Analysis of the Set of GABA$_A$ Receptor Genes in the Human Genome*

Joseph Simon‡§, Hironobu Wakimoto§§, Norihisa Fujita¶, Marc Lalande, and Eric A. Barnard‡**

From the ‡Department of Pharmacology, University of Cambridge, Cambridge CB2 1PD, United Kingdom, the §Department of Bioscience and Bioinformatics, Ritsumeikan University, Noji-higashi, Kusatsu 525-8577, Japan, and the ¶Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, Connecticut 06030

The genes of the ionotropic $\gamma$-aminobutyric acid receptor (GABR) subunits have shown an unusual chromosomal clustering, but only now can this be fully specified by analyses of the human genome. We have characterized the genes encoding the 18 known human GABR subunits, plus one now located here, for their precise locations, sizes, and exon/intron structures. Clusters of 17 of the 19, distributed between five chromosomes, are specified in detail, and their possible significance is considered. By applying search algorithms designed to recognize sequences of all known GABR-type subunits in species from man down to nematodes, we found no new GABR subunit is detectable in the human genome. However, the sequence of the human orthologue of the rat GABR $\beta 3$ receptor subunit was uncovered by these algorithms, and its gene can be analyzed. Consistent with those search results, orthologues of the $\beta 4$ and $\gamma 4$ subunits from the chicken, not cloned from mammals, were not detectable in the human genome by specific searches for them. The relationships are consistent with the mammalian $\theta$ subunit being derived from the $\beta$ line and $\epsilon$ from the $\gamma$ line, with mammalian loss of $\beta 4$ and $\gamma 4$. In their structures the human GABR genes show a basic pattern of nine coding exons, with six different genomic mechanisms for the alternative splicing found in various subunits. Additional noncoding exons occur for certain subunits, which can be regulatory. A dicysteine loop and its exon show remarkable constancy among all GABR subunits and species, of deduced functional significance.

The GABA$_A$ receptors are GABA-gated ion channels that are well known to mediate most of the inhibitory signaling in the vertebrate central nervous system. They also operate at a number of peripheral sites. These receptors have been linked to a variety of human pathologies and developmental abnormalities (1, 2) and to human cognitive pathways (3), and they are the targets of a significant overall percentage of current pharmacological usage and design (4, 5). For the pursuit of all of these aspects of this receptor activity, a full knowledge of the GABA$_A$ receptor genes in the human genome, knowledge which hitherto has been only partial, will clearly be an important aid.

The GABA$_A$ receptors are formed as combinations of homologous subunits (6); these are from 420 to 632 amino acids in length, of which 19 (Fig. 1) have been cloned from mammals but not all from man (reviewed in Refs. 7 and 8), each clearly being attributed to a different gene. This repertoire of subunits is drawn upon in a restricted combinatorial system (7, 8) to assemble as pentamers (9), which are the active receptors. The set of 19 GABA$_A$ receptor (GABR) subunits is the largest of any among the mammalian ion channel receptors, and the arrangement in the genome of their genes, as will be discussed, is unusual and can itself be of intrinsic significance. The mapping of these genes began in 1989 with the assignment of $\alpha 1$ to chromosome 5q34 and $\alpha 3$ to X q28, and most significantly, $\alpha 2$ and $\beta 1$ were found together on 4p12 (10, 11). The distribution of 14 of these genes in clusters on human chromosomes, as known from further cytogenetic, linkage, and genomic cloning studies, was reviewed in 1999 (12, 13). Four GABR genes were known to be close together on chromosome 4, four on chromosome 5, three on chromosome 15, and three on X. The $\pi$ subunit gene was reported to occur separately on chromosome 5q. The order of the genes within the clusters was not always univocally known by the methods previously available.

From analysis of the human genome data base and with some experimental studies based on it, we can now address a number of issues that have been outstanding, as follows. 1) Are there further subunits of human GABA$_A$ receptors that, due to a greater divergence, have gone unrecognized? 2) Not every one of the genes for the 19 subunits as known from rodents and man combined had been previously found and located in the human genome. Furthermore, it is only now that the detailed locations of all of the human GABA$_A$ receptor (GABR) genes can be fully specified, together with the precise arrangement within each of five gene clusters that are present. 3) The cluster of three GABR genes on chromosome 15q, in particular, is incompletely known, due to residual refractory gaps there in the genome mapping and other ambiguities, so further investigation of this cluster was needed. 4) Certain functional GABA$_A$ receptor subunits described from nonmammalian vertebrates do not appear to correspond to any known mammalian, notably chicken $\beta 4$ and $\gamma 4$. Similarly, from invertebrates a homologous functional GABA-gated cation channel is known. Can human orthologues of those nonmammalian types now be found? 5) Are there common features of the exon/intron organization across the entire family of GABR genes? 6) The genomic origin of alternatively spliced GABR mRNAs was not

* This work was supported by the Wellcome Trust. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Both authors contributed equally to this work.

** To whom correspondence should be addressed: Dept. of Pharmacology, University of Cambridge, Tennis Court Rd., Cambridge CB2 1PD, UK. Tel.: 44-1223-347876; Fax: 44-1223-334040; E-mail: eb247@cam.ac.uk.

§ The abbreviations used are: GABA$_A$, $\gamma$-aminobutyric acid, type A; GABA, $\gamma$-aminobutyric acid; EST, expressed sequence tag; GABR, ionotropic $\gamma$-aminobutyric acid receptor; GLR, glycine receptor; nAChR, nicotinic acetylcholine receptor; 5-HT$_3$, 5-hydroxytryptamine type 3 receptor; utr, untranslated region; ZAC, zinc activated cation channel.

41422 This paper is available on line at http://www.jbc.org
often defined previously. A surprising variety of genomic mechanisms for creating multiple forms of the GABR subunits can be specified.

EXPERIMENTAL PROCEDURES

Consensus Sequence Searches—The amino acid sequences in and immediately neighboring the Cys loop (as defined under “Results”) were aligned for each receptor family as specified. For the vertebrate GABR family, 87 GABR subunit full-length independent sequences from vertebrate species were used in determining a consensus Cys loop motif for that set (form A, as defined under “Results”). Later the total of those sequences, including others from the nonmammalian vertebrates that could be retrieved and verified from the latest EMBL and treEMBL data bases, rose to 100, and it was found that all 100 matched the form A consensus. Taking positions conserved in all of the GABR sequences, a GABR Cys loop consensus was written in nucleotide form, including all codon redundancies and with bidirecionality. This consensus was used as a position-specific profile to write an algorithm for a search program (available on request), as described under “Results,” designed to be run against the human or any other genome data base. For each of the other three Cys loop receptor families considered, all the human sequences were aligned to obtain three family-specific forms of the consensus. In another case where invertebrate GABR-like sequences were also considered, those were aligned in protein form to identify the consensus around the Cys loop region. In all of the GABR sequences, vertebrate and invertebrate, this agreed completely with a second minimal form of the consensus (form E, as defined under “Results”); form E was also applied to the human genome search. This type of search program registers as hits only cases where all of the specified positions match and does not require probability judgements. The hits in the GABR search were identified in their contig locations if on known subunits and where on an unknown sequence this was identified or characterized by a BLASTN search of its contig.

Sequence Alignments and Other Search Methods—Full DNA and protein sequences were aligned and analyzed by using the Vector NTI Suite 8.0 sequence analysis package (Invitrogen). For the dendrogram construction, an algorithm based on a Neighbor-joining method was used, with evaluation by bootstrapping to 100,000 replications, plotted (unrooted) by the NJ-Plot program. A parallel tree, constructed instead by a Maximum Parsimony method with 10,000 bootstrap replications, gave a topology that overall was quite similar, but this showed low bootstrap scores (53 and 60%) at two of the final branch points and was not used. For the search applications, where not specified, on-line software packages, FASTA, WU-BLAST 2.0, NCBI-BLAST 2.0 or ClustalW, from EBI, NCBI, or the Wellcome Trust Sanger Institute (Hinxton, UK), were used. The significance of hits in BLAST or FASTA searches was assessed by their p values (as defined in the BLAST program), with a cut-off at e⁻¹⁰. Signal peptide cleavage sites were taken from the original publications where predicted or, if not, from predictions in data bases where given; otherwise they were assigned by standard methods so that the mature protein sequences were always used. For prediction of exon/intron boundaries the following splice site consensus sequences were used: for the most commonly occurring GT-AG type introns, GT/AG at the 5’-splice site and (C/T)AG at the 3’-splice site; for the minority AT-AC type introns, ATATCT(T/C) at the 5’-splice site and (C/T)AC at the 3’-splice site (14). For the sequences used from invertebrates, for *Drosophila* sp. or *Caenorhabditis elegans*, searches were made in the Flybase or Wormbase data bases, respectively, for reported anion-channel receptor-like sequences, and those showing overall clear homology to vertebrate GABR subunits were selected. For *Anopheles gambiae* the EBI completed genome in Ensemble was used similarly. For other invertebrates the Pasteur Institute LGIC classification (15) was used, and this was much extended by searches made with the same criteria in the Uniprot/TreEMBL data base.

Exon Analyses—The subunit cDNA sequences retrieved from EMBL data base were run against the human genome or against relevant contigs to identify exons, except in cases where there was sufficient independent published evidence, as cited here, to confirm a genome browser automated analysis of exons. Where cDNA information was incomplete, the Grail Express exon prediction program (originally from Oak Ridge National Laboratories) within the Vector NTI package was applied for exon scanning, combined with splice site analysis.

FIG. 1. Dendrogram showing the sequence relationships of 19 human GABA<sub>A</sub> receptor subunits. Sequences used are those of the mature protein after signal peptide removal. A distantly related outgroup marker, the nAChR δ subunit (as shown), was used for refinement of the plot. The length of horizontal branches connecting any two GABA<sub>A</sub> receptor subunits represents the fractional divergence in their amino acid sequence. The scale bar corresponds to 20% sequence divergence. Numbers at the junctions are bootstrap values for 100,000 replications.

RESULTS AND DISCUSSION

The genes encoding 19 subunits already known from their cloned and expressed cDNAs, as reported either from man, or from rat or mouse (where previously not identified in man), were located in the human genome, as described below. The full-length human protein sequences verified there were used to construct a dendrogram of them all (Fig. 1). Only the mature polypeptide sequences (see “Experimental Procedures”) were used in the multiple alignment. In this plot the ε, θ, π, and δ subunits are relatively isolated, but ε appears as related more to γ subunits than others, θ as more related to β subunits, whereas δ and π each form a separate single subunit subgroup (Fig. 1). The six questions raised in the Introduction were considered in relation to this family of genes.

Are There Previously Undetected Subunits of GABA<sub>A</sub> Receptors Encoded in the Human Genome?

Conventional nucleotide/protein BLAST searches of the genome with known GABR cDNA sequences gave no significant hit other than on known GABR subunit or glycine receptor α subunit (GLRA) sequences (data not shown). As the query here the human α1, α4, θ, and π and the chicken β4 sequences were
receptor sequences was constructed, initially using the amino acids of the consensus A of Fig. 2. These were then used in nucleotide sequence, with full inclusion of codon redundancy. Recorded hits in this system will be clear total matches, not probabilities. The present form of this probe depends upon the consensus lying within one exon. That situation was confirmed to hold in analyses (to be described below) of the genomic structures of all human GABR subunit genes and in similar examinations of the GABR genes from analyzed nonmammalian vertebrates.

A genome search made with consensus A gave no match at any unknown sequence, although it retrieved all of the predicted human GABR genes. Therefore, we addressed the question whether some unidentified human GABR subunit sequence might exhibit a change in the Cys loop region that is so aberrant that it has not occurred in any of the 100 known vertebrate cases. In the general case, six amino acid positions is the minimum number that we found could be specified in such a position-specific algorithm without producing a prohibitive level of random hits in a genome search. We sought a 6-point consensus that would cover, in addition, known invertebrate GABR sequences, where greater variation occurs and where any other ancestral GABR line might be revealed. In the invertebrates the pharmacological distinctions seen between the GABR subunits in vertebrates have been found not to be generally applicable, and the distinction between GABA and glycine receptors can become blurred in some cases (23). An alignment was then made of the known invertebrate GABR-like proteins using the available full-length subunit sequences of the anion channel type that have a discernible overall homology to human GABR subunits. For most of those, expression and GABA responses have been reported, but some have not been tested, e.g. those in the recently released full genome (EBI data base) of the A. gambiae mosquito. For the many putative GABR-related subunit sequences (−40) encoded in the completed C. elegans genome caution is needed, because evidence on their activity in assembled GABR receptors is as yet mostly lacking and the exon organization, where deduced, can differ considerably from that of known GABR genes in other phyla; some there could not yet be classified. For all C. elegans sequences that could be assigned as possible GABR-related subunits, as well as for 29 other invertebrate cases (from molluscs, crustacea, insects, arachnids, and parasitic worms), a GABR Cys loop region 6-position consensus E (Fig. 2) was found to be completely conserved.

Consensus E thus makes the smallest number of assumptions in this series and was used in the final screen. It is comprehensive, so it will also see other members of the superfamily (Fig. 2), but their features elsewhere in the sequence will be readily recognized. This search yielded 118 hits on sequences present in the human genome. 41 were on known Cys loop subunits as follows: 17 on GABR, 14 on nAChR (discussed below), 5 on GlyR, and 4 on 5-HT3R subtypes. The 41st match led to the human GABR β3 subunit gene, previously not characterized and now detailed below. A 19th known GABR subunit, γ3, gave no hit here because the region containing the γ3 Cys loop is still incomplete in the current human genome data base, as explained below. 49 of the other hits could be discarded, due to multiple stop codons in the apparent Cys loop sequence or to a clear location within an intron of a known gene or to the consensus nucleotides occurring in simple tandem repeats.

This left 28 possible candidates. All of these were located very far from any known human GABR gene. Each was analyzed. By using the most general consensus sequences for splicing sites (see “Experimental Procedures”), where exons were
predicted none was at (or even close to) the 83 bp in length which was always found in known GABR Cys loop exons, as analyzed below. In the hit regions, the predicted exon was either much shorter or longer, ranging up to 109 bp. Five of the candidates could be discarded because they again contained a stop codon within the consensus Cys loop sequence or had a repeat of the consensus in tandem, with no separating exon boundary. In a sixth case, the orientation of the hit sequence was opposite that of the coding DNA at that point. For the rest, despite their varying exon length, a detailed analysis was nevertheless made by the Grail Express genomic structure prediction program. This showed that the candidate again either lies within an annotated known gene, always from a very different family and always in a known intron therein or spanning an intron boundary, or it lies in a predicted intron of an unidentified gene, which in any case has a total coding length much too small for a GABR subunit. All of those 28 hits were chance nonspecific alignments. Where the hit was on an intron large enough to contain an entire GABR gene, the Grail Express analysis was further applied separately on that intron to show that no smaller genes were enclosed and to confirm that no exon with the consensus Cys loop sequence was present. In sum, every one of the 118 hits from the minimal consensus sequence E could be accounted for, and no previously unidentified human GABR subunit gene could be detected in the genome.

As noted above, the general GABR consensus sequence E also recognized 14 human nAChR subunits. 16 nAChR subunits are currently recognized in mammals: α1–7, α9–10, β1–4, γ, δ, and ε, with an α8 sequence known so far from chicken but not from man (for sequences see LGIC data base (15)). In all vertebrate β1 nAChR genes known, exceptionally, the Cys loop lies in 2 exons. The consensus B of Fig. 2 matches the protein sequence of all 16 known human nAChRs (plus the avian α8); when that was used in the genome search it retrieved all known human nAChRs, except β1, but (apart from two pseudogenes) not an α8 and no other gene. We confirmed the absence in the human genome of a sequence like the chicken nAChR α8 by a BLAST search with that α8 sequence, with negative results. For comparison, in the pufferfish (Fugu rubripes), 28 expressed nAChR subunit sequences (24) were recently found (all of which would be found by consensus b of Fig. 2). Like the high multiplicity of nAChR-like sequences and GABR-like sequences seen in the C. elegans genome, this suggests that diversification in the nAChR and GABR series is notably less in mammals than can occur in some lower species. Whether the known mammal-specific interspersed elements (e.g. LINE-1) could operate to restrict such diversity merits investigation.

**The GABR Form of the Cys Loop Is Present in Invertebrate GABA-gated Cation and Glutamate-gated Anion Channels**

A GABA-gated cation channel, EXP-1, has been identified recently from C. elegans, which is homologous to the vertebrate GABR subunits and not to vertebrate AChR or glutamate receptors (25). EXP-1 has, however, a cation-selective pore segment. Evidence that similar responses occur in certain rat brain neurons was cited (25). The Cys loop in EXP-1 still matches our GABR consensus E. It even matches the vertebrate GABR Cys loop fuller consensus A, apart from the initial Thr being Ser, a common minor variation in C. elegans, but not found in Cys loop consensus forms for the other cation channel receptor families, nAChR, 5-HT3R, and ZAC. Hence, we searched by BLAST with the EXP-1 coding sequence the full human genome and also the newly completed mouse genome.

All significant hits not found on known Cys loop genes were analyzed by the Grail Express program and shown to be from accidental similarities in noncoding sequences. Hence, GABA-gated cation channels cannot be found in these mammals.

Conversely, glutamate-gated anion channels are known from C. elegans, from parasitic nematodes, and from Drosophila (26–28). That subfamily also maintains overall homology to vertebrate GABA<sub>a</sub> receptors and is structurally totally different from the family of glutamate-gated cation channels (26). Those invertebrate glutamate-gated anion channel sequences also possess the GABR-type Cys loop consensus E. From these comparisons and those noted above for invertebrate GABR-like receptors of varied agonist specificities, we concluded that this specific form of a Cys loop framework is diagnostic for the basic GABA<sub>a</sub> receptor structure and independent of the variations in the binding site or in the pore structure.

**Locations and Clusters of All GABA<sub>a</sub> Receptor Genes**

**Chromosome 4 Cluster**—This contains two α, one β, and one γ genes, confirming results from earlier methods (29). We can now specify their relationships (Fig. 3A). The order of these along the chromosome is the opposite of that deduced by Russek (13) but is as mapped by Bailey et al. (12). The differing orientations of transcription of the four genes, however, confirm those deduced by Russek (13) on genomic DNA clones. No genes of other types lying between the γ1 and α2 genes can be detected, but the wide region that separates the α2 and α4 genes contains the cytochrome oxidase 7B2 gene and at least one other unidentified (but non-GABR) gene. The gap between the α4 and β1 genes is the smallest here and contains no other gene.

**Chromosome 5 Cluster**—The two α, one β, and one γ genes here (13, 30) are organized as shown in Fig. 3B. Their locations and separations are now corrected. Only one non-GABR gene (a member of the glutaredoxin family) has been detected so far in this cluster, between α6 and α1.

**Revised Specification of the Chromosome 15 Cluster**—The start of this cluster lies in band q13.2; it known to contain (16) an α, a β, and a γ gene (Fig. 4A). The first is the GABRB3 gene. This had been found previously to contain 10 exons (17), of which the first two are alternatively expressed (31). These 10 exons were confirmed across three adjacent contigs in the genome path. The nine introns range in size from 100 to 175,000 bp.

The region of this cluster unfortunately contains two of the rare persistent gaps in the genome tile path, presently described as “regions whose sequence cannot be reliably resolved with current technology” (e.g. UCSC Genome Gateway, latest release, genome.ucsc.edu). Due to gaps, the exon-intron structure is not yet fully covered in the genome data base for the α5 and γ3 GABR genes. This causes important omissions or errors, even well beyond the gaps themselves, within the annotations of those two genes in the genome data bases. Therefore, we have used independent clones of genomic DNA segments obtained from this region, together with the known α5 and γ3 cDNA sequences, to map in more detail than before the exon-intron structures of GABRA5 and GABRG3 and the gene separations in this cluster (Fig. 4). The GABRA5 gene was found previously to have four exons in its 5′-utr region (32). To determine the exact location of these genes along chromosome 15 and to check for any others, we used selected phase clones of human chromosome 15 DNA (see “Experimental Procedures”). We found that some sequences from these phase clones aligned, with 95–100% matches, within two new larger contigs recently deposited by the Whitehead-MIT Genome Sequencing Group into the GenBank<sup>TM</sup> and Ensembl genome data bases.
To place the fosmid AC145196 (NT_079554) on the map, we aligned with the AC135999 cosmid (already mapped) and found an overlap (24.1 kbp) as represented in Fig. 4B. Alignment of the /H9251 cDNA sequence with our probe sequence b showed that b contains the first five exons of the GABRA5 gene. Probe b lies within the AC145196 fosmid sequence, and exons 6–8 of /H9251 were also found here (Fig. 4B). We located a remaining gap (27.7 kbp) that interrupts the present genome tile path, but we found that it lies wholly within intron 8. Intronic probe a, derived like the other probes from GABRA5 genomic DNA, was found to be within intron 8; probe a overhangs at its 3’-end by 892 bp the 5’-end of AC145196 and extends into the gap for 1934 bp, reducing that gap to 25.8 kbp. Sequences of two other clones (not marked in Fig. 4) which should from their origin be close to probe a gave no match at all in the genome, so they may be from intron 8 but within the gap. Probes c–f were shown to align in cosmid AC135999 (Fig. 4B). Probe e (our λ329 sequence) was shown also to overlap by 2.4 kbp with the next contig, AC145196. Probe e was used to locate the 5’-end of the GABRA5 gene and to exclude any further 5’-utr noncoding exons there. Probes d–f were located outside the GABRA5 gene. By combined use of those clones and alignments of the exon sequences derived earlier from genomic DNA clones (17, 32), we could locate precisely all of the exons of GABRA5 in the genome (Fig. 4C). This confirmed the number of the exons as 13. The sizes of the exons and introns are exact or close matches to the estimates in the previous results (17, 32), except that we can now specify the lengths of the noncoding exons 1–3 (each previously given as a minimum) and of intron 8. The translation starts for the GABRA5 protein at nucleotide 74 of exon 5 (24,661,758 bp on the chromosome). The transcriptional start of the GABRA5 gene (24,659,419 bp on the chromosome) has now been located at 93,260 bp downstream from the GABRB3 gene; the size found for the GABRA5 transcription unit was much increased over earlier estimates to 97,355 bp.

The γ3 gene was previously shown, using standard cloning methods for exonic sequences, to contain 10 exons (GenBank™ accession number AF228453). However, their arrangement in the genome within 15q13 was not determined until now, due to the absence (as noted above) of exon 5 and most of introns 4 and 5 in the present tile path of the genome and the consequent perturbation of the annotations for the rest of the gene. We found three contigs (Fig. 4A) covering this region (apart from the gap). By using BLASTN to interrogate them with the γ3 cDNA sequence, we could identify nine exons of that gene, all within those contigs. Exon 5, which includes the Cys loop, was not retrieved (because it lies in the present gap) but only inferred from the cDNA sequence. The translated sequence resumes from exon 6 to its end, and the gene is deduced to have 10 coding exons (Fig. 4D). Exon 1 encodes only the signal peptide plus 16 bp of 5’-utr; a preceding noncoding 11th exon has not been excluded.

The X Chromosome Cluster—This cluster is located in band Xq28 and comprises the genes of ε (GABRE), α3, and θ (GABRQ) subunits. That band location is as found for GABRE (33), for α3 (10, 11), and for θ (34, 35). We have derived the precise locations, separations, and sizes of the genes there (Fig. 3E). There was disagreement previously on the order of these three genes relative to the centromere (cf. Fig. 5 of Ref. 13 with...
Fig. 6 of Ref. 33); we found the order in the genome (Fig. 3) that predicted (33) from PCR mapping on YACs and cosmids containing X chromosome markers. The two intergenic distances are 2- and 2.8-fold lower than those estimated by the earlier methods. Nevertheless, these three genes are separated by four or more unrelated genes.

Fig. 4. GABR gene structures in the cluster on chromosome 15. Annotations are as for Fig. 3, except that noncoding exons are also shown where marked. The lines or boxes showing genes, probes, contigs, exons, or introns are not proportional in length to those items, except where noted. A, the detailed arrangement of the three genes. Also shown are the genome contigs (plus one deposited additional sequence, AC145196) used for our analyses. Overlaps of the contigs are not shown. The two presently residual gaps in the genome data base sequence are located here showing the nearest intron boundaries and with exon 5 as yet of uncertain position. B, positioning by probe sequences (from phage clones derived from chromosome 15 DNA) which define the start and 5'-region of the α5 gene up to the region bordering the first of the two gaps. The probe sequences α–f are specified under “Experimental Procedures.” The lengths in this panel are shown on the scale. An exception to that, due to the small exon lengths, occurs for the arrows indicating the locations, relative to the probes and contigs, of the exons 1–5 (contained in probe b) and exons 6–8; there only the tip of the arrow marks the end of exon 5 or exon 8. C, the structure of the α5 gene. Exons 1–4 are entirely in the 5'-utr (broken line); exons 1–3 become expressed separately in three alternative full-length transcripts, each also containing exons 4–13. The dotted line denotes the region of presently unknown genomic sequence. D, the structure of the α3 gene. If there is a preceding exon of 5'-utr DNA, it has not yet been identified in any transcript. The broken line indicates the distance between known exons 4 and 6, because exon 5 and its surroundings are in the residual gap in the genome sequence, the position of which is marked by the dotted line.
The \( \theta \) and \( \epsilon \) Genes and Their Relationship to the \( \beta 4 \) and \( \gamma 4 \) Subunits—Much change in the evolution of \( \theta \), from the variation in amino acid sequences seen between the mouse and human \( \theta \) subunits (exceptional for the highly conserved mammalian GABR family), has been proposed by Sinkkonen et al. (36). A unique feature of the mammalian \( \theta \) subunit sequences is a long insertion in the second intracellular loop before the last transmembrane domain (one addition of 86 amino acids, plus two smaller insertions nearby, in human \( \theta \). We noted that when this specialization is subtracted, the mature subunit \( \theta \) shares 50% amino acid identity with the chicken \( \beta 4 \) sequence.

Most surprisingly, the latter is its closest relative in the entire GABR family known from vertebrates. Indeed, the \( \theta \) gene was originally termed human \( \beta 4 \) (13, 33, 35) because its structural similarity to chicken \( \beta 4 \) was recognized. \( \theta \) in heterologous expression resembles known \( \beta \) subunits in forming an \( \alpha 9 \) combination at the cell surface (and not \( \beta 0 \)), but differs in that this combination does not produce active receptors until a \( \gamma \) subunit is added (34).

Only three human \( \beta \) subunits have been found by cDNA cloning, having mostly similar properties when in receptor combinations (5, 7); furthermore, no \( \theta \) subunit has been found in any submammalian species. Therefore, we asked whether a human \( \beta 4 \) exists undetected in the genome or whether the mammalian \( \theta \) subunit could have evolved as a modified subtype from the \( \beta 4 \) subunit line. The chicken \( \beta 4 \) subunit DNA sequence was used to search the full human genome for homologous sequences. The \( \beta 2 \), \( \beta 3 \), \( \beta 1 \), and \( \theta \) genes were thus identified as the closest relatives (in that order), followed less closely by other known GABR genes, but down to nonsignificant alignments, no unknown gene that could be a GABR gene was thus found, i.e. a human \( \beta 4 \) cannot be found.

It is therefore possible that, as was suggested previously (36), the \( \beta 4 \) subunit has evolved into \( \theta \), although by its nature this hypothesis cannot be directly established. When the chicken genome is completed, it will be tested to see if a \( \theta \) gene is found there. We have observed (data not shown), however, that the exon/intron organization of the human \( \theta \) gene is exceptionally similar to that of the chicken \( \beta 4 \) gene. The exon lengths and the positions of the boundaries are identical in \( \theta \) and \( \beta 4 \) (disregarding the first exon containing the signal peptide and the final exon where the unique 86-residue insertion in \( \theta \) falls, as noted above). The boundary amino acids are totally conserved (other than one Ile/Leu change). The 16 exonic nucleotides around each boundary show overall 24% change between these two subunits, compared with 36% or more from chicken \( \beta 4 \) to the three human \( \beta \) subunits and much less than that between chicken \( \beta \) subunits and other human non-\( \beta \) subunits. Hence, \( \theta \) behaves as being in the \( \beta \) family in its gene organization and is closest to \( \beta 4 \) there.

The human \( \epsilon \) gene has chicken \( \gamma 4 \) as its nearest vertebrate relative by overall sequence identity, slightly closer than human \( \gamma 3 \). Three \( \gamma \) subunit genes are known in man (Figs. 3 and 4) and also in the chicken, where no \( \gamma 3 \) subtype has so far been found but a \( \gamma 4 \) subunit has, which in functional properties is quite distinct from other \( \gamma \) subunits and hence has been regarded as a fourth \( \gamma \) subtype (37, 38). Avian \( \gamma 4 \) has relatively low protein sequence identity (64–67%) with human \( \gamma \) subunits and much less with all other human subunits except \( \epsilon \).

We searched the full human genome with the chicken \( \gamma 4 \) sequence; the closest partial alignment was with the \( \gamma 2 \) gene but \( \epsilon \) was close, followed by \( \gamma 3 \) and \( \gamma 1 \), and at much lower \( p \) values, a variety of other GABR and GLR subunits, and then weak hits on noncoding sequences.

It is not known if \( \epsilon \) exists outside the mammals. Sinkkonen et al. (36) suggest that the avian \( \gamma 4 \) may have evolved into mammalian \( \epsilon \), with a change (both in sequence and in function) that is unusually rapid for GABR subunits. The organization of the three GABR genes (\( \epsilon - \alpha 3 - \theta \)) on the human X chromosome (Fig. 3E), plus their same order found on the mouse X chromosome (35, 39), when compared with the other clusters would be consistent with \( \epsilon \) being derived from the \( \gamma \) line and \( \theta \) from the \( \beta \) line of GABR genes of an ancestral chordate. However, species specialization within the \( \gamma 3 \) subgroup might have occurred independently.

In summary, we can now say that no new human GABR gene, which could be a true orthologue of \( \gamma 4 \) or of \( \beta 4 \), can be found. The latter conclusion is reinforced by the absence of any new \( \beta \)- or \( \gamma \)-like sequences in the general search (above) using the Cys loop consensus. Although an absolute negative conclusion cannot be proven here, for all of these reasons we conclude that there is an extremely low probability that any previously unknown GABR gene is present in the human genome.

The \( \rho \) Subunit Genes—Fig. 1 shows that these lie in a separate group, sharing only 28–38% amino acid sequence identity with the other subunits. The human \( \rho 1 \) and \( \rho 2 \) genes (GABRR1 and GABRR2) were both earlier mapped to the chromosome region 6q13-q16 (40, 41). A \( \rho 3 \) subunit has been known only in the rat (42).

The locations of the GABRR1 and GABRR2 genes can now be seen in the genome to be at 6q15 and are in fact very close, with no recognized intervening gene (Fig. 3D). As described above, we located with the consensus A screen a human GABR \( \rho \)-type gene, lying on chromosome 3. To obtain the full gene there, we searched the human genome with the rat GABRR3 protein sequence of Ogurasu and Shingai (42) as a query sequence (TBLASTX), and we retrieved at a high probability level nine \( \rho \)-type exons. These were back-translated into DNA sequences, and with these new probes we searched the human genome again (BLASTN), and we located all of them as exons of one complete gene (Fig. 5B). This lies within the AC026100 contig at the point which we had located, by the consensus A screen, on chromosome 3; its position is detailed in Fig. 5C. Fig. 5A presents the 467-amino acid sequence of the human \( \rho 3 \) subunit. It includes a stretch encoding 99 amino acids with homology to rat \( \rho 3 \), which had been noted on human chromosome 3 as within a noncoding sequence (41). The human \( \rho 3 \) shares 81% identity with the rat \( \rho 3 \) subunit, with the changes being mainly in the signal peptide and in a 21-residue segment following from that, as well as in a 26-residue segment in the last intracellular loop.

Analysis of the genome now allows us to compare the genomic organization for the three human \( \rho \) subunits. The \( \rho 3 \) gene has nine exons, shown in Fig. 5B. The \( \rho 2 \) gene also has nine exons; corresponding exons in \( \rho 2 \) and \( \rho 3 \) are of essentially identical lengths. It is also interesting to find in the genomes of the mouse and rat a \( \rho 2 \) gene (RefSeq accession numbers NM_008076 and NM_017292, respectively), each again adjacent to a \( \rho 1 \) gene and again with nine exons for \( \rho 2 \). The nine \( \rho 2 \) exons are of corresponding sizes and are virtually identical in amino acid sequence in man, mouse, and rat. Their exon boundaries also show a high degree of conservation in all three \( \rho \) genes. The human \( \rho 1 \) gene, however, has 10 exons; the extra exon is inserted as an exon 2, with exons 3–10 of \( \rho 1 \) being homologous and equivalent to exons 2–9, respectively, of the \( \rho 2 \) and \( \rho 3 \) genes. This pattern is conserved in the rat and mouse \( \rho 1 \) genes, each having 10 exons, and their respective coding lengths are again identical in the three species. Indeed, in all three \( \rho \) subtypes in man, all of the exons (apart from the first and last, which contain some untranslated sequence, and excluding the extra exon in \( \rho 1 \)) do not differ by more than four nucleotides in their corresponding lengths. Hence, despite the
great conservation of the genomic organization throughout this set, it is the two genes that are adjacent on one chromosome (γ1 and γ2) that differ in exon number, thus suggesting a more recent separation.

The acquisition of an extra exon in the γ1 gene of this pair raised the possibility that this segment modifies the γ1 receptor function and may be a site of alternative mRNA splicing, to give 9-exon and 10-exon forms. Two functional alternative
forms have been reported in the ρ subunits, and in fact these are in π1 (43). The extra exon 2 is indeed spliced out, as discussed below. We also searched the genome data bases for any further transcripts that would align with the ρ subunit genes, but no differences in expressed exons appeared, with one exception. A transcript annotated as human ρ1A (Ensembl Vega transcript, accession number Q9YGG4) was found to contain new and very different forms of exons 1 and 2. However, when these were used in a TBLASTX search of the genome, those new sequences were not found, indicating that they are artifactual.

The δ and π Subunit Genes—The π subunit is unusual in being expressed primarily in peripheral tissues, e.g. human uterus, lung, thymus, and to a greatly limited extent in the brain (44). It has been little studied as yet but was shown to have the capacity to combine with an αβ or an αβδ set of subunits to confer upon them new functional properties (45). The gene, GABRP, was located by radiation hybrid mapping to be on chromosome 5q near the telomere (12), and in the genome we found it in 5q35 (Fig. 3C). This is outside the 4-member GABR cluster in 5q34 (Fig. 3B).

In the GABRP gene we found 10 exons, but exon 1 contains only the 5′-utr sequence, whereas exon 2 has further 5′-utr sequence (43 bp) and then encodes the first 18 amino acids of the signal peptide. Exon 10 encodes the final 100 amino acids followed by 1789 bp of 3′-utr sequence, including a polyadenylation signal and poly(A) tail sequences. Hence this gene conforms to the usual GABR pattern of nine coding exons.

The π gene location is at 8.6 Mbp distal to the end of the cluster in chromosome 5q of the genes for the β2, α6, α1, and γ2 subunits, which, for comparison, covers a total length of 1.1Mbp (Fig. 3B). Between that cluster and the π gene other genes occur, on the order of 25 genes or more, some as yet unidentified but all clearly unrelated to GABR genes, whereas within the cluster only one non-GABR gene can be detected. This makes it uncertain whether or not the π gene was directly related in evolutionary origin to the GABR genes in that cluster. We approached this question by species comparisons. A mouse cDNA sequence with high homology to the human π cDNA sequence has been deposited in the RefSeq data base (Ref. Seq accession number NM_146017), and although the mouse transcript has not yet been reported as being translated and functional it has been assumed in the data base to be derived from an orthologous mouse gene, Gabrp, in view of its 93% predicted amino acid identity to the human π subunit. A BLAST search with this cDNA (99% match on the exons) locates the Gabrp gene on cosmid AL669814 on mouse chromosome 11 in the A5 sub-band (at 33.91–33.94 Mbp). Indeed, this is in the region syntenic to human chromosome 5q33–35. Furthermore, four mouse GABR genes, for the same four GABR subunits that form the cluster in that region of human chromosome 5, were found in the adjacent region of mouse chromosome 11, i.e. band B1.1. Those four genes form a similar cluster and are in the same order and orientation in the two species. The mouse GABR gene lies beyond the γ2 end of the cluster, as on the human chromosome, and at a distance of 8.6 Mbp in both species. The complete conservation of this arrangement of five genes during evolution from a common ancestor of mouse and man supports the concept that the π gene is not in this vicinity by chance but by divergence from an ancestral precursor of that cluster.

There is also genomic evidence that the π subunit exists in lower vertebrates. In the chicken, a homologous EST of 660 bp has been reported (EMBL data base accession number BM440925). In the pufferfish (F. rubripes) genome (currently partially sequenced), we found by searches an incomplete reading frame (SINFRUG00000148202) comprising 5.9 kb from a putative orthologue of the human GABRP gene. Both the predicted avian and fish protein sequences gave clearly the highest match to the π subunit in reciprocal BLAST analysis against the human genome. The pufferfish gene partial sequence has eight predicted exons, which correspond to the region containing human exons 3–10 (the start of the coding region up to part of exon 3 being not yet sequenced in the Fugu genome). The introns are all small, as generally in Fugu. The predicted protein sequence of the pufferfish shares 62% identity with the human sequence over that region, with almost all of the variation being in the long second intracellular loop of each, generally variable. Exons 4–8 have identical lengths in the fish and human sequences. The chicken nucleotide partial sequence corresponds to the final exon 10 of the human and fish π genes. The coding sequence within that exon (which is followed by a lengthy 3′-utr in each case) shares 74% identity at amino acid level between man and bird.

The δ gene, GABRD, is isolated, being the sole GABR gene on chromosome 1. It was located in band p36.3 by Windpassinger et al. (46), who also obtained cDNA sequences to deduce that it has 11 exons, 10 of which are coding. Three of these were proposed to be expressed alternatively, although three corresponding subunit cDNAs were not isolated from tissues. We have confirmed in the genome that this gene is in 1p36.3, but only eight exons are currently annotated as such in data bases, with seven of these in full agreement with the last seven exons reported in Ref. 46. First, by BLASTN search of the genome with cDNA sequences covering the other four reported exons, we found in the GABRD gene that a noncoding exon is indeed present in the 5′-utr, being in fact the first exon of the gene; this corresponds to the reported (46) GABRD “1B” sequence (except for the first 12 nucleotides of 1B, which are not present in the GABRD gene). Second, exon 2 of GABRD starts with the sequence of the “5′-variant exon 1C” of Ref. 46, which contains a further 301 bp of 5′-utr and then encodes the initial 37 amino acids including the signal peptide. However, there is no adjacent intron following, and exon 2 in fact continues with a further 115 bp that encode the next 39 amino acids. Then the remaining seven exons previously described, together with their introns, were confirmed as completing the gene. The last exon has both the C-terminal coding sequence and a considerable length of 3′-utr, including a polyadenylation signal and poly(A) tail. The updated location and genomic structure, with a total of nine exons, are shown in Fig. 6.

The apparent 11th GABRD exon (46) was termed 1A, as an initial exon expressed as an alternative to 1C and thus introducing a different signal peptide. This sequence had been obtained only from ESTs. However, after various BLAST searches with that exon 1A as a query, we cannot find the 1A
sequence or anything approaching it in the human genome or in the contig involved. Windpassinger et al. (46) had also noted that they could not find exon 1A in the then available genome data bases, attributing this to gaps in the map. Therefore, we have omitted it from the GABRD structure. However, some mystery remains, because the second version of the signal peptide that is encoded by that proposed alternative exon 1A (46) would have 60% identity to the rat and mouse signal peptides, as known from their cDNAs. In particular, 1A would translate to give an unusual motif in the center of the signal peptide, LLXPLLLLC, which does occur identically there in those two rodent δ subunits. In contrast, the signal peptide encoded by exon 1C, as found also in the presently known human δ subunit cDNA, has no homology in the signal region to the rat or mouse δ subunits. Therefore, we cannot fully exclude an additional coding exon, which would be same as or similar to exon 1A of Ref. 46 and which can be spliced in as an alternative to exon 2. More cloning of human δ cDNAs from different tissues is needed to test for this possibility. If found, it must still be explained why it is not detected in the genome by normal procedures, because no gap in the tile path in this region is now shown. A population difference is a conceivable cause.

Another δ transcript with a different 5’-end was also proposed, based upon ESTs and 5’-extensions made by PCR of partial δ cDNAs (46). That would be without the aforementioned 1B 5’-utr sequence. We find that the latter segment constitutes exon 1, so alternative splicing might occur there to give the a and b transcripts indicated in Fig. 6. Evidence is still lacking for tissue expression of the entire native δ subunit mRNA in alternative forms, and likewise for their translation and functionality.

Location of a Gene Related to the Three Known Human Glycine Receptor α-Subunit Genes—Our search with the 8-point Cys loop algorithm (Fig. 2A) gave a strong match at an unknown gene sequence on the X chromosome (q-arm). The predicted protein sequence around this Cys loop has a Thr at 5 positions before the first Cys and a Met at 2 positions after the second Cys, a signature of the glycine receptor α subunits (Fig. 2D). For the three human glycine receptor α subunits, α1–α3, which have previously been cloned, genes have been located on chromosomes 5, X (p-arm), and 4, respectively (47). Those three genes (GLRA1–3) also gave strong matches with the 8-point (2A) Cys loop screen. A fourth glycine receptor α subunit cDNA from the mouse has been cloned and expressed (48) but has not been found in man, and the question whether a human glycine receptor α4 exists has been left open (49). None of the human genome data bases or browsers currently lists a GLRA4 gene.

Therefore, the mouse α4 protein sequence of Harvey et al. (48) was used as a query sequence (TBLASTX) to search the human genome. We retrieved nine peptide sequences that show a high degree of homology to the mouse Glra4 exons. Nucleotide probes based on these peptide sequences were then used as the basis of a nucleotide search (BLASTN), and we were able to locate all nine probes as exons of one gene within two adjacent contigs on Xq22.3 (Fig. 7). These encode a sequence closely related to the mouse glycine receptor α4 subunit, but this has a stop codon just before the predicted 4th transmembrane domain (Fig. 7) and is thus a pseudogene. This gene lies far from the GABR cluster present on Xq (Fig. 3E), which starts at 47 Mbp downstream (Fig. 7). It is noteworthy that the mouse gene for the functional glycine α4 subunit is also located on the X chromosome (49) in a region that is, in fact, syntenic to that in Fig. 7.

Six Genomic Mechanisms for Creation of Multiple Forms of GABR Subunits by Alternative Splicing

With four of the GABA_4 receptor subunits, two alternatively spliced functional proteins have been found earlier to be expressed in tissues, as noted below for β2, γ2, γ3, and ρ1. However, for nine GABR subunits in all, alternative splicing of the pre-mRNA can be found, and we here relate this to their genomic structures.

An alternative, functional “long” form (γ2L) with an insertion in the second intracellular loop was reported previously in the γ2 subunit and another in β2 (β2L) (Table I) but at different positions along that loop. The exon spliced in is now identified as exon 9 or 10, respectively, in the human genome (Table I).

For the β2 case, the subunit is exceptionally conserved, with all its other coding exons identical in human and rat genomes other than one codon difference (Ser/Thr-372). When compared with that conservation of the short form (seen even in the chicken (50) β2 protein, 98% identical to human), the reported lack of detection (50) of a rat or bovine version of the alternative spliced long (β2L) mRNA could mean that this additional exon present in the human gene is of low biological significance or has a special role in man. Therefore, we searched (TBLASTN) the rat and mouse genomes; a β2 subunit gene exists there on rat chromosome 10 and on mouse chromosome 11. Although those rodent genes are annotated as having only 9 exons, our searches showed first that each gene has (as in man, Table I) an additional initial exon that contains only 5’-utr sequence. Second, a genomic sequence exists within each that corresponds to the spliced human β2 exon 10 in sequence and position and with the same boundaries. Its start is at 27,795,098 bp along rat chromosome 10 but encoding 31 amino acids in place of the 38 in man and in mouse. The rat sequence is IFYKDKIQNRTQQSLWDPT-RWTYHYFSLY-, whereas the dashes denote a 6-amino acid internal deletion and a C-terminal single one and boldface type denotes the five differences, all relative to the human. The mouse exon 10 differs from human exon 10 only at two positions, 1(M) and 22(D). Therefore, there are 11 exons in GABRB2 in all three species. Four of the five potential protein kinase sites in human and mouse exon 10 (Table I) are retained in the rat. All of the various features conserved support the view that the known alternative splicing of exon 10 in the human β2 transcript is of functional importance.

Five other types of alternative splicing found in the GABA_4 receptors are described in Table I. Type 2 is seen in the β3 gene. It is known (31) that the relative abundance of the two forms of β3 mRNA varies with the tissue, that each 5’-utr in question contains functional promoter activity, and that the first exon 5’-utr sequence is almost identical in human and rat DNA. Those results suggest that this alternative splicing has a role in the control of expression of the mature β3 subunit. Furthermore, the full exon/intron structure of GABRB3 was reported by Glatt et al. (17), revealing 10 exons, confirmed now in the genome. Nevertheless, each expressed form of this subunit again uses nine exons, as in other subunits.

For type 3, alternative splicing creates transcripts that differ only in their 5’-utr sequences and promoters. Six α2 subunit forms can thus arise (Table I), all being functional and differentially expressed in development in the rat (51), and we found equivalent exons for all such forms in the human α2 gene. The human α5 gene exhibits a similar phenomenon, not found in other GABR genes (Table I). By searching the relevant genome regions, we found 12 exons (3 noncoding) in the human α2 gene and confirmed 13 in α5 with details shown in Fig. 4C. Their 5’-utr sequences are, again, unusually conserved between rat and human, consistent with a regulatory function.
In type 4, alternative forms are expressed with several coding exons (contiguous except in β2) spliced out together. For the ρ1 subunit, two variants of the original ρ1 were cloned (43). One is active and has a type 1 deletion (Table I), which we found to correspond to the loss of exon 2. The other is inactive, losing a 150-amino acid segment of the N-terminal domain. We found this is due to the loss of exons 2–5, so that a basis for the proposed alternative splicings exists. The active ρ1 form lacking only exon 2 would be equivalent in exonic structure to ρ2 and ρ3 (Fig. 5) and therefore may well have some biological significance.

For the α4 subunit, a very severe, inactivating truncation arises in both the human and mouse α4 transcripts from the deletion of 929 bp and an introduced frameshift (52). This product corresponds to the splicing out of exons 3–8 of the nine α4 exons. Ma et al. (52) suggest that such inactive truncated forms, where expressed as polypeptides, may have a regulatory function in the assembly or trafficking of some GABA_A receptors.

Proteins that are much truncated are also predicted to arise from alternative splicing of the ε gene mRNA. In the studies of X chromosome genomic DNA clones by Wilke et al. (33), as noted earlier, the human GABRE gene was deduced to contain eight exons, exceptional among GABR genes. However, Sinkkonen et al. (36) found nine exons for the ε subunit by screening libraries with fragments of the ε cDNA. We have confirmed a total of 9 ε subunit exons in the genome, all nine being coding exons. Four expressed ε variants can be found, the first being the full-length functional transcript (33, 53–55). In variant 2, exons 1–3 are spliced out, with protein translation then starting at an alternative Met. In variant 3 that same splicing occurs in combination with deletion of residues 127–158 from the center of exon 4, using an internal cryptic site there for intron-independent splicing. In variant 4 only exon 1 is deleted. The mRNAs of variants 2 and 3 have indeed been found by Wilke et al. (33) in several peripheral tissues. The full-length human ε expression is high in heart and lung and is also present in brain (33, 54, 55), and the same is true for the rat (36); in human heart it is confined to the electrical conducting system and is located throughout it (54), suggesting a specialized non-neuronal role there. The other variants lack the signal peptide and also (for forms 2 and 3) lack a significant region of the extracellular N-terminal domain, but roles for those in combination with other (full-length) subunits or in regulation of GABA_A receptor trafficking remain possibilities.

Another origin of two forms of a GABR subunit can be an intronless mRNA splicing at an internal cryptic consensus splice site (type 5). This was originally found for two chicken β4 transcripts (56). It also occurs with the human γ3 gene, where Poulsen et al. (57) found a second transcript having an insert of only 18 bp, which is similar in its position and in introducing a protein kinase C site to that created by conventional splicing in γ2 mRNA. For the γ3 case, the secondary splicing occurs at a cryptic site in the immediately adjacent intronic sequence; we confirmed that no corresponding extra coding exon exists in this gene (Fig. 4D). No transcript which has been conventionally spliced at an exon/intron boundary has been reported for γ3, but it cannot yet be excluded that a 9-exon form is also expressed in some locations.

Alternative splicing has thus been reported for the pre-mRNAs of two of the three γ subunits, but no such additional full-length γ1 transcript (nor an indicative EST) has been reported. However, by analysis of the GABRG1 gene, we found only the nine exons that correspond, in their lengths and positions, to the nine exons in GABRG2 that encode the short form of the γ2 protein (data not shown).

Finally, a novel form of GABR alternative splicing (type 6) has been discovered recently in the human γ2 gene product, independent of its exon-9 splicing described above. In this form, an additional 120-bp exon is created by a weak splice site inside intron 5 by an Alu repetitive element residing there (58). The protein product may affect the trafficking of active GABRA1 receptors (58). We can regard this phenomenon as inserting a pseudo-exon 5a. The 300-bp Alu sequence is specific to primates and is common in man, producing ~5% of human alternative splicing. It is also present in introns of several other GABR genes: α1, α3, α5, β1, γ1, ρ1, and π. If in any of those genes a favorable intronic splice site is created, this could give rise to a well expressed alternative GABR form not predictable from the exon content.

**Common Features of the GABR Gene Clusters**

The clusters have now been specified more fully than was previously possible. It is clear that there is a common pattern of GABR gene arrangement over chromosomes 4, 5, 15, and X of β-[α]-α-γ, where there are either 1 or 2 α genes and where in one case (on X) a β gene can be replaced by the related θ and the γ by the related ε (36) (Figs. 4 and 5). The GABR set appears to exhibit the most extensive degree of clustering of the genes relative to their total number which is known (excluding the special case of the exceptionally numerous olfactory and gustatory receptors) for any one receptor family. Because the chromosome 15 cluster has the same pattern as the clusters on chromosomes 4 and 5 but lacking a second α subunit gene, we searched the intervening sequences within that cluster (see above) for any unidentified GABR-like gene and found none.
The δ gene and the three ρ genes lie outside those clusters (Figs. 3, 5, and 6). The ρ subunits can co-assemble in a pool separate from that of the other GABR subunits and produce a very distinctive pharmacology (59). The chromosomal isolation of the ρ genes, taken together with the latter features, suggests that an ancestral ρ separated from the other GABR genes at a very early stage in GABR evolution. This is supported by the position of the ρ set in dendrograms of vertebrate/invertebrate GABR subunits; there the mammalian ρ is much closer to the RDL GABR subtype (which has some ρ-like pharmacological features) from insects (23) and from C. elegans (60) than to other mammalian GABR subunits.

The transcriptional orientations of the GABR genes have been observed previously to differ within a cluster (13, 17). We could now confirm directly that in the chromosones those orientations are in a uniform pattern in the chromosome 4 clusters, with the β gene (or θ, in the X-chromosome cluster) running head-to-head to its α subunit neighbor. This may indicate that a common regulatory element lies between that pair of α and β genes. Some support for this idea comes from knock-out mice in which the α6 gene was disrupted by neomycin gene insertion into its exon 8 (61). The mRNA and protein levels of the α1 and β2 subunits in the forebrain were selectively decreased, those being the products of the two chromosomal neighbors of GABRA6. One possible interpretation of the GABR gene clusters would be that there is an inbuilt mechanism there for some co-ordination of their expression, but if so this must be regionally or developmentally restricted, because GABR subunit co-assembly in general is not at all restricted to combine subunits from each clustered gene set. It has been noted previously, however, that some common GABR co-assemblies are indeed produced from co-clustered genes, e.g. α1 β2 γ2 is the most abundant GABR subtype (62). Another interpretation of the clusters (13, 29, 30) is that they arose from an ancestral αβγ receptor set by a series of duplications, sequence divergences, and chromosomal translocations. There is evidence, for example, that human chromosomes 4 and 5 had a common single chromosome ancestor (63), which could explain the similar pattern of their GABR clusters. So far as the GABR genes have been mapped to mouse chromosomes, they form the same clusters with the same order and spacing, in diverse regions that are syntenic to the translocated regions carrying them on the human chromosomes. This is exemplified by the results we described for the mouse equivalent of the human chromosome 5 cluster. Hence these clusters remain stable in some chromosome rearrangements, giving credence to the hypothesis that they survived throughout earlier ones. The two interpretations of the clustering are not mutually exclusive.

### Common Features of the GABR Genes

A common pattern persisting through at least 17 of the GABR genes was found, i.e. with nine coding exons expressed in a major transcript. Only two of the genes are apparent exceptions: GABRD has 8 exons and GABRG3 has 10 coding exons with no known alternative exon splicing. The overall sizes of the genes (or transcription units), where estimated at all in the past, were sometimes in error. They cover a surprising range for one family, from 8 (GABRD) to 601 kbp (GABRG3) (Figs. 3–6). The cluster on chromosome 15 is exceptional, with α, β, and γ genes all having very large introns, which may be a sign of a different evolutionary history to the other clusters.

The ubiquitous presence of the disulfide-bridged loop structure provided an essential tool for this study. Our analyses of

---

**TABLE I**

Genomic origins of alternatively spliced forms of human GABA<sub>A</sub> receptors

| Type | Origin | Subunit | Total exons<sup>a</sup> | Sequence differences | Effects |
|------|--------|---------|-------------------------|---------------------|---------|
| 1.   | Insertion or omission of an exon | β2 | 11 | 38-aa insertion in second IC loop (70)<sup>b</sup> | Introduces potential sites for PKA, PKC, CPMK II, CK1, CK2 |
|      |        | γ2 | 10 | 8-aa insertion in second IC loop (71–73) in human, rat, and mouse. This is the ninth exon | Introduces site phosphorylated by PKC and CPMK II (71, 74) |
|      |        | ρ1 | 10 | Omission of exon 2 (51 bp) (43) (and adding Ser from split codons) | Functional in oocyte expression (43) |
| 2.   | 1 exon exists in 2 alternative forms<sup>c</sup> | β3 | 10 | Splicing in of alternative exons 1 or 1a, with different promoters in their 5′-utrs (31) | Two alternative signal peptides of equal length but 80% different, with different ATG starts (31) |
| 3.   | Different 5′-utr exons with specific promoters | α2 | 12 | Three 5′-utr exons variably spliced, each with its own promoter (51) | Six alternative mRNAs, as seen in rat (51), all differing in 5′-utr, only 1–3 exons spliced in alternatively, producing three mRNAs (32) |
|      |        | α5 | 13 | Four 5′-utr exons variably deleted, each with its own promoter (32) | Three new variants formed (33, 54) |
| 4.   | Loss of multiple exons | ε | 9 | Exons 1–3 and part of 4 can be variably deleted, with an alternative start Met | One inactivating variant, but regionally and developmentally expressed in brain; up to 42% of brain ε mRNA (52) |
|      |        | α4 | 9 | Exons 3–8 can be deleted together (52) | Translation not investigated |
|      |        | β2 | 11 | Exons 3 and 10 both deleted<sup>d</sup> | Inactive in oocytes. Translation in vivo is not investigated |
|      |        | ρ1 | 10 | Second new variant, with exons 2–5 deleted (43) | Introduces PKC site |
| 5.   | Alternative use of cryptic site for intronless splicing | γ3 | 10 | 6-aa insertion in second IC loop (57) | Protein product cannot reach cell surface (58) |
| 6.   | Use of cryptic splice site at intron ALu sequence | γ2 | 12 | Insertion of new 120-bp exon within intron 5 (58) | |

<sup>a</sup> The total number of exons, coding and non-coding, as determined or confirmed here on the genome.

<sup>b</sup> The β2 gene has 11 exons: exon 1 has 74 bp of 5′-utr (unreported), exon 2 has 152 bp of 5′-utr plus 77 bp coding for the signal peptide, exon 10 is expressed only in the "β2L" form.

<sup>c</sup> This may occur also with the δ subunit, but expression in vivo of the second exon has not yet been confirmed; see text on the δ subunit.

<sup>d</sup> In the human genome we located, using Grail Express analysis, a similar set of three 5′-utr exons with very high homology to the rat set, but alternative transcripts formed with these have not been investigated so far.

<sup>e</sup> Ensembl sequence ENSESTT00000025025.

---

**GABA<sub>A</sub> Receptor Genes in the Human Genome**

Only forms that have been shown to be transcribed in vivo are included. The abbreviations used are as follows: CMPK, calmodulin-dependent Ser/Thr protein kinase; CK, casein kinase; IC, intracellular; aa, amino acid; PKC, protein kinase C; PKA, cAMP-dependent protein kinase.
all human GABR genes also found that the exon encoding the Cys loop and surrounding sequence has an invariant length, i.e. 83 bp. This holds further in all GABR subunit gene structures known in the vertebrates. This precise conservation of exon length is unique among the exons of GABR genes and again suggests a common and essential role for this structure in all vertebrate GABA<sub>δ</sub> receptors.

The role of the Cys loop has been studied in the nAChRs, where evidence has been found that it acts as an essential core in the tertiary folding of the major N-terminal domain of each subunit and that it guides, by a conformational change in the loop, the interactions of the subunits to form the active oligomer (64, 65). In addition, two recent mutagenesis studies, on Cys loop in GABA<sub>δ</sub> and in glycine receptors (66, 67), have provided direct evidence for a role in the loop in relaying the agonist binding conformational changes to the channel domain. We deduce that these requirements must impose an invariant framework on the Cys loop structure in the entire GABR family and throughout its evolutionary history.

Furthermore, there is a striking absolute conservation, throughout the known vertebrate and invertebrate GABR, of a single associated residue outside the loop, namely Ser at 6 places downstream (Fig. 2E). This Ser is likewise maintained in two of the other families of vertebrate Cys loop receptors (Fig. 2). This Ser lies in loop B of the currently structurally based 6-loop model of the ligand binding domain, first found in the nAChRs (68) and extended also (69) to vertebrate GABRs. The essentiality of this Ser in the human GABR γ2 subunit has been demonstrated very recently by its mutation, which prevents active receptor formation with wild-type α and β subunits (58). We note that all of the evidence cited here supports the concept of a constant and essential role for the framework (consensus E) that we have used in the analysis.

Acknowledgments—We thank Drs. D. Bentley and G. Leschziner (Wellcome Trust Sanger Institute, Hinxton, UK) for helpful discussions.

Note Added in Proof—A preceding version of the human genome (used in the online advance version of this paper) showed an arginine codon at position 433 of the sequence of the glycine receptor α subunit, but we have recently found that this has since been corrected in the genome to a stop codon. In support of this, Dr. Robert Harvey has informed us that he has found this pseudogene sequence in human DNA and mRNA samples. The statement on this gene is corrected herein.

REFERENCES
1. Baulac, S., Huberfeld, G., Gourfinkel-An, I., Mitropoulou, G., Beranger, A., Prud’homme, J.-F., Baulac, M., Brice, A., Bruzzone, R., and LeGuern, E. (1999) Nature 399, 70–73
2. Kim, H. Y., and Olsen, R. W. (2001) FEBS Lett. 502, 213–216
3. Leventhal, A. G., Wang, Y., Pu, M., Zhou, Y., and Ma, Y. (2003) J. Biol. Chem. 278, 22225–22230
4. Mo, Y., and Wainer, H. W. (1999) Trends Pharmacol. Sci. 20, 187–192
5. Nayeem, N., Green, T. P., Martin, I. L., and Barnard, E. A. (1994) Handb. Exp. Pharmacol. 119, 291–314
6. Barnard, E. A. (2001) Mol. Pharmacol. 59, 356–370
7. Barnard, E. A., Skolnick, P., Olsen, R. W., Mohler, H., Sieghart, W., Biggio, G., Dichter, R. B., Sattelle, D. B., and Ffrench-Constant, R. H. (1996) FEBS Lett. 392, 222–227
8. Barnard, E. A., Darlison, M. G., and Seeburg, P. (1987) Trends Neurosci. 10, 509–512
9. Nayeem, N., Green, T. P., Martin, I. L., and Barnard, E. A. (1994) Handb. Exp. Pharmacol. 119, 270–309
10. Buckle, V. J., Fujita, N., Ryder-Cook, A. S., Derry, J. M. J., Barnard, P. J., Atack, J. R., Dawson, G. R., McKernan, R. M., Whiting, P. J., and Rosahl, T. W. (2001) Neuron 29, 2094–2099

11. Bell, M. V., Bloomfield, J., McKinley, M., Patterson, N. M., Barnard, M. G., and Johnson, K. J. (1999) J. Mol. Biol. 290, 839–844
12. Bailey, M. E. S., Albrecht, B. E., Johnson, K. J., and Darlison, M. G. (1999) Biochim. Biophys. Acta 1447, 307–312
13. Martinez-Torres, A., Vanzea, E. A., Panicker, M. M., and Miledi, R. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 2019–2024
14. Ogurusu, T., and Shingai, R. (1995) FEBS Lett. 365, 77–81
15. Ortells, M. O., and Lunt, G. G. (1995) Trends Biochem. Sci. 20, 356–361
16. Greger, V., Knoll, J. H., Woolf, E., Glatt, K., Tyndale, R. F., Deforey, T. M., Muelhardt, C., and Becker, C.-M. (1998) J. Neurosci. 18, 5911–5923
17. Glatt, K., Glatt, H., and Lalande, M. (1997) Genomics 41, 63–69
18. Barnard, E. A., Darlison, M. G., and Seeburg, P. (1987) Trends Neurosci. 10, 509–512
19. Barnard, E. A. (1992) Trends Biochem. Sci. 17, 368–374
20. Ortelts, M. O., and Lunt, G. G. (1995) Trends Neurosci. 18, 121–127
21. Betz, H. (1999) Neuron 23, 385–392
22. Karlin, A., and Jorgensen, E. M. (2003) Nat. Neurosci. 6, 1145–1152
23. Hosie, A. M., Arinstein, K., Sattelle, D. B., and Ffrench-Constant, R. H. (1997) Trends Neurosci. 20, 573–583
24. Jones, A. K., Elgar, G., and Sattelle, D. B. (2003) Genomics 82, 441–451
25. Beg, A., and Jorgensen, E. M. (2003) Nat. Neurosci. 6, 1183–1193
26. Vassilakos, D. K., Elliston, K. O., Pares, P. S., Hamelin, M., Arena, J. P., Schaeffer, J. M., Van der, Ploeg, L. H., and Culy, D. F. (1997) J. Mol. Evol. 44, 501–508
27. Yates, D. M., Portillo, V., and Wolstenholme, A. J. (2003) Int. J. Parasitol. 33, 1183–1193
28. Lodmunder, S. W., Warren, V. A., Williams, B. S., Zheng, Y., Hunt, D. C., Ayer, M. B., Wallace, M. A., Chaudhary, A. G., Egan, M. A., Menike, P. T., Dean, D. C., Garcia, M. L., Culy, D. F., and Smith, M. M. (2002) Biochemistry 41, 6548–6560
29. Melean, P. J., Farh, D. H., and Russell, S. J. (1995) Genomics 26, 580–586
30. Kostrzewa, M., Kriis, B. W., Dixon, J. M. J., Eppelt, K., Kohler, A., Grady, D. L., Steinberger, D., Fairweather, N. D., Merz, R. K., Monaco, A. P., and Muller, U. (1998) Eur. J. Hum. Genet. 6, 266–274
31. Baulac, S., Huberfeld, G., Gourfinkel-An, I., Mitropoulou, G., Beranger, A., Prud’homme, J.-F., Baulac, M., Brice, A., Bruzzone, R., and LeGuern, E. (1999) Nature 399, 70–73
66. Absalom, N. L., Lewis, T. M., Kaplan, W., Pierce, K. D., and Schofield, P. R. (2003) J. Biol. Chem. 278, 50151–50157
67. Kash, T. L. (2003) Nature 421, 272–275
68. Le Novere, N., Grutter, T., and Changeux, J. P. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 3210–3215
69. Cromer, B. A., Morton, C. J., and Parker, M. W. (2002) Trends Biochem. Sci. 27, 280–287
70. McKinley, D. D., Lennon, D. J., and Carter, D. B. (1995) Brain Res. Mol. Brain Res. 28, 175–179
71. Whiting, P. J., McKernan, R. M., and Iversen, L. L. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 330–337
72. Kofuji, P., Wang, J. B., Moss, S. J., Huganir, R. L., and Burt, D. R. (1991) J. Neurochem. 56, 713–715
73. Cheng, A. T., Loh, E. W., Cheng, C. Y., Wang, Y. C., and Hsu, Y. P. (1997) Biochem. Biophys. Res. Commun. 238, 683–685
74. McDonald, B. J., and Moss, S. J. (1994) J. Biol. Chem. 269, 18111–18117
75. Davies, P. A., Wang, W., Hales, T. G., and Kirkness, E. F. (2003) J. Biol. Chem. 278, 712–717
Analysis of the Set of GABA<sub>A</sub> Receptor Genes in the Human Genome
Joseph Simon, Hironobu Wakimoto, Norihisa Fujita, Marc Lalande and Eric A. Barnard

J. Biol. Chem. 2004, 279:41422-41435.
doi: 10.1074/jbc.M401354200 originally published online July 16, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M401354200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 75 references, 20 of which can be accessed free at http://www.jbc.org/content/279/40/41422.full.html#ref-list-1