5/\beta 3$: 770 MYA. The age of the precocious divergences have been approximatively infered from the divergence of nematodes (1000 MYA: Vanfleteren et al. 1994). M: muscle subunit; N: neuronal subunit. The grey branches represent subunits putatively expressed in neurons. The black branches represent subunits possibly expressed in muscle.

specificities of expression took place in the two phyla. Accordingly, a transgene with the avian $\alpha2$ 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 $\alpha2$ nor to that of $\alpha4$).

In the same way, the duplications $\alpha3/\alpha6$ and $\beta2/\beta4$ occurred a little before, or a little after, the split between the teleost and the tetrapod lineages. In the rat brain, $\alpha3$ and $\alpha4$ 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. $\alpha4$ and $\beta2$ 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 occured 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 and the Direction des Recherches et Etudes Techniques (contract no. 90/142).

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 266:11192-98.

- 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 Op 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). $\alpha 3, \alpha 5$, and $\beta 4$: 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. Sem 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 [^{3}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 $\alpha 5$ 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 ($\alpha 7$) is developmentally regulated and forms a homo-oligomeric channel blocked by α-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 Boston Berlin. pp 48-64.
- Daubas P, Devillers-Thiéry A, Geoffroy B, Martinez S, Bessis A, Changeux JP (1990). Differential expression of the neuronal acetylcholine receptor $\alpha 2$ 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 $\alpha 2$-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). $\beta 3$: 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, Maryland.
- 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, Lewis JA, Sattelle DB (1993). Acetylcholine receptor molecules of the nematode Caenorhabditis elegans. In: Pichon Y (ed) EXS 63, Comparative Molecular Neurobiology. Birkhäuser, Basel Boston Berlin. 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 unconventionnal allosteric proteins. Curr Op Struct Biol 4 (sous presse).
- 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 $\beta 2$ subunit. J Neurosci 13: 1551-1568.
- Jonas P, Baumann A, Merz B, Gundelfinger ED (1990). Structure and developmental expression of the Dα2 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 Op 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 Boston Berlin. 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- subunnits contribute to the agonist sensivity 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 a 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 Op 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, Schloss P, Betz H, Schmitt B (1990). Heterogeneity of Drosophila nicotinic receptors: SAD, a novel developmentally regulated a-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 a-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 E, Wada K, Boulter J, Deneris E, Heinemann S, Patrick J, Swanson LW (1989). Distribution of $\alpha 2, \alpha 3, \alpha 4$ and $\beta 2$ 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 ($\alpha 5$) in the rat central nervous system. Brain Res 526: 45-53.
- 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.
- Walker RJ, Colquhoun L, Holden-Dye L (1992). Pharmacological profiles of the GABA and acetylcholine receptors from the nematode, Ascaris suum. Acta Biol Hungar 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 (1994). Developmental regulation of nicotinic receptor subunit mRNAs in the rat central and peripheral nervous system. J Neurosci, submitted.

1 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 than other homologous characters.

Orthology Two orthologs are homologs arisen by speciation (the "same" gene in two different species).

Paralogy Two paralogs are homologs 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.

Plesiomorphy A plesiomorphy is an ancestor character. It is not useful to reconstruct phylogenies

Autapomorphy An autapomorphy is a derived character, present in one descendant only. It is not useful to reconstruct phylogenies.

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

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