J Mol Evol (1998) 47:323–333

JOURNAL OF MOLECULAR EVOLUTION



# Identification of Major Phylogenetic Branches of Inhibitory Ligand-Gated Channel Receptors

**Hong Xue**

Department of Biochemistry, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong

Received: 15 April 1997 / Accepted: 11 March 1998

**Abstract.** The gene superfamily of ligand-gated ion channel (LGIC) receptors is composed of members of excitatory LGIC receptors (ELGIC) and inhibitory LGIC receptors (ILGIC), all using amino acids as ligands. The ILGICs, including $GABA_A$, Gly, and GluCl receptors, conduct $Cl^-$ when the ligand is bound. To evaluate the phylogenetic relationships among ILGIC members, 90 protein sequences were analyzed by both maximum-parsimony and distance matrix-based methods. The strength of the resulting phylogenetic trees was evaluated by means of bootstrap. Four major phylogenetic branches are recognized. Branch I, called BZ, for the majority of the members are known to be related to benzodiazepine binding, is subdivided into IA, composed of all $GABA_A$ receptor α subunits, and IB, composed of the γ and ε subunits, which are shown to be tightly linked. Branch II, named NB for non–benzodiazepine binding, and consisting of $GABA_A$ receptor β, δ, π, and ρ subunits, is further subdivided into IIA, containing β subunits; IIB, containing δ, and π subunits; and IIC, containing ρ subunits. Branch IIIA, composed of vertebrate Gly receptors, is loosely clustered with Branch IIIB, composed of invertebrate GluCl receptors, to form Branch III, which is designated NA for being non–GABA responsive. Branch IV is called UD for being undefined in specificity. The existence of primitive forms of $GABA_A$ receptor non-β subunits in invertebrates is first suggested by the present analysis, and the identities of sequences p25123 from *Drosophila melanogaster*, s34469 from *Lymnaea stagnalis*, and u14635 and p41849 from *C. aenorhabditis elegans* are determined to be different from their previously given annotations. The proposed branching classification of ILGICs provides a phylogenetic map, based on protein sequences, for tracing the evolutionary pathways of ILGIC receptor subunits and determining the identities of newly discovered subunits on the basis of their protein sequences.

**Key words:** Classification — Evolution — γ-Aminobutyric $acid_A$ — Glutamate — Glycine — Ligand-gated channel — Phylogeny — Receptor

## Introduction

Ligand-gated ion channel (LGIC) receptors form a gene superfamily (Betz 1990) composed of excitatory cation channels gated by acetylcholine (nACh) and serotonin ($5HT_3$) and inhibitory anion channels gated by γ-aminobutyric acid ($GABA_A$), glycine (Gly), and glutamate (GluCl). The members of the superfamily share high degrees of amino acid sequence similarities and are therefore believed to have similar three-dimensional (3-D) structures (Unwin 1993). They are likely to be hetero- or homopentamers, with a large N-terminal extracellular domain followed by transmembrane and cytoplasmic domains (Karlin 1993). As members of the LGIC superfamily are often important pharmaceutical targets, their 3-D structural resolution would greatly facilitate the study of drug–receptor interactions. However, the oligomeric and membrane-bound nature of these receptors has so far obstructed all attempts to obtain suitable starting

*Correspondence to:* H. Xue; *e-mail:* hxue@usthk.ust.hk

**Table 1.** Fifty-eight sequences of ILGICs from vertebrates

| Receptor | Subunit | Accession No. | | | | | |
|---|---|---|---|---|---|---|---|
| | | Human | Bovine | Rat | Mouse | Chicken | Goldfish |
| $GABA_A$ | α1 | p14867 | p08219 | p18504 | — | p19150 | — |
| | α2 | p47869 | p10063 | p23576 | p26048 | — | — |
| | α3 | p34903 | p10064 | p20236 | p26049 | — | — |
| | α4 | p48169 | p20237 | p28471 | — | — | — |
| | α5 | p31644 | — | p19969 | — | — | — |
| | α6 | s81944 | — | p30191 | p16305 | — | x94342 |
| | γ1 | — | — | p23574 | — | — | — |
| | γ2 | p18507 | p22300 | p18508 | p22723 | p21548 | — |
| | γ3 | — | — | p28473 | p27681 | — | — |
| | γ4 | — | — | — | — | p34904 | — |
| | ε | y07637 | — | — | — | — | — |
| | β1 | p18505 | p08220 | p15431 | p50571 | — | — |
| | β2 | p47870 | — | p15432 | — | — | — |
| | β3 | p28472 | — | p15433 | — | p19019 | — |
| | β4 | — | — | — | — | p24045 | — |
| | δ | — | — | p18506 | p22933 | — | — |
| | π | u95367 | — | u95368 | — | — | — |
| | ρ1 | p24046 | — | p50572 | — | — | — |
| | ρ2 | p28476 | — | p47742 | — | — | — |
| | ρ3 | — | — | p50573 | — | — | — |
| Gly | α1 | p23415 | — | p07727 | — | — | — |
| | α2 | p23416 | — | p22771 | — | — | — |
| | α3 | — | — | p24524 | — | — | — |
| | β | p48167 | — | p20781 | p48168 | — | — |

materials for 3-D structural studies by either X-ray or nuclear magnetic resonance (NMR). The lack of information about the origin of LGIC receptors has also hampered the application of homology modeling to elucidate the tertiary structure of these receptors.

As a result of the combined power of patch clamping and molecular genetics, new ion channel receptors from various phylogenetic sources are constantly being cloned and characterized (Betz 1990). These advances along with progress in genomic sequencing, including the complete sequencing of four entire genomes from three phylogenetic kingdoms (Garret 1996), have made possible an evolutionary analysis of the LGIC receptor superfamily. The molecular evolution of nACh receptors, an excitatory LGIC (or ELGIC) multigene family, has been the topic of several publications (Novére and Changeux 1995; Ortells and Lunt 1995; Gundelfinger 1995), in contrast to the much less discussed inhibitory LGIC (or ILGIC) receptors. Recently, Ortells and Lunt (1995) analyzed the evolutionary history of the whole superfamily of LGIC receptors, including 47 $GABA_A$ and Gly receptors, but not the GluCl receptors. Their study was based on nucleic acid sequences of the receptor subunits. In contrast, the present study attempts to infer phylogenies from protein sequences, since the greater conservation of amino acid sequences should allow us to bring more information to bear on ancient origins of lineages. Only the phylogeny of ILGICs as a subissue of the phylogeny of the whole LGIC superfamily is addressed. Accordingly, 90 protein sequences of ILGIC members, including $GABA_A$, Gly, and GluCl receptor subunits, are statistically analyzed for their evolutionary relationships using both cladistic and phenetic methods supported by the use of the resampling method, bootstrap, to test the robustness of resulting tree nodes. Known ELGIC receptor subunits are used as the outgroup, which offers the best basis for placing the root on the phylogenetic trees. The level of sequence density employed has allowed the delineation of four major phylogenetic branches. This branching classification provides a basis for not only tracing the evolutionary history of ILGICs, but also predicting the ligand specificity of newly discovered ILGIC receptor subunits.

## Materials and Methods

### *Sequence Data*

Protein sequences were obtained from the SwissProt (release 34.0) or PIR (release 50.0) databases or deduced from DNA sequences in the GenBank (release 97.0) database. We used the tools provided with the Sequence Analysis Software Package, GCG, Version 8.1.0 (Genetic Computer Group 1994), EGCG (Rice 1996), and Entrez available at the National Center of Biotechnology Information for database search and sequence conversion. The sequences employed are listed in Table 1 for vertebrates and Table 2 for invertebrates.

### *Alignment of Sequences*

Multiple sequence alignments were performed by means of the program PILEUP from the GCG package, according to the empirical scor-

**Table 2.** Thirty-six sequences of ILGICs from invertebrates

| Lineage | Species (common name) | Acc. No.[a] | Receptor (alternative name) | Subunit | Suggested identity[b] |
|---|---|---|---|---|---|
| Arthropoda | *Drosophila melanogaster* (fruit fly) | **x78349** | LGIC (Grd, Gly-like) | | GA α/γ-like |
| | | p25123 | GABA (LCCH3, gab_drome) | β | GL/GA |
| | | q08832 | GABA (gab3_drome) | β-like | GA β |
| | | m69057 | GABA (Rdl; cyclodiene resistant) | | GL/GA |
| | | u02042 | GABA | | GL/GA |
| | | u58776 | GluCI | | GluCI α |
| Arthropoda | *Aedes aegypti* (yellow fever mosquito) | s33744 | GABA-like (Rdl; cyclodiene resistant) | | GL/GA |
| Mollusca | *Lymnaea stagnalis* (great pond snail) | p26714 | GABA (gab_lymst) | β | GA β |
| | | s34469 | GABA-like | ζ | GL |
| Nematoda | *Caenorhabditis elegans* | u40187 | GABA-like | | GA/GL |
| | | z50016 | GABA-like | | GA/GL |
| | | u14525 | GluCl | β | GluCl β |
| | | s50864 | GluCl | α | GluCl α |
| | | **u41113** | ILGIC (Cegbr 3) | | GluCl α |
| | | **u40573** | ILGIC (Cegbr 2) | | GluCl α |
| | | **u64840** | LGIC-like | | GluCl α |
| | | u14635 | GABA/Gly-like | | GluCl |
| | | z70270 | GABA | β-like | ILGIC |
| | | z74040 | GABA | β-like | ILGIC |
| | | p41849 | Gly-like (yo99_caeel) | | ILGIC |
| | | u28929 | Gly | α-like | ILGIC |
| | | z68217 | Gly-like | | ILGIC |
| | | u40422 | GABA-like | | ILGIC |
| | | z46791 | GABA | β1-like | ILGIC |
| | | z50027 | GABA-like | | ILGIC |
| | | u64843 | LGIC-like | | ILGIC |
| | | u40948 | LGIC-like | | ILGIC |
| | | z49888 | LGIC-like | | ILGIC |
| | | u40573 | LGIC-like (Cegbr2) | | — |
| | | u42836 | LGIC-like | | — |
| | | u59743 | GluCl | | — |
| | | u59744 | GluCl | | — |
| | | u64840 | LGIC-like | | — |
| Nematoda | *Haemonchus contortus* | x73584 | ILGIC-like | | Ga α/γ-like? |
| | | y09796 | GluCl | β | GluCl β |
| Nematoda | *Onchocerca volvulus* | u59745 | GluCl | | GluCl α |

[a] Accession No.: underlined, suggested identity is different from previous annotation; boldface, newly suggested identity.
[b] GA, $GABA_A$; GL, GluCl/Gly; ILGIC, undifferentiated ILGIC; —, not included in the present analysis.

ing matrix PAM250 (Dayhoff 1979), with a gap creation weight of 6.0 and a gap length weight of 0.01. Most variable sequences from the N and C termini are excluded from further analysis. The alignments were cross-checked using a more sensitive program, CLUSTAW (Thompson et al. 1994), which made no difference to the final results.

### *Phylogenetic Analyses*

Multiple alignments of amino acid sequences were analyzed using the PHYLIP 3.57c software package (Felsentein 1993). For cladistic analysis, the program PROTPARS based on the maximum-parsimony algorithm (Fitch 1971) was used to construct phylogenetic trees. For phenetic analysis, the program FITCH (Fitch and Margoliash 1967) was employed. Based on sequence alignments corrected distances were calculated according to the PAM250 scoring matrix using PRODIST of the PHYLIP package and scaled in expected historical events per site (Dayhoff 1979). Gaps were treated as missing data and excluded from the calculations. The matrices were then used to construct additional trees by the least-squares method of program FITCH from the PHYLIP package (Fitch and Margoliash 1967). The strength of the tree topology was tested by bootstrap analysis (Felsenstein 1985) with either 100 or 1000 (as specified in the figure legends) replications employing the program SEQBOOT. Majority-rule consensus trees were obtained with the program CONSENSE.

## Results and Discussion

Ninety-four protein sequences (Tables 1 and 2) were retrieved from databases. They all have some highly characteristic sequence motif, for example, a 15-residue cysteine (Cys) loop in the N-terminal domain, and all are annotated as ligand-gated $Cl^-$ ion channel receptors regardless of the ligand being GABA, Gly, Glu, or, in some cases, unidentified. Eighty-eight sequences of ninety-five were aligned with PILEUP, five of the remaining seven sequences bearing excessive variations and two others being published recently (Hedblom and

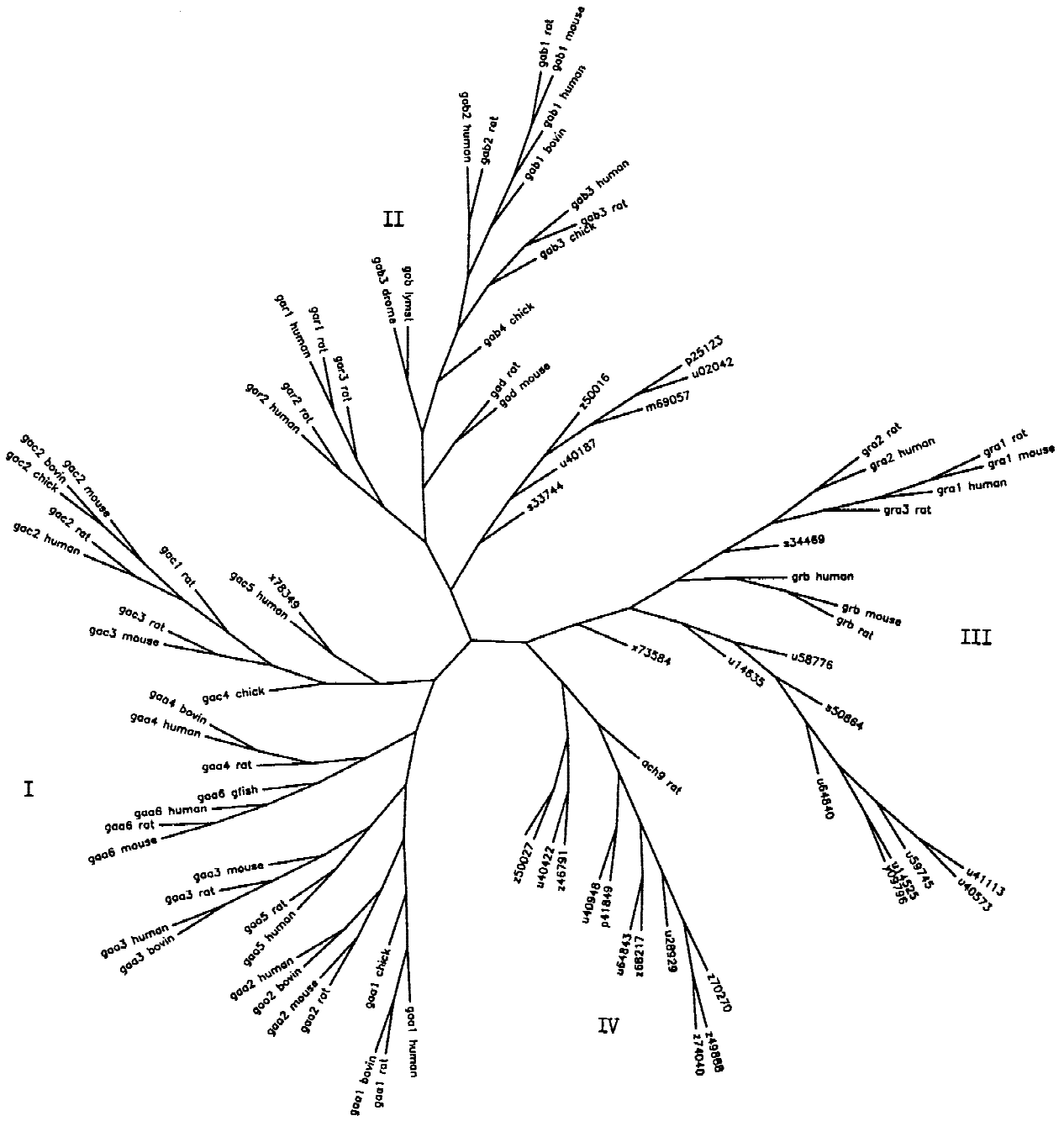


**Fig. 1.** Bootstrap majority-rule consensus tree obtained from 100 replicates (SEQBOOT, PROTPARS, and CONSENSE programs) of 88 ILGIC sequences using nACh receptor sequence ach9_rat as an outgroup with 80 informative sites. For clarity, bootstrap numbers are omitted. Abbreviations used in Figs. 1–4: GAB, $GABA_A$ receptor; GLY, Gly receptor; GLU, Glu receptor; a, α subunit; b, β subunit; g, γ subunit; d, δ subunit; e, ε subunit; p, π subunit; r, ρ subunit; chick, chicken; gfish, goldfish; drome, *Drosophila melanogaster;* lymst, *Lymnaea stagnalis;* caeel, *Caenorhabditis elegans;* haeco, *Haemonchus contortus;* oncvo, *Onchocerca volvulus.*

Kirkness 1997). Two most highly conserved regions are revealed by the alignment. The first region covers the 15-residue Cys loop, while the second spans the three tentative transmembrane segments M1–M3. Other regions of the sequences are generally more variable than these two. For example, a LGIC-like sequence (Accession No. x78349 (Harvey et al. 1994) from *Drosophila melanogaster* has a long insertion of about 70 residues between the two regions. The consensus tree obtained from 100 bootstrap replicates is shown in Fig. 1.

To carry out a thorough search of tree topologies, the 90 sequences were divided into smaller groups and subjected to further extensive analysis. Results from analyzing various combinations of sequences are summarized in Table 3. Only one example of such group analysis is provided graphically (Figs. 2 and 3). This group is a

**Table 3.** Branch classification of ILGIC receptor subunits

| Branch | Name[a] | Members |
|---|---|---|
| I | BZ | $GABA_A$ receptor α and γ subunits plus the new class, ε subunit |
| IA | | $GABA_A$ receptor α subunits |
| IB | | $GABA_A$ receptor γ and ε subunits |
| II | NB | $GABA_A$ receptor β, δ, and ρ subunits plus the new class, π subunit |
| IIA | | $GABA_A$ receptor β subunits |
| IIB | | $GABA_A$ receptor δ and π subunits |
| IIC | | $GABA_A$ receptor ρ subunits |
| III | NA | Gly and GluCl receptor subunits |
| IIIA | | Gly receptor subunits |
| IIIB | | GluCl receptor subunits |
| IV | UD | p41849, u28929, u40422, u40948, u64843, z46791, z50027, z68217, z70270, z74040, z49888 |
| Intermediates | | |
| I-like | | x78349 |
| I/II/III-like | | m69057, u02042, s33744, u40187, z50016, p25123 |

[a] BZ, relevant to benzodiazepine binding; NB, not involved in benzodiazepine binding; NA, nonresponsive to GABA; UD, specificity undefined.

collection of 30 representative sequences (Figs. 2 and 3). This sample collection included one sequence from each subtype of the GABA, and Gly receptors, along with the nACh receptor subunit ach9_rat as an outgroup. Sequences from rat were chosen from orthologs wherever feasible. For $GABA_A$ receptor γ4 and β4, for which no rat orthologs were available, chicken sequences were employed. Seven invertebrate sequences were also included due to their being significantly different from the vertebrate sequences. Among them, three were from *Drosophila melanogaster*, i.e., gab_drome, x78349, and u58776. Two of them, gab_lymst and s34469, were from *Lymnaea stagnalis*, while the other two, u14525 and y09796, were from *Caenorhabditis elegans* and *Haemonchus contortus*, respectively. Other group-analysis results are available on request.

### *Branches of ILGICs*

Four major branches were delineated by both the maximum-parsimony (Figs. 1 and 2) and the distance matrix-based (Fig. 3) methods. Branches I and III each can be further subdivided into two secondary branches, A and B, while Branch II is divided into A, B, and C (Table 3). Alignment with either ach1_caeel or 5ht3_mouse as an outgroup did not change these tree topologies. The ranges of pairwise distances between and within branches are presented in Table 4 to provide a numerical estimation of the relationships among the sequences. While the interbranch distances point to the relative closeness between branches, the intrabranch distances shown on the diagonal in Table 4 are indicative of the relative closeness between members of the same branch. The striking difference in pairwise distance between orthologs (Table 5) suggests an inconstancy in evolutionary rates of ILGICs. Therefore no attempt was made to estimate the divergence times using the molecular clock assumption.

### *$GABA_A$ Receptor Subunits*

Fifty-three $GABA_A$ receptor sequences (Tables 1 and 2), forty-four from mammals, five from chicken, one from goldfish, two from fruit fly, and one from the great pond snail, were analyzed. Among these are two newly discovered subunit classes, namely, ε (Davies et al. 1997) and π (Hedblom and Kirkness 1997) subunits. All known subtypes of $GABA_A$ receptor subunits can be divided into the two monophyletic Branches I and II (Figs. 2 and 3). Branch I comprises all known $GABA_A$ receptor α and γ subunit sequences, plus the new class, the ε subunit. As the α and γ subunits are known to carry, respectively, the principal and the complementary parts of benzodiazepine sites, Branch I is called BZ. Branch I is further divided into two monophyla, named IA for all α subunits and IB for all γ and ε subunits. Branch II, composed of $GABA_A$ receptor β, δ, and ρ subunits as well as the new class, the π subunit, is designated NB because a majority of the members of this branch are known not to be directly relevant to benzodiazepine binding. Under Branch II, all β subunits clearly form the monophylum IIA, and all ρ subunits form another monophylum, IIC. Bootstrap did not strongly support the assignment of δ subunits as either IIA or IIC members. They are therefore assigned to a separated secondary branch, IIB. The newly discovered π subunit is assigned to IIB based on our analysis.

Most recently, Davies et al. (1997) suggested the occurrence of a new class of human $GABA_A$ receptor subunits, ε, based on its pharmacological property of conferring insensitivity to the potentiating effects of the intravenous anesthetic agents propofol, pentobarbital, and pregnanolone. The sequence of the ε subunit is nearly identical to that of a putative GABA-gated chloride channel subunit expressed in the human cardiac conduction system (Garret et al. 1997), with the exception that the former has one extravaline (Val) residue at position 261. The latter was not included in our analysis, to avoid redundancy. The ε subunit is shown by our analysis (Figs. 2 and 3) to form a monophyletic group with all the other γ subunits. In this sense, it may be regarded as a type of γ subunit rather than a separate class.

### *Gly and GluCl Receptor Subunits*

Branch III is named NA for being non–GABA responsive; it is essentially a collection of chloride channel receptors gated by either Gly or Glu. The fact that

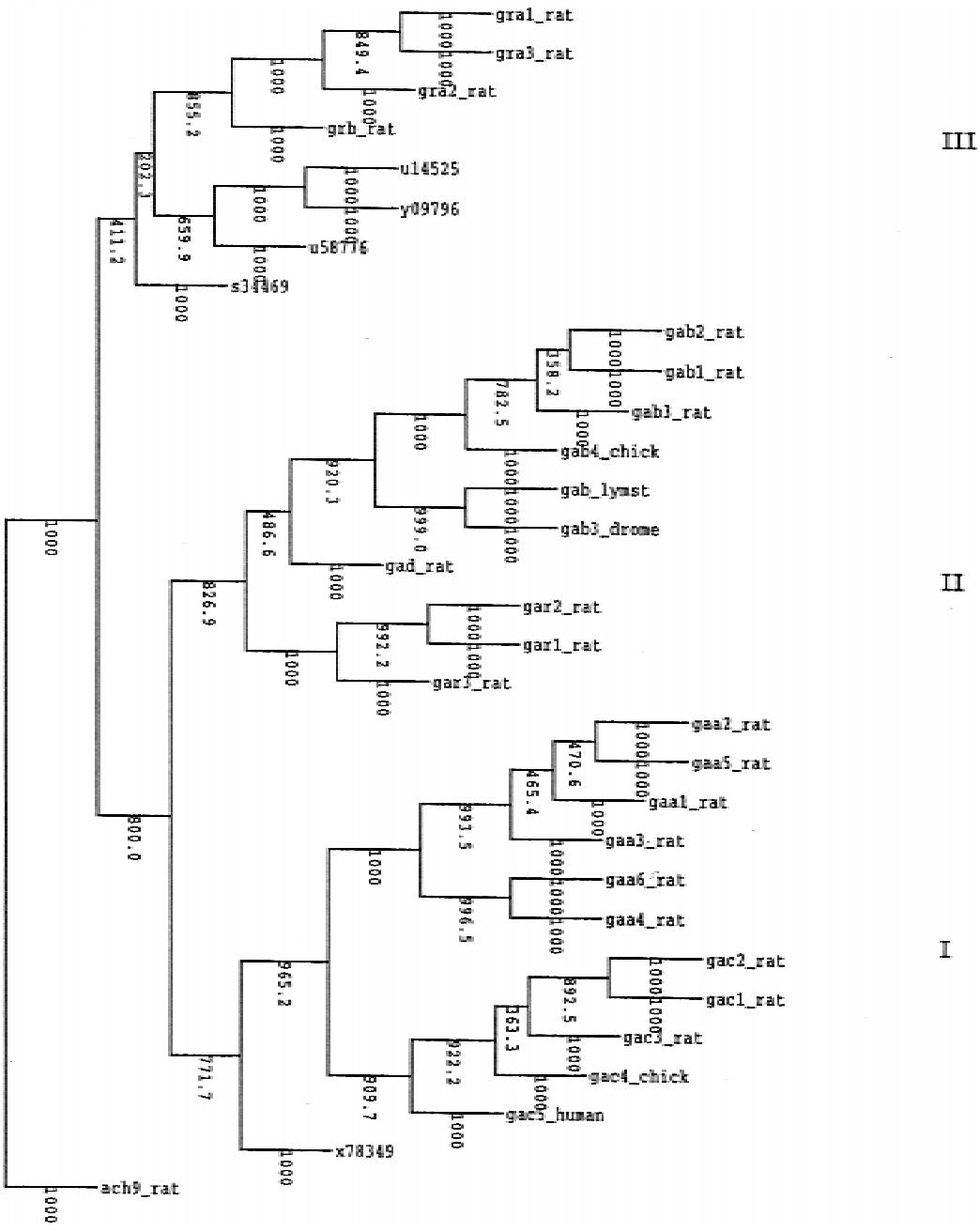


**Fig. 2.** Bootstrap majority-rule consensus tree obtained from 1000 replicates (SEQBOOT, PROTPARS, and CONSENSE programs) of 30 ILGIC sequences along with nACh receptor sequence ach9_rat as an outgroup. The reconstruction is based on the alignment shown in Appendix 1 with 267 informative sites. The *numbers beneath the branches* give the bootstrap values of 1000 replications.

$GABA_A$ and Gly receptors are closely related has long been noted from their high sequence similarity. This close relationship is statistically confirmed by the present analysis. In an evolutionary analysis of the whole LGIC receptor superfamily, including ELGICs and ILGICs, Ortells and Lunt (1995) suggested that Gly receptors were derived from $GABA_A$ receptors. Although the present study does not provide strong support for this hypothesis, the closeness between $GABA_A$ and Gly receptors is evident from the pairwise distance values (Table 4). It has been noted that the sequences of GluCl receptors are related to those of Gly and $GABA_A$ receptors (Cully et al. 1994). Figures 2 and 3 further suggest that GluCl receptors are closer to Gly receptors than they are to $GABA_A$ receptors.

The clustering of Gly and GluCl receptors into Branch III is indicated by both cladistic (Fig. 2) and phenetic (Fig. 3) analyses, but only weakly supported by bootstrap. This looseness in the clustering as shown by the low bootstrapping rate and the relatively large distances (Table 4) between Gly and GluCl suggest that contemporary Gly and GluCl receptors are derived from a rela-

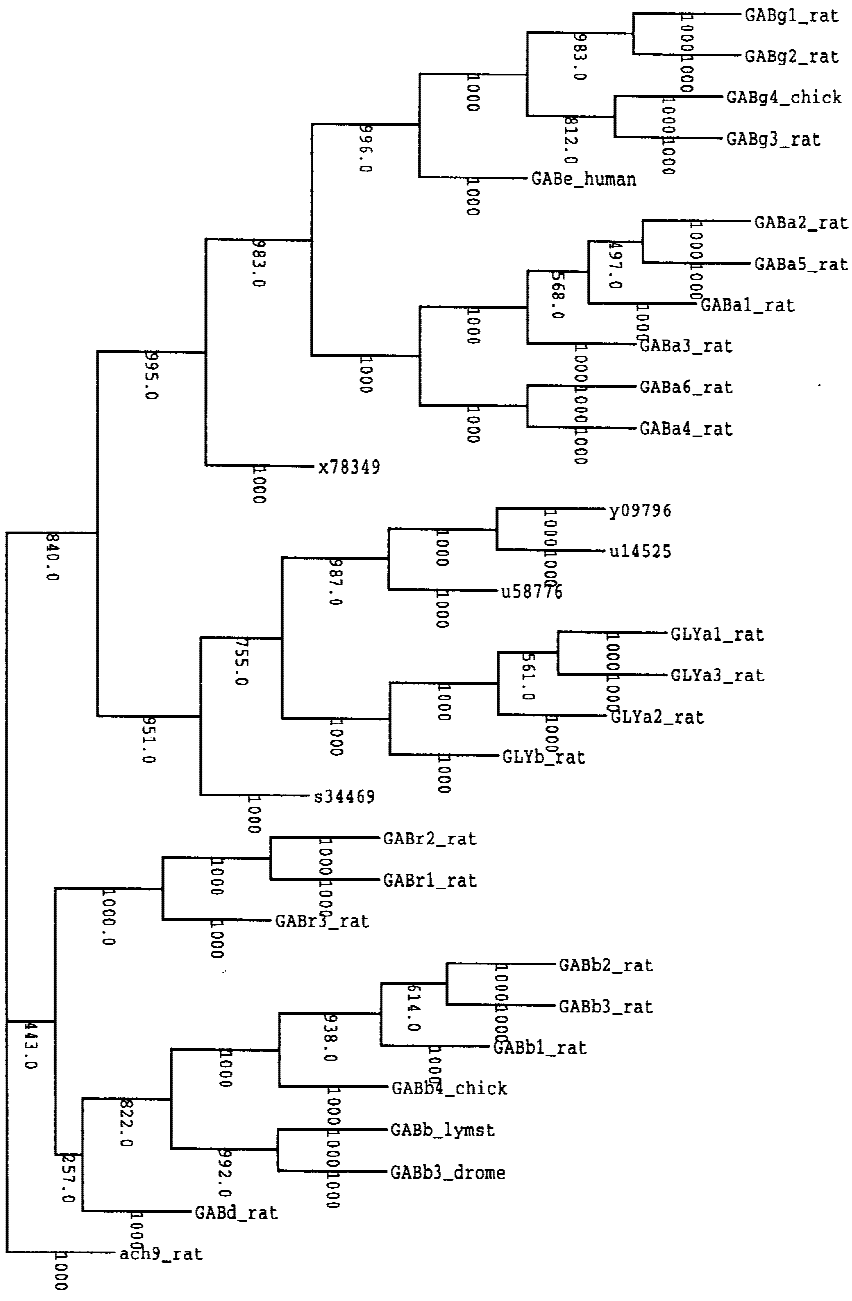


**Fig. 3.** Bootstrap majority-rule consensus tree obtained from 1000 replicates (SEQBOOT, PROTDIST, FITCH, and CONSENSE programs) of 30 ILGIC sequences along with nACh receptor sequence ach9_rat as an outgroup. The reconstruction is based on the alignment shown in Appendix 1 with 267 informative sites. The *numbers beneath the branches* give the bootstrap values of 1000 replications.

tively remote common ancestor. The internal stabilities of the two secondary branches, IIIA, with all the Gly receptors, and IIIB, with all the GluCl receptors, are strongly established by bootstrapping. This confers substantial confidence in the identification of any potential Gly or GluCl receptors on the basis of protein sequence alone.

So far no GluCl receptor has been discovered in vertebrates, and likewise no Gly receptor could be confirmed in invertebrates (Laughton et al. 1994). Three sequences from *C. elegans*, namely, p41849, u28929, and z68217, were previously annotated as Gly or Gly-like receptor subunits (Table 2). Instead, they are found to be members of Branch IV (Table 3). This would suggest that the specialization of Gly and GluCl receptors must have occurred after the divergence of invertebrates and vertebrates. Given the clustering of Branch IIIA with Branch IIIB, Gly and GluCl receptors possibly represent the vertebrate and invertebrate counterparts of one another. This agrees with the suggestion of Vassilatis et al. (1997), based on their gene structure comparisons and phylogenetic analysis, that invertebrate GluCl receptors may be orthologous to the vertebrate Gly receptors, although no outgroup was employed in their analysis.

### *ILGIC Receptor in Invertebrates*

Invertebrate ILGIC sequences are listed in Table 2. The sequences include those obtained by Wilson et al. (1994)

**Table 4.** Distances between ILGIC branches

| | Branch | | | | | | |
|---|---|---|---|---|---|---|---|
| Branch | IA | IB | IIA/B | IIC | IIIA | IIIB | IV |
| IA | 0.66–61.62 | 70.45–93.69 | 96.32–126.38 | 105.13–135.82 | 95.89–145.25 | 95.55–167.90 | 124.47–177.95 |
| IB | | 1.08–40.47 | 93.44–111.01 | 110.31–125.35 | 98.35–130.09 | 103.54–152.80 | 117.35–162.44 |
| IIA | | | 0.21–98.88 | 93.18–114.42 | 93.55–121.01 | 97.98–150.86 | 119.23–169.79 |
| IIB | | | | 5.44–48.72 | 109.12–129.29 | 104.40–154.24 | 121.70–159.81 |
| IIIA | | | | | 0.22–78.30 | 87.86–156.69 | 109.05–160.20 |
| IIIB | | | | | | 17.58–115.59 | 100.69–195.52 |
| IV | | | | | | | 51.0–178.53 |

**Table 5.** Distances between orthologs

| | Distance | |
|---|---|---|
| Subunit | Human/rat | Rat/mouse |
| $GABA_A$ receptor | | |
| α1 | 1.55 | — |
| α2 | 3.16 | 1.56 |
| α3 | 3.94 | 1.85 |
| α4 | 11.18 | — |
| α5 | 4.89 | — |
| α6 | 9.02 | 2.98 |
| β1 | 1.92 | 0.21 |
| β2 | 0.21 | — |
| β3 | 2.79 | — |
| γ2 | 1.08 | 1.08 |
| δ | — | 0.45 |
| ρ1 | 5.44 | — |
| ρ2 | 9.49 | — |
| Gly receptor | | |
| α1 | 2.48 | 0.22 |
| α2 | 1.11 | — |
| β | 2.66 | 0.81 |

from a 2.2-Mb fragment of *C. elegans* chromosome III (Wilson et al. 1994), with at least 24 open reading frames encoding proteins bearing sequence similarity to LGIC receptors. All 11 current members of Branch IV (Tables 2 and 3) are from *C. elegans.* It remains to be seen whether the completion of additional genome sequencing will reveal additional Branch IV members that could shed more light on the nature of this phylogenetic branch. The distances between Branch IV and each of other three branches are of the same order of magnitude (Table 4). In view of the relatively large distances between Branch IV and the other branches, the separations of the three other branches from Branch IV were evidently remote historical events. The large intrabranch distances within Branch IV suggest that the original divergences of the various Branch IV members were likewise ancient occurrences.

The classification scheme developed and illustrated in Figs. 1–3 suggests the nature of many of these invertebrate sequences (Tables 2 and 3) whose specifications were hitherto unknown. These include the assignment of u41113, u40573, and u64840 from *C. elegans* to Branch IIIB (Figs. 2 and 3, Tables 2 and 3), thereby implying a GluCl identity, and x78349 from *Drosophila melanogaster* to Branch 1, thereby implying a $GABA_A$ α/γ identity, although the large insertion in its extracellular portion may well have altered its ligand specificity.

Instances where the present branching classification results in a revision of earlier annotations include the placement of s34469 from the great pond snail, previously designated the $GABA_A$ receptor ζ subunit (Hutton et al. 1993), closer to Gly/GluCl than to $GABA_A$ receptors by both the PROTPARS (Fig. 2) and the FITCH (Fig. 3) procedures; also, P41849 from *C. elegans* is grouped with other Branch IV sequences instead of Gly sequences as had been suggested (Swiss-Prot annotation). In yet another example, *C. elegans* u14635 (Wilson et al. 1994), previously annotated as a $GABA_A$/Gly subunit (Swiss-Prot annotation), is assigned to Branch IIIB (Figs. 2 and 3, Table 3), which implies a potential GluCl identity.

The monophylum consisting of p25123, m69057, and u02042 from *Drosophila melanogaster,* u40187 and z50016 from *C. elegans,* and s33744 from *Aedes aegypti,* represents an intermediate group. All six sequences were previously annotated as GABA or GABA-like, but the present study suggests that they are more likely to be evolutionary intermediates rather than well-differentiated $GABA_A$ receptor subunits. Two of these six sequences, namely, m69057 and s33744, are known to confer cyclodiene resistance (Table 2).

In invertebrates, hitherto only β subunits have been encountered, and there has been no suggestion made of the occurrence of $GABA_A$ non-β subunits in invertebrates. In the present analysis, however, the *Drosophila melanogaster* sequence x78349 is found to cluster with the non-β Branch I sequences (Figs. 2 and 3). The distances between x78349 and Branch I vertebrate sequences are 93.66–120.73, which are smaller than the distances of 123.83–135.79 between x78349 and other *Drosophila melanogaster* ILGICs. Therefore x78349 is closer to Branch I than to other *Drosophila melanogaster* sequences, with the major difference between x78349 and Branch I sequences arising from a single large insertion at its amino-terminal domain (Harvey et al. 1994). In line with this tentative identification of x78349

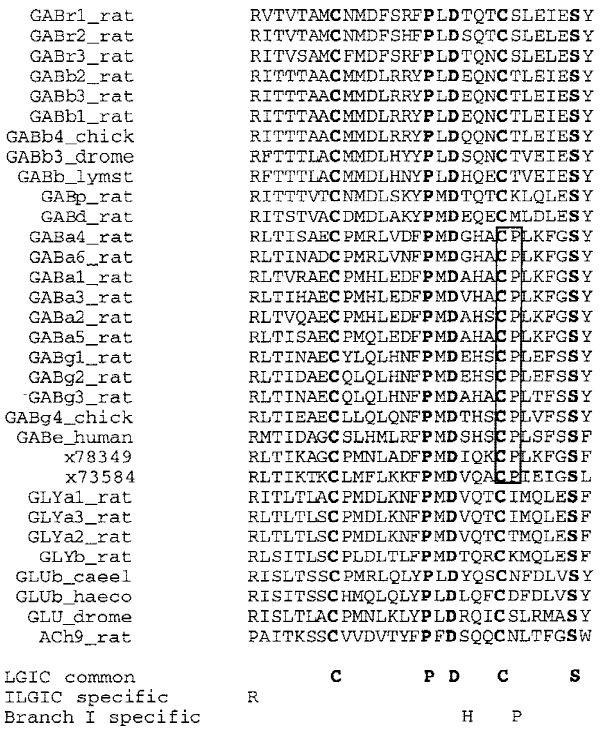


**Fig. 4.** The neighborhood of the conserved Cys–Cys loop in 31 representative ILGIC subunits aligned with the outgroup ACh9_rat. Residues common to both ILGIC and ELGIC are in *boldface.* They are indicated at the *bottom,* along with residues specific to ILGIC or Branch I members. The Cys–Pro doublet characteristic to Branch I is *boxed.*

as a primitive form of the $GABA_A$ receptor $\alpha/\gamma$ subunit, sequence x73584 from the nematode *Haemonchus contortus* also displays $\alpha/\gamma$-like properties, showing in the C–C loop region the characteristic Cys–proline (Pro) doublet that is unique for Branch I sequences (Fig. 4). Based on x78349 and supported by x73584, the present analysis therefore provides evidence for the first time of the existence of primitive non-$\beta$ $GABA_A$ receptor subunits among invertebrates.

*Acknowledgments.* I would like to thank Dr. J. Tze-Fei Wong for expert discussion and critical review of the manuscript and Dr. J. Felsenstein for provision of the PHYLIP program. The computational facility of SEQNET at Daresbury Laboratory, Science & Engineering Research Council, Warrington, UK, has been most helpful in the completion of this project. Support by the Hong Kong Government Industry Department is gratefully acknowledged.

## References

Antkiewicz-Michaluk L, Guidotti A, Kruger KE (1988) Molecular characterization and mitochondrial density of a recognition site for peripheral-type benzodiazepine ligands. Mol Pharmacol 34:272–278

Arena JP, Liu KK, Paress PS, Cully DF (1991) Avermectin-sensitive chloride currents induced by *Caenorhabditis elegans* RNA in *Xenopus oocytes.* Mol Pharmacol 40:368–374

Arena JP, Liu KK, Paress PS, Schaeffer JM, Cully DF (1992) Expression of a glutamate-activated chloride current in *Xenopus oocytes* injected with *Caenorhabditis elegans* RNA: evidence for modulation by avermectin. Mol Brain Res 15:339–348

Betz H (1990) Ligand-gated ion channels in the brain: the amino acid receptor superfamily. Neuron 5:383–392

Cully DF, Vassilatis DK, Liu KK, Paress PS, Van der Ploeg LHT, Schaeffer JM, Arena JP (1994) Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans.* Nature 371:707–711

Cully DF, Paress PS, Liu KK, Schaeffer JM, Arena JP (1996) Identification of a *Drosophila melanogaster* glutamate-gated chloride channel sensitive to the antiparasitic agent avermectin. J Biol Chem 271:20187–20191

Davies PA, Hanna MC, Hales TG, Kirkness EF (1997) Insensitivity to anaesthetic agents conferred by a class of $GABA_A$ receptor subunit. Nature 385:820–823

Dayhoff MO (1979) Atlas of protein sequence and structure, Vol 5, Suppl 3, National Biomedical Research Foundation, Washington, DC

Felsentein 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. Department of Genetics, University of Washington, Seattle

Ffrench-Constant RH, Rocheleau TA, Steichen JC, Chalmers AE (1993) A point mutation in a *Drosophila melanogaster* GABA receptor confers insecticide resistance. Nature 363:449–451

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

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Garret M, Bascles L, Boue-Grabot E, Sartor P, Charron G, Bloch B and Margolskee RF (1997) An mRNA encoding a putative GABA-gated chloride channel is expressed in the human cardiac conduction system. J Neurochem 68:1382–1389

Garrett RA (1996) Genomes: *Methanococcus jannaschii,* the golden fleece. Curr Biol 6:1377–138

Geary TG, Klein RD, Vanover L, Bowman JW, Thompson DP (1992) The nervous system of helminths as targets for drugs. J Parasitol 78:215–230

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

Gundelfinger ED (1995) Evolution, desensitization of LGIC receptors. Trends Neurosci 18:297

Harvey RJ, Schmitt B, Hermans-Borgmeyer I, Gundelfinger ED, Betz H, Darlison MG (1994) Sequence of a *Drosophila melanogaster* ligand-gated ion-channel polypeptide with an unusual amino-terminal extracellular domain. J Neurochem 62:2480–2483

Hedblom E and Kirkness EF (1997) A novel class of GABAA receptor subunit in tissues of the reproductive system. J Biol Chem 272: 15346–15350

Holden-Dye L, Krogsgaard-Larsen P, Nielsen L, Walker RJ (1989) GABA receptors on the somatic muscle cells of the parasitic nematode, *Ascaris suum:* stereoselectivity indicates similarity to a $GABA_A$-type agonist recognition site. Br J Pharmacol 98:841–850

Hutton ML, Harvey RJ, Earley FG, Barnard EA, Darlison MG (1993) A novel invertebrate $GABA_A$ receptor-like polypeptide: sequence and pattern of gene expression. FEBS Lett 326:112–116

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

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

Laughton DL, Amar M, Thomas P, Towner P, Harris P, Lunt GG, Wolstenholme AJ (1994) Cloning of a putative inhibitory amino acid receptor subunit from the parasitic nematode *Haemonchus contortus.* Receptors Channels 2:155–163

Le Novére N, Changeux JP (1995) Molecular evolution of the nicotinic

acetylcholine receptor: an example of multigene family in excitable cells. J Mol Evol 40:155–172

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

Martin RJ, Pennington AJ, Duittoz AH, Robertson S, Kusel JR (1991) The physiology and pharmacology of neuromuscular transmission in the nematode parasite. *Ascaris suum.* Parasitology 102 (Suppl):S41–S58

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

Ortells MO, Lunt GG (1995) Evolutionary history of the ligand-gated ion-channel superfamily of receptors. Trends Neurosci 18:121–127

Rich P et al. (1996) The EGCG extension to the Wisconsin Package Version 8.1.1, Sanger Center, Hinxton, England

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Unwin N (1993) The nicotinic acetylcholine receptor at 9 Å resolution. J Mol Biol 229:1101–1124

Walker RJ, Holden-Dye L (1991) Evolutionary aspects of transmitter molecules, their receptors and channels. Parasitology 102 (Suppl):S7–S29

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

Wieland HA, Lüddens H, Seeburg PH (1992) Molecular determinants in $GABA_A$/BZ receptor subtypes. In: Biggio G, Concas A, Costa E (eds) GABAergic synaptic transmission. Raven Press, New York, pp 29–40

Wilson R, Ainscough R, Anderson K, et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.* Nature 368:32–38

## Appendix

```
                1                                                  50
  gar1_rat  GGPAIPVGVD VQVESLDSIS EVDMDFTMTL YLRHYWKDER LSPGRLVKKI
  gar2_rat  GGPAIPVGVD VQVESLDSIS EVDMDFTMTL YLRHYWRDER LAFGRLVKKI
  gar3_rat  GGSPVPVGID VQVESIDSIS EVNMDFTMTF YLRHYWKDER LSFRRLIQKI
  gab2_rat  GGPPVAVGMN IDIASIDMVS EVNMDYTLTM YFQQAWRDKR LSYNRVADQL
  gab3_rat  GGPPVCVGMN IDIASIDMVS EVNMDYTLTM YFQQYWRDKR LAYNRVADQL
  gab1_rat  GGPPVDVGMR IDVASIDMVS EVNMDYTLTM YFQQSWKDKR LSYNRVADQL
 gab4_chick GGNPVTVGMS IHISSIDQIS EVNMDYTITM YFQQSWRDKR LAYNRVADQL
 gab3_drome GGEPLHVGMD LTIASFDAIS EVNMDYTITM YLNQYWRDER LAFGDFAEKI
  gab_lymst GGAPLEIGIE VILASFDSIS EVDMDYTITM YLNQYWRDER LQFGAFAEKI
    gad_rat GGPPVNVALA LEVASIDHIS EANMEYTMTV FLHQSWRDSR LSYSRFVDKL
   gaa4_rat GGPVTEVKTD IYVTSFGPVS DVEMEYTMDV FFRQTWIDKR LKYNMMVTKV
   gaa6_rat GGAVTEVKTD IYVTSFGPVS DVEMEYTMDV FFRQTWTDER LKFNLMVSKI
   gaa1_rat GERVTEVKTD IFVTSFGPVS DHDMEYTIDV FFRQSWKDER LKFNLMASKI
   gaa3_rat GDAVTEVKTD IYVTSFGPVS DTDMEYTIDV FFRQTWHDER LKFNLLASKI
   gaa2_rat GDSITEVFTN IYVTSFGPVS DTDMEYTIDV FFRQKWKDER LKFNSMASKI
   gaa5_rat GERITQVRTD IYVTSFGPVS DTEMEYTIDV FFRQSWKDER LRFNLLASKI
   gac1_rat GVRPTVIETD VYVNSIGPVD PINMEYTIDI IFAQTWFDSR LKFSNMVGKI
   gac2_rat GVKPTLIHTD MYVNSIGPVN AINMEYTIDI FFAQTWYDRR LKFSNMVGKI
   gac3_rat GIKPTVIDVD IYVNSIGPVS SINMEYQIDI FFAQTWTDSR LRFSNMVGLI
 gac4_chick GIKPTFIDVD IYVNSIGPVS VIQMEYTIDI FFAQTWYDRR LRFTNMVSRI
 gac5_human GEKPTVVTVE IAVNSLGPLS ILDMEYTIDI IFSQTWYDER LCYGNVVSQL
      x78349 GGPPATIEVD IMVRSMGPIS EVDMTYSMDC YFRQSWVDKR LAFVSMLARI
   gra1_rat KGPPVNVSCN IFINSFGSIA ETTMDYRVNI FLRQQWNDPR LAYPSMLDSI
   gra3_rat KGPPVNVTCN IFINSFGSIA ETTMDYRVNI FLRQKWNDPR LAYPSMLDSI
   gra2_rat KGPPVNVTCN IFINSFGSVT ETTMDYKVNI FLRQQWNDSR LAYPSMLDSI
    grb_rat KGIPVDVVVN IFINSFGSIQ ETTMDYRVNI FLRQKWNDPR LKLPTMYKCL
    u14525  TEGAVNVRVN IMIRMLSKID VVNMEYSIQL TFREQWIDPR LAYPHVKKSL
    y09796  PMGPVTVRVN IMIRMLSKID VVNMEYSMQL TFREQWLDSR LAYPHIKSNL
    u58776  TDGPAIVRIN LFVRSIMTIS DIKMEYSVQL TFREQWTDER LKFLTEANRV
    s34469  KLEPARIQVL LYVSSIDAVN EASMDFTVGI LLHLRWTDTR IYHSENIKKV
   ach9_rat TDAVLNVTLQ VTLSQIKDMD ERNQILTAYL WIRQTWHDAY LTWRIPSDLV

                101                                               150
  gar1_rat  RFPLDTQTCS LEIESYAYTE DDLMLYWKKG NDSLKTHTTT KLAFYSSTGW
  gar2_rat  HFPLDSQTCS LELESYAYTD EDLMLYWKNG DESLKTHTTS RLAFYSSTGW
  gar3_rat  RFPLDTQNCS LELESYAYNE EDLMLYWKHG NKSLNTSASS GLAFYSSTGW
  gab2_rat  RYPLDEQNCT LEIESYGYTT DDIEFYWRGD DNAVTGKLIT KKVVF.STGS
  gab3_rat  RYPLDEQNCT LEIESYGYTT DDIEFYWRGG DKAVTGRLVS RNVVF.ATGA
  gab1_rat  RYPLDEQNCT LEIESYGYTT DDIEFYWNGG EGAVTGKMVS KKVEF.TTGA
 gab4_chick RYPLDQQNCT LEIESYGYTV DDIVFFWQGN DSAVTGRLVS REVVF.TTGS
 gab3_drome YYPLDSQNCT VEIESYGYTV SDVVMYW..K PTPVRGETND RKERL.ATGV
  gab_lymst NYPLDHQECT VEIESYGYTM DDIVLYWLND RGAVTGATIN KIEEL.STGD
    gad_rat KYPMDEQECM LDLESYGYSS EDIVYYWSEN QEQIHGRFTT ELMNFKSAGQ
   gaa4_rat DFPMDGHACP LKFGSYAYPK SEMIYTWTKG PEKSVEQTVS SETIKSITGE
   gaa6_rat NFPMDGHACP LKFGSYAYPK SEIIYTWKKG PLYSVEQTVS SETIKSNTGE
   gaa1_rat DFPMDAHACP LKFGSYAYTR AEVVYEWTRE PARSVVQTVD SGIVQSSTGE
   gaa3_rat DFPMDVHACP LKFGSYAYTK AEVIYSWTLG KNKSVEHVVG TEIIRSSTGE
   gaa2_rat DFPMDAHSCP LKFGSYAYTT SEVTYIWTYN PSDSVQQSIG KETIKSSTGE
   gaa5_rat DFPMDAHACP LKFGSYAYPN SEVVYVWTNG STKSVVQTVG TENISTSTGE
   gac1_rat NFPMDEHSCP LEFSSYGYPK NEIEYKWKKP SVEVADLRNS TEISHTISGD
   gac2_rat NFPMDEHSCP LEFSSYGYPR EEIVYQWKRS SVEVGDLRNT TEVVKTTSGD
   gac3_rat NFPMDAHACP LTFSSYGYPK EEMIYRWRKN SVEAADLRNT TEIVTTSAGD
 gac4_chick NFPMDTHSCP LVFSSYGYPR EEIVYRWRRY SIEVSDLRNT SEVLRTGAGE
 gac5_human RFPMDSHSCP LSFSSFSYPE NEMIYKWENF KLEINEVSNK TEII.TTPGD
      x78349 DFPMDIQKCP LKFGSFGYTT SDVIYRWNKE RPPVAILSGT ITLETNHPSE
   gra1_rat NFPMDVQTCI MQLESFGYTM NDLIFEWQEQ GAVQVARYCT KH..Y.NTGK
   gra3_rat NFPMDVQTCI MQLESFGYTM NDLIFEWQDE APVQVARYCT KH..Y.NTGK
   gra2_rat NFPMDVQTCT MQLESFEYTM NDLIFEWLSD GPVQVAGYCT KH..Y.NTGK
    grb_rat LFPMDTQRCK MQLESFGYTT DDLRFIWQSG DPVQLEGNCT KY..YKGTGY
    u14525  LYPLDYQSCN FDLVSYAHTM NDIMYEWDPS TPVQLKADCT SH...TNTGS
    y09796  LYPLDLQFCD FDLVSYAHTM KDIVYEWDPL APVQLKDDCT SH...TNTGS
    u58776  LYPLDRQICS LRMASYGWTT NDLVFLWKEG DPVQVVDYCN SK...TNTGE
    s34469  NYPFDKQTCH ILIMSFGYSD QDLVLDWMNL TTADDLFCNR RY..HQKAGN
   ach9_rat YPPFDSQQCN LTPGSWTYNG NQVDIFNALD SGDLSDAVKN VISYGCCSEP

                 51                                               100
  gar1_rat  WVPDMFFVHS KRSFIHDTTT DNVMLRVQPD GKVLYSLRVT VTAMCNMDFS
  gar2_rat  WVPDVFFVHS KRSFTHDTTT DNIMLRVFPD GHVLYSMRIT VTAMCNMDFS
  gar3_rat  WVPDIFFVHS KRSFIHDTTV ENIMLRVHPD GNVLFSLRIT VSAMCFMDFS
  gab2_rat  WVPDTYFLND KKSFVHGVTV KNRMIRLHPD GTVLYGLRIT TTAACMMDLR
  gab3_rat  WVPDTYFLND KKSFVHGVTV KNRMIRLHPD GTVLYGLRIT TTAACMMDLR
  gab1_rat  WVPDTYFLND KKSFVHGVTV KNRMIRLHPD GTVLYGLRIT TTAACMMDLR
 gab4_chick WLPDTYFLND KKSFLHGVTV KNRMIRLHPD GTVLYGLRIT TTAACMMDLR
 gab3_drome WVPDTFFAND KNSFLHDVTE RNKLVRLGGD GAVTYGMRFT TTLACMMDLH
  gab_lymst WVPDTFLAND KNSFLHDITE KNKMVRLYGN GSLVYGMRFT TTLACMMDLH
    gad_rat WLPDTFIVNA KSAWFHDVTV ENKLIRLQPD GVILYSIRIT STVACDMDLA
   gaa4_rat WTPDTFFRNG KKSVSHNMTA PNKLFRIMRN GTILYTMRLT ISAECPMRLV
   gaa6_rat WTPDTFFRNG KKSIAHNMTT PNKLFRLMHN GTILYTMRLT INADCPMRLV
   gaa1_rat WTPDTFFHNG KKSVAHNMTM PNKLLRITED GTLLYTMRLT VRAECPMHLE
   gaa3_rat WTPDTFFHNG KKSVAHNMTT PNKLLRLVDN GTLLYTMRLT IHAECPMHLE
   gaa2_rat WTPDTFFHNG KKSVAHNMTM PNKLLRIQDD GTLLYTMRLT VQAECPMHLE
   gaa5_rat WTPDTFFHNG KKSIAHNMTT PNKLLRLEDD GTLLYTMRLT ISAECPMQLE
   gac1_rat WIPDTFFRNS RKSDAHWITT PNRLLRIWSD GRVLYTLRLT INAECYLQLH
   gac2_rat WIPDTFFRNS KKADAHWITT PNRMLRIWND GRVLYTLRLT IDAECQLQLH
   gac3_rat WIPDTIFRNS KTAEAHWITT PNQLLRIWND GKILYTLRLT INAECQLQLH
 gac4_chick WIPDTFFRNS KRADSHWITT PNQLLRIWND GKVLYTLRLT IEAECLLQLQ
 gac5_human WIPDTFFRNS KRTHEHEITM PNQMVRIYKD GKVLYTIRMT IDAGCSLHML
      x78349 WKPDTYFYNG KQSYLHTITT PNKFVRIYQN GRVLYSSRLT IKAGCPMNLA
   gra1_rat WKPDLFFANE KGAHFHEITT DNKLLRISRN GNVLYSIRIT LTLACPMDLK
   gra3_rat WKPDLFFANE KGANFHEVTT DNKLLRIFKN GNVLYSIRLT LTLSCPMDLK
   gra2_rat WKPDLFFANE KGANFHDVTT DNKLLRISKN GKVLYSIRLT LTLSCPMDLK
    grb_rat WKPDLFFANE KSANFHDVTQ ENILLFIFRD GDVLVSMRLS ITLSCPLDLT
    u14525  WIPDTFFPTE KAAHRHLIDM ENMFLRIYPD GKILYSSRIS LTSSCPMRLQ
    y09796  WIPDTFFPTE KAAHRHLIDT DNMFLRIHPD GKVLYSSRIS ITSSCHMQLQ
    u58776  WMPDLFFSNE KEGHFHNIIM PNVYIRIFPN GSVLYSIRIS LTLACPMNLK
    s34469  WVPDIFFPNE KKGSFHDIMT QNQMMRLYQG GTILYISRLS MTLSCPMDLI
   ach9_rat WRPDIVLYNK ADDESSEPVN TNVVLRY..D GLITWDSPAI TKSSCVVDVT

                151                                               200
  gar1_rat  YNRLYINFTL RRHIFFFLLQ TYFPATLMVM LSWVSFWIDR RAVPARVPLG
  gar2_rat  YNRLYINFTL RRHIFFFLLQ TYFPATLMVM LSWVSFWIDH RAVPARVSLG
  gar3_rat  YYRLFINFVL RRHIFFFVLQ TYFPAMLMVM LSWVSFWIDR RAVPARVSLG
  gab2_rat  YPRLSLSFRL KRNIGYFILQ TYMPSILITI LSWVSFWINY DASAARVALG
  gab3_rat  YPRLSLSFRL KRNIGYFILQ TYMPSLLTTI LSWVSFWINY DASAARVALG
  gab1_rat  YPRLSLSFRL KRNIGYFILQ TYMPSTLITI LSWVSFWINY DASAARVALG
 gab4_chick YLRLSLSFRI KRNIGYFILQ TYMPSILITI LSWVSFWINY DASAARVALG
 gab3_drome YQRLSLSFKL QRNIGYFVFQ TYLPSILIVM LSWVSFWINH EATSARVALG
  gab_lymst YQRLSLIFQL QRNIGYFIFQ TYLPSILIVM LSWVSFWINH EATSARVALG
    gad_rat FPRLSLHFQL RRNRGVYIIQ SYMPSVLLVA MSWVSFWISQ AAVPARVSLG
   gaa4_rat YIVMTVYFHL RRKMGYFMIQ TYIPCIMTVI LSQVSFWINK ESVPARTVFG
   gaa6_rat YVIMTVYFHL QRKMGYFMIQ IYTPCIMTVI LSQVSFWINK ESVPARTVFG
   gaa1_rat YVVMTTHFHL KRKIGYFVIQ TYLPCIMTVI LSQVSFWLNR ESVPARTVFG
   gaa3_rat YVVMTTHFHL KRKIGYFVIQ TYLPCIMTVI LSQVSFWLNR ESVPARTVFG
   gaa2_rat YTVMTAHFHL KRKIGYFVIQ TYLPCIMTVI LSQVSFWLNR ESVPARTVFG
   gaa5_rat YTIMTAHFHL KRKIGYFVIQ TYLPCIMTVI LSQVSFWLNR ESVPARTVFG
   gac1_rat YIIMTIFFDL SRRMGYFTIQ TYIPCILTVV LSWVSFWINK DAVPARTSLG
   gac2_rat YVVMSVYFDL SRRMGYFTIQ TYIPCTLIVV LSWVSFWINK DAVPARTSLG
   gac3_rat YVVMTIYFEL SRRMGYFTIQ TYIPCILTVV LSWVSFWIKK DATPARTTLG
 gac4_chick YMVMTVSFDL SRRMGYFAIQ TYIPCILTVV LSWVSFWIKR DSTPARTSLG
 gac5_human FMVMTIFFNV SRRFGYVAFQ NYVPSSVTTM LSWVSFWIKT ESAPARTSLG
      x78349 YSMLMVNFHL QRHMGNFLIQ VYGPCCLLVV LSWVSFWLNR EATADRVSLG
   gra1_rat FTCIEARFHL ERQMGYYLIQ MYIPSLLIVI LSWISFWINM DAAPARVGLG
   gra3_rat FTCIEVRFHL ERQMGYYLIQ MYIPSLLIVI LSWVSFWINM DAAPARVALG
   gra2_rat FTCIEVKFHL ERQMGYYLIQ MYIPSLLIVI LSWVSFWINM DAAPARVALG
    grb_rat YTCVEVIFTL RRQVGFYMMG VYAPTLLIVV LSWLSFWINP DASAARVPLG
    u14525  YGCLRMQLLF KRQFSYYLVQ LYAPTTMIVI VSWVSFWIDL HSTAGRVALG
    y09796  YACLRMQLTL KRQFSYYLVQ LYGPTTMIVI VSWVSFWIDM HSTAGRVALG
    u58776  YSCLKVDLLF RREFSYYLIQ IYIPCCMLVI VSWVSFWLDQ GAVPARVSLG
    s34469  HSCIQAEFHL ARNIGFYIVQ MYIPSMLIVM LSWISFWLTV NSVPGRVSLG
   ach9_rat YPDVTFTLLL KRRSSFYIVN LLIPCVLISF LAPLSFYL.P AASGEKVSLG
```

```
            201                                                 250
  gar1_rat  ITTVLTMSTI ITGVNASMPR VSYIKAVDIY LWVSFVFVFL SV.LEYAAVN
  gar2_rat  IMTVLTMSTI ITGVNASMPR VSYIRAVDIY LWVSFVFVFL SV.LEYAAVN
  gar3_rat  ITTVLTMSTI VTGVSASMPQ VSYVKAVDVY MWVSSLFVFL SV.IEYAAVN
  gab2_rat  ITTVLTMTTI NTHLRETLPK IPYVKAIDMY LMGCFVFVFM AL.LEYALVN
  gab3_rat  ITTVLTMTTI NTHLRETLPK IPYVKAIDMY LMGCFVFVFL AL.LEYAFVN
  gab1_rat  ITTVLTMTTI STHLRETLPK IPYVKAIDIY LMGCFVFVFL AL.LEYAFVN
gab4_chick  VTTVLTMTTI NTHLRETLPK IPYVKAIDVY LMGCFVFVFL AL.LEYAFVN
gab3_drome  ITTVLTMTTI STGVRSSLPR ISYVKAIDIY LVMCFVFVFA AL.LEYAAVN
 gab_lymst  ITTVLTMTTI SNGVRSSLPR ISYVKAIDIY LVMCFVFVFA AL.LEYAAVN
   gad_rat  ITTVLTMTTL MVSARSSLPR ASAIKALDVY FWICYVFVFA AL.VEYAFAH
  gaa4_rat  ITTVLTMTTL SISARHSLPK VSYATAMDWF IAVCFAFVFS AL.IEFAAVN
  gaa6_rat  ITTVLTMTTL SISARHSLPK VSYATAMDWF IAVCFAFVFS AL.IEFAAVN
  gaa1_rat  VTTVLTMTTL SISARNSLPK VAYATAMDWF IAVCYAFVFS AL.IEFATVN
  gaa3_rat  VTTVLTMTTL SISARNSLPK VAYATAMDWF MAVCYAFVFS AL.IEFATVN
  gaa2_rat  VTTVLTMTTL SISARNSLPK VAYATAMDWF IAVCYAFVFS AL.IEFATVN
  gaa5_rat  VTTVLTMTTL SISARNSLPK VAYATAMDWF IAVCYAFVFS AL.IEFATVN
  gac1_rat  ITTVLTMTTL STIARKSLPK VSYVTAMDLF VSVCFIFVFA AL.MEYGTLH
  gac2_rat  ITTVLTMTTL STIARKSLPK VSYVTAMDLF VSVCFIFVFS AL.VEYGTLH
  gac3_rat  ITTVLTMTTL STIARKSLPR VSYVTAMDLF VTVCFLFVFA AL.MEYATLN
gac4_chick  ITTVLTMTTL STISRKHLPR VSYITAMDLF VSVCFIFVFA AL.MEYATLN
gac5_human  ITSVLTMTTL GTFSRKNFPR VSYITALDFY IAICFVFCFC AL.LEFAVLN
    x78349  ITTVLTMTFL GLEARTDLPK VSYPTALDFF VFLSFGFIFA TI.LQFAVVH
  gra1_rat  ITTVLTMTTQ SSGSRASLPK VSYVKAIDIW MAVCLLFVFS AL.LEYAAVN
  gra3_rat  ITTVLTMTTQ SSGSRASLPK VSYVKAIDIW MAVCLLFVPS AL.LEYAAVN
  gra2_rat  ITTVLTMTTQ SSGSRASLPK VSYVKAIDIW MAVCLLFVFA AL.LEYAAVN
   grb_rat  IFSVLSLASE CTTLAAELPK VSYVKALDVW LIACLLFGFA SL.VEYAVVQ
    u14525  VTTLLTMTTM QSAINAKLPP VSYVKVVDVW LGACQTFVFG AL.LEYAFVS
    y09796  VTTLLTMTTM QAAINAKLPP VSYVKVVDVW LGACQTFVFG AL.LEYAFVS
    u58776  VTTLLTMATQ TSGINASLPP VSYTKAIDVW TGVCLTFVFG AL.LEFALVN
    s34469  LLTVLTMTTQ SSSVNAALPR VSYTKAIDVW MSTCLVFVFA AL.LEFAVVN
  ach9_rat  VTILLAMTVF QLMVAEIMPA SENVPLIGKY YIATMALITA STALTIMVMN

            251                                                 280
  gar1_rat  YNTHAIDKYS RIIFPAAYIL FNLIYWSIFS
  gar2_rat  YNTHAIDKYS RLIFPAFYIV FNLIYWSVFS
  gar3_rat  YNNHVIDTYS RIVFPVVYII FNLFYWGIYV
  gab2_rat  YDVNAIDRWS RIFFPVVFSF FNIVYWLYYV
  gab3_rat  YDVNAIDRWS RIVFPFTFSL FNLVYWLYYV
  gab1_rat  YDVNSIDKWS RMFFPITFSL FNVVYWLYYV
gab4_chick  YDVSTIDKWS RIIFPITFGF FNLVYWLYYV
gab3_drome  YDVNIIDKYS RMIFPISFLA FNLGYWLFYI
 gab_lymst  YDVNTIDKYA RLMPPLLFII FNTSYWSVYL
   gad_rat  FDADTIDIYA RAVFPAAFAA VNIIYWAAYT
  gaa4_rat  YGTSKIDKYA RILFPVTFGA FNMVYWVVYL
  gaa6_rat  YGTSKIDQYS RILFPVAFAG FNLVYWIVYL
  gaa1_rat  YSVSKIDRLS RIAFPLLFGI FNLVYWATYL
  gaa3_rat  YSVSKVDKIS RIIFPVLFAI FNLVYWATYV
  gaa2_rat  YSVSKIDRMS RIVFPVLFGT FNLVYWATYL
  gaa5_rat  YSISKIDKMS RIVFPILFGT FNLVYWATYL
  gac1_rat  YRIAKIDSYS RIFFPTAFAL FNLVYWVGYL
  gac2_rat  YRIAKMDSYA RIFFPTAFCL FNLVYWVSYL
  gac3_rat  YDVSELDSYS RVFFPTSFLL FNLVYWVGYL
gac4_chick  YHISRLDSYS RVFFPTAFLL FNIVYWIAYL
gac5_human  FHVYRLDNYS RVVFPVTFFF FNVLYWLVCL
    x78349  YSVSKIDRAS RIVFPLLFIL INVFYWYGYL
  gra1_rat  FRAKKIDKIS RIGFPMAFLI FNMFYWIIYK
  gra3_rat  FRAKKIDTIS RACFPLAFLI FNIFYWVIYK
  gra2_rat  FRAKRIDTIS RAAFPLAFLI FNIFYWITYK
   grb_rat  VAAKRIDLYA RALFPFCFLF FNVIYWSIYL
    u14525  YLPAKIDFYA RFVVPLAFLA FNVIYWVSCL
    y09796  YLPAKIDYYA RFCVPLGFLA FNAIYWTSCL
    u58776  YRSKRIDVIS RITFPLVFAL FNLVYWSTYL
    s34469  VYAIYVDMTA RVVFPICFII FIMSYWLYYV
  ach9_rat  IKGSEWKKVA KVIDRFFMWI FFAMVFVMTV
```