injection of beta2 mRNA along with an alpha2, alpha3, or alpha4 mRNA (*2, 3*). Although the subunit composition of neuronal nAChRs in vivo has not been determined, the idea that the neuronal receptors are composed of two different subunits is consistent with findings in a recent study (*21*). On the basis of the stoichiometry of the *Torpedo* electric organ receptor, we predict that the neuronal receptor is a pentameric structure.

Detailed studies of in situ hybridization histochemistry (*22*) show that alpha2, alpha3, and alpha4 transcripts are coexpressed with beta2 transcripts (*3*) in many brain regions. This result suggests that the functional combinations observed in oocytes could also occur in vivo. However, the studies also show that in some regions, beta2 and alpha2, alpha3, and alpha4 transcripts are not coexpressed. This observation raises the possibility of the existence of another alpha-type subunit and another beta-type subunit. It would seem, therefore, that there are more than three distinct populations of neuronal nAChRs.

Our studies indicate that the alpha2 gene product functions as a neuronal nAChR subunit with pharmacological features different from those of the alpha3 and alpha4 subunits and that the alpha2 type of receptor differs from all other neuronal nAChRs studied to date. Neuronal nAChRs are a heterogeneous population with respect to composition, distribution, and functional properties.

REFERENCES AND NOTES

1. J. Boulter *et al.*, *Nature (London)* **319**, 368 (1986); D. Goldman, D. Simmons, L. Swanson, J. Patrick, S. Heinemann, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 4076 (1986); D. Goldman *et al.*, *Cell* **48**, 965 (1987).
2. J. Boulter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7763 (1987).
3. E. S. Deneris *et al.*, *Neuron*, in press.
4. P. Nef, C. Oneyser, T. Barkas, M. Ballivet, in *Nicotinic Acetylcholine Receptor: Structure and Function*, A. Maelicke, Ed. (Springer-Verlag, Berlin, 1986), pp. 417–422.
5. R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4853 (1978).
6. M. Kozak, *Nucleic Acids Res.* **12**, 857 (1984).
7. G. I. Bell *et al.*, *ibid.* **14**, 8427 (1986).
8. G. von Heijne, *ibid.*, p. 4683.
9. For review, see R. M. Stroud and J. Finer-Moore, *Annu. Rev. Cell Biol.* **1**, 317 (1985); S. Heinemann *et al.*, in *Nicotinic Acetylcholine Receptor: Structure and Function*, A. Maelicke, Ed. (Springer-Verlag, Berlin, 1986), pp. 359–387.
10. T. Takai *et al.*, *Nature (London)* **315**, 761 (1985).
11. I. Hermans-Borgmeyer *et al.*, *EMBO J.* **5**, 1503 (1986).
12. G. Grenningloh *et al.*, *Nature (London)* **328**, 215 (1987); P. R. Schofield *et al.*, *ibid.*, p. 221.
13. C. J. Brandl and C. M. Deber, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 917 (1986).
14. F. Hucho, W. Oberthur, F. Lottspeich, *FEBS Lett.* **205**, 137 (1986); K. Imoto *et al.*, *Nature (London)* **324**, 670 (1986); J. Giraudat *et al.*, *Biochemistry* **26**, 2410 (1987).
15. M. Ballivet *et al.*, in preparation.
16. P. N. Kao *et al.*, *J. Biol. Chem.* **259**, 11662 (1984).
17. R. D. Marshall, *Biochem. Soc. Symp.* **40**, 17 (1974).
18. J. Patrick and B. Stallcup, *J. Biol. Chem.* **252**, 8629 (1977); S. T. Carbonetto, D. M. Fambrough, K. J. Muller, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1016 (1978).
19. P. M. Ravdin and D. K. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2072 (1979); L. S. Higgins and D. K. Berg, *J. Neurosci.* **7**, 1792 (1987).
20. M. Mishina *et al.*, *Nature (London)* **307**, 604 (1984); *ibid.* **321**, 406 (1986).
21. P. Whiting *et al.*, *ibid.* **327**, 515 (1987); P. Whiting, F. Esch, S. Shimasaki, J. Lindstrom, *FEBS Lett.* **219**, 459 (1987).
22. E. Wada *et al.*, in preparation.
23. F. Sierra, A.-C. Pittet, U. Schibler, *Mol. Cell. Biol.* **6**, 4067 (1986).
24. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
25. F. Sanger, S. Nicklen, A. R. Coulson, *ibid.*, p. 5463; J. Messing, B. Gronenborn, B. Muller-Hill, P. H. Hofschneider, *ibid.*, p. 3642.
26. J. Boulter *et al.*, *J. Neurosci.* **5**, 2545 (1985).
27. J. Finer-Moore and R. M. Stroud, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 155 (1984); H. R. Guy, *Biophys. J.* **45**, 249 (1984).

28. We thank F. Sierra for the rat genomic library; S. Halvorsen and D. Berg for the toxin Bgt3.1; R. Stalder, A. O'Shea, A. Jensen, and K. Evans for their technical assistance; P. Mason and G. Martin for their helpful comments on techniques; and B. Cessna for her help in the preparation of this manuscript. This work was supported by grants from the Swiss National Science Foundation and the National Institutes of Health, a contract from the U.S. Army Medical Research and Development Command, and grants to the Salk Institute from the Noble Foundation and the Weingart Foundation. J.C. is a postdoctoral fellow of the Muscular Dystrophy Association, and E.S.D. is a postdoctoral fellow of the National Institutes of Health.

4 December 1987; accepted 8 February 1988

Exon-Intron Organization in Genes of Earthworm and Vertebrate Globins

SISSY M. JHIANG, JAMES R. GAREY, AUSTEN F. RIGGS

The structure of an invertebrate, intron-containing globin gene has been determined as part of a study of the evolution of hemoglobin. The gene encoding chain *c* of *Lumbricus terrestris* hemoglobin has the two-intron, three-exon structure characteristic of vertebrate globin genes, and the exact positions of the splice junctions are conserved. The two introns interrupting the coding sequence are longer than those of known hemoglobins but shorter than myoglobin introns. The gene encodes a secretory preglobin containing a 16-residue signal peptide, as expected for an extracellular hemoglobin. However, no intron separates the DNA encoding the signal sequence from that of the globin sequence. The 3′ untranslated region of the *Lumbricus* gene is much longer than those of the genes for other hemoglobins and is similar to those found for myoglobins.

THE GENES OF VERTEBRATE GLOBINS are characterized by the presence of three exons separated by two introns (*1*). A third intron, which is predicted to have existed in the ancestral globin gene and to have been lost during animal evolution (*2*), exists in the genes for plant globins (*3–5*). The structures of invertebrate globin genes need to be determined in order to fix the time of loss of the third intron (Fig. 1A). The only invertebrate globin gene characterized to date is that of the insect *Chironomus thummi*, which surprisingly contains no introns (*6*). We have investigated the organization of a gene coding for a chain of earthworm hemoglobin to determine if this gene is similar in intron organization to those of vertebrates, plants, or the insect *Chironomus*.

The extracellular hemoglobin of the earthworm *Lumbricus terrestris* is composed of about 200 polypeptides of six kinds (*7*) that are arranged to form hexagonally shaped molecules with a molecular weight of 3.8 million (*8*). The amino acid sequences (*9, 10*) of the four major globin chains (*a*, *b*, *c*, and *d*) are homologous with those of vertebrates. Two additional chains of unknown structure appear to be necessary for assembly (*11*). The globin chains are synthesized in the chloragogen cells that line the gut (*12, 13*).

A λgt10 complementary DNA (cDNA) library was prepared from polyadenylated RNA of the chloragogen cells. The 1119-bp insert of a clone was sequenced (*14*) and found to encode chain *c*. Comparison with the known protein sequence (*10*) shows the presence of an NH_2-terminal signal peptide of 16 residues. The remainder of the open reading frame corresponds exactly to the sequence of chain *c* determined by protein analysis.

The 3′ untranslated regions of known globin cDNAs range from 88 to 295 bp (*15,*

S. M. Jhiang and A. F. Riggs, Department of Zoology, University of Texas, Austin, TX 78712.
J. R. Garey, Department of Hematology, University of Utah, School of Medicine, Salt Lake City, UT 84132.

16). Those of seal and human myoglobins are 548 and 531 bp (*15, 17*). The 3′ untranslated region of the cDNA for *Lumbricus* chain *c* is 559 bp (*14*), which is closer in length to those of myoglobins than those of hemoglobins and is the longest found for globins examined to date (Table 1). It is not known whether the length of the 3′ untranslated region has any biological significance.

A 4-kb Xba I–Eco RI fragment that contained the gene coding for chain *c* from *Lumbricus* was isolated from a partial genomic library in λgt10. Comparison of the coding region with the sequence of the cDNA showed that the coding sequence of chain *c* is interrupted by two introns of about 1150 and 1250 bp (Fig. 1B). Alignment (*10*) of the amino acid sequence of *Lumbricus* chain *c* with other globins shows that the two introns in the *c* gene are in exactly the same positions as in the genes for vertebrate globins.

The two-intron, three-exon gene organization, which has been strictly conserved in vertebrate hemoglobins and myoglobins, is also conserved in the *Lumbricus* chain *c* gene. This finding suggests that the three-exon, two-intron organization of genes encoding vertebrate globins was established by the time of divergence of annelids and the common ancestor to chordates and echinoderms (*18, 19*). Echinoderms were well established by the early Cambrian period (*18*), some 600 million years ago, and annelids were probably established by this time also, although worm castings and impressions provide the only fossil evidence for their early presence (*19*). This conclusion is consistent with the fact that the genes for vertebrate myoglobins and hemoglobins have the same organization and are believed to have diverged about 600 million years ago (*15*). Therefore, the loss of the third intron must have occurred early in the evolution of invertebrate globin genes, and the loss of all introns from the genes for *Chironomus* globin may be anomalous. We anticipate that introns will be found in globin genes of other noninsect invertebrates and perhaps in those of other insects as well.

Table 1. Lengths (in base pairs) of noncoding regions of genes for various globins. NT, untranslated messenger RNA sequence.

Gene	5′ NT	Intron			3′ NT	Reference
		1	Central	2		
Globin						
Lumbricus		1150		1250	559	
Vertebrate α	34–40	108–171		103–338	95–109	(*15*)
Vertebrate β	45–53	116–192		573–906	88–132	(*15*)
Plant	49	119–169	99–334	159–680	144–174	(*4, 15*)
Myoglobin						
Seal	70–72	4800		3400	548	(*15*)
Human	70	5800		3600	531	(*17*)

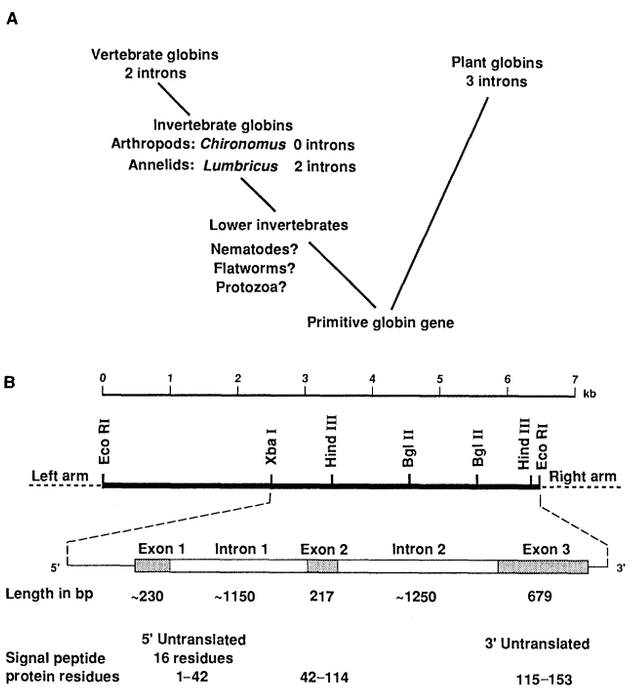

Fig. 1. (**A**) Distribution of the number of introns in genes of globins of vertebrates, invertebrates, and plants. (**B**) The organization of the gene of chain *c* of *Lumbricus* globin deduced from DNA sequencing and mapping of restriction sites. Polyadenylated RNA from chloragogen cells was the template for double-stranded cDNA synthesis with oligo (dT) as primer (*25*). Eco RI linker-ligated double-stranded cDNA was cloned into the Eco RI–digested λgt10. An oligonucleotide-primed, ^{32}P-labeled, single-stranded cDNA was used for screening. Eight clones that showed strong hybridization signals were plaque-purified. DNA was isolated from each clone and digested with Eco RI to release inserts, which were subcloned into Bluescript (Stratagene Cloning Systems, San Diego, California). The cDNA inserts were sequenced from double-stranded templates with T3 and T7 primers (Stratagene) and by the use of specific primers (*26*). A Southern blot of Eco RI–digested genomic DNA prepared from *Lumbricus* sperm was hybridized with a ^{32}P-labeled RNA probe derived from a cDNA clone for chain *c*. Four bands, 5 to 10 kb in size, were found. These fragments were electroeluted and cloned into λgt10. The ^{32}P-labeled RNA transcripts derived from the cDNA clone were used to screen the partial genomic library and to identify a globin gene. The restriction map of the 6.5-kb genomic DNA insert was established by single and double digestion. The entire fragment was sequenced except intron 2, 120 bp of intron 1, and 200 bp upstream from the TATA box, which is 45 bp upstream from the ATG start signal. The first intron starts before the G of the AAG codon for Lys at position 42 of the protein sequence. The second intron starts after the AAG codon for Lys at position 114.

The length of the introns in genes of vertebrate globins is relatively constant (Table 1). The *Lumbricus* chain *c* gene has introns longer than those of vertebrate hemoglobin genes but shorter than those found in myoglobin genes. The intermediate intron length and the long 3′ untranslated region of the *Lumbricus* gene suggest an evolutionary relation between earthworm hemoglobin and vertebrate myoglobins. Protein analysis shows that invertebrate globins are more diverse than vertebrate globins. The invertebrate globins differ from vertebrate globins more than vertebrate myoglobins differ from vertebrate hemoglobins (*21*). However, the time of divergence of vertebrate myoglobins and invertebrate globins is uncertain. Our data suggest that the time of the origin of annelids and the

Why have the positions of intron junctions in animal globins been so highly conserved? Intron-exon splice junctions tend to map at protein surfaces (*20*), which means that insertions and deletions can be accommodated without structural damage. However, the junctions map within helices at the surface of subunits in tetramers of vertebrate hemoglobins, but not at the surfaces of the tetramers. The splice junctions are in helices that form critical intersubunit contacts in the tetramer. Any insertions or deletions associated with "sliding" of these splice junctions in globin genes would be expected to disrupt the helices and modify the subunit contacts and should therefore be selected against. Comparisons among eight different globin chains from annelids (*10*) indicate that insertions and deletions probably have occurred exclusively in regions between helices or at their ends and not in the neighborhood of splice junctions, emphasizing the conservation of these regions.

time of vertebrate myoglobin divergence were close.

Many eukaryotic genes that code for secretory proteins have introns between the signal peptide coding region and that of the mature protein (*22*). Neither the *Lumbricus* nor the *Chironomus* globin gene has an intron in this position. There is speculation that the absence of this intron in the *Chironomus* globin genes indicates that the total loss of all introns in *Chironomus* globins was caused by a single conversion event rather than by separate stochastic events (*23*). The absence of this intron in the *Lumbricus* chain *c* gene suggests that the lack may be a common feature shared by other genes of extracellular invertebrate hemoglobins.

Our results suggest that the third (central) intron of globin genes was lost at least 600 million years ago. A third intron should be sought in the genes for globins of lower invertebrates, particularly those of nematodes, flatworms, and protozoa (Fig. 1A). All these phyla have members with hemoglobins (*24*). It is possible that a third intron may not be found at all in the invertebrates, because comparison of animal and plant globins suggests a possible common ancestor 1500 million years ago (*4*).

REFERENCES AND NOTES

1. T. Maniatis, E. F. Fritsch, J. Aner, R. M. Lawn, *Annu. Rev. Genet.* **14**, 145 (1980).
2. M. Gō, *Nature (London)* **291**, 90 (1981).
3. E. Ø. Jensen, K. Paludan, J. J. Hyldig-Nielsen, P. Jørgensen, K. A. Marcker, *ibid.*, p. 677.
4. J. Landsmann *et al.*, *ibid.* **324**, 166 (1986).
5. D. Bogusz *et al.*, *ibid.* **331**, 178 (1988).
6. M. Antoine and J. Niessing, *ibid.* **310**, 795 (1984).
7. J. M. Shlom and S. N. Vinogradov, *J. Biol. Chem.* **248**, 7904 (1973).
8. O. H. Kapp, S. N. Vinogradov, M. Ohtsuki, A. V. Crewe, *Biochim. Biophys. Acta* **704**, 546 (1982).
9. F. Shishikura, J. W. Snow, T. Gotoh, S. N. Vinogradov, D. A. Walz, *J. Biol. Chem.* **262**, 3123 (1987).
10. K. Fushitani, M. S. A. Matsuura, A. F. Riggs, *J. Biol. Chem.*, in press.
11. S. N. Vinogradov, S. D. Lugo, M. G. Mainwaring, O. H. Kapp, A. V. Crewe, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 8034 (1986).
12. E. Lindner, *Z. Zellforsch.* **66**, 891 (1965).
13. S. M. Jhiang, J. R. Garey, A. F. Riggs, in *Invertebrate Oxygen Carriers*, B. Linzen, Ed. (Springer-Verlag, Berlin, 1986), pp. 469–472.
14. The sequence has been deposited in GenBank, access number J03082 (GenBank, Los Alamos National Laboratory, Los Alamos, NM 87545).
15. A. Blanchetot, V. Wilson, D. Wood, A. J. Jeffreys, *Nature (London)* **301**, 732 (1983).
16. J. R. Garey and A. F. Riggs, *J. Biol. Chem.* **261**, 16446 (1986).
17. P. Weller, A. J. Jeffreys, V. Wilson, A. Blanchetot, *EMBO J.* **3**, 439 (1984).
18. D. Nichols. *Symp. Zool. Soc. London* **20**, 209 (1967).
19. S. C. Morris, in *The Origins and Relationships of Lower Invertebrates*, S. C. Morris, J. D. George, R. Gibson, H. M. Platt, Eds. (Oxford Univ. Press, Oxford, 1985), pp. 350–352.
20. C. S. Craik, S. Sprang, R. Fletterick, W. J. Rutter, *Nature (London)* **299**, 180 (1982).
21. R. E. Dickerson and I. Geis, in *Hemoglobin: Structure, Function, Evolution, and Pathology* (Benjamin-Cummings, Menlo Park, CA, 1983), pp. 76–82.
22. R. Breathnach and P. Chambon, *Annu. Rev. Biochem.* **50**, 349 (1981).
23. R. Lewin, *Science* **226**, 328 (1984).
24. S. N. Vinogradov, *Comp. Biochem. Physiol.* **82B**, 1 (1985).
25. T. V. Huynh, R. A. Young, R. W. Davis, in *DNA Cloning: A Practical Approach*, D. Glover, Ed. (IRL Press, Oxford, 1985), vol. 1, pp. 49–78.
26. E. C. Strauss, J. A. Kobori, G. Siu, L. E. Hood, *Anal. Biochem.* **154**, 353 (1986).

27. We thank K. Fushitani for providing the amino acid sequences of the chains of *Lumbricus* globin prior to publication and G. M. Pfaffenbach for helpful discussion. Supported by NSF grant DMB-8502857, Welch Foundation grant F-213, and NIH grant GM28410.

15 December 1987; accepted 2 March 1988

Oral *Salmonella typhimurium* Vaccine Expressing Circumsporozoite Protein Protects Against Malaria

JERALD C. SADOFF, W. RIPLEY BALLOU, LOUIS S. BARON, WILLIAM R. MAJARIAN, ROBERT N. BREY, WAYNE T. HOCKMEYER, JAMES F. YOUNG, STANLEY J. CRYZ, JONATHAN OU, GEORGE H. LOWELL, JEFFREY D. CHULAY

Immunization with radiation-attenuated malaria sporozoites induces potent cellular immune responses, but the target antigens are unknown and have not previously been elicited by subunit vaccines prepared from the circumsporozoite (CS) protein. A method is described here for inducing protective cell-mediated immunity to sporozoites by immunization with attenuated *Salmonella typhimurium* transformed with the *Plasmodium berghei* CS gene. These transformants constitutively express CS antigens and, when used to immunize mice orally, colonize the liver, induce antigen-specific cell-mediated immunity, and protect mice against sporozoite challenge in the absence of antisporozoite antibodies. These data indicate that the CS protein contains T cell epitopes capable of inducing protective cell-mediated immunity, and emphasize the importance of proper antigen presentation in generating this response. Analogous, orally administered vaccines against human malaria might be feasible.

ANIMALS AND HUMANS CAN BE PROtected against malaria by immunization with radiation-attenuated sporozoites. The basis of this immunity is complex and includes a humoral immune response to immunodominant repeat epitopes on the circumsporozoite (CS) protein, and cell-mediated immune responses to sporozoite or exoerythrocytic antigens (*1*). We previously reported that protective immunity induced by irradiated *Plasmodium berghei* sporozoites is mediated by antigen-specific T cells (*2*). These T cell populations may participate in cytotoxic or lymphokine-mediated killing of exoerythrocytic parasites in the liver (*3*). Unlike live sporozoites (*4*), killed sporozoites do not invade cells and are unable to induce protective immunity (*5*). Intracellular targeting of CS antigens may be necessary for induction of protective cell-mediated immune responses. We therefore investigated the use of attenuated *Salmonella* as vectors to present CS antigen.

Soon after oral inoculation, attenuated *Salmonella* vaccine strains are found in regional and systemic lymphatics where they are ingested by macrophages (*6*). Immunity induced by these strains is cell-mediated and its induction requires a period of intracellular survival by the organism (*7–9*). We used a plasmid containing the *P. berghei* CS protein gene to transform *S. typhimurium* WR4017, an avirulent strain with impaired ability to multiply within macrophages (*10*). This transformant vaccine constitutively expressed CS antigen, and induced antigen-specific cell-mediated immune responses in mice and protection from sporozoite challenge in the absence of measurable antibodies to CS protein. These studies identify the CS protein as a target antigen for protective cell-mediated immunity against sporozoites and demonstrate the feasibility of constructing analogous oral vaccines against human malarias.

The transformant (WR4017/pMGB2) was constructed with an expression plasmid containing the entire *P. berghei* CS protein coding region (Fig. 1). Female BALB/c mice (6 to 8 weeks old) were immunized by subcutaneous (sc) injection of 10^4 bacteria or by oral administration of 10^9 or 10^{10}

J. C. Sadoff, W. R. Ballou, L. S. Baron, J. Ou, G. H. Lowell, J. D. Chulay, Departments of Bacterial Diseases, Immunology, and Bacterial Immunology, Walter Reed Army Institute of Research, Washington, DC 20307-5100.
W. R. Majarian, R. N. Brey, W. T. Hockmeyer, Praxis Biologics, Rochester, NY 14623–1943.
J. F. Young, Molecular Genetics, Smith Kline and French Laboratories, Swedeland, PA 19406–2799.
S. J. Cryz, Swiss Serum and Vaccine Institute, Berne, Switzerland.