Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 6486–6490, September 1987
Evolution

Early genes that were oligomeric repeats generated a number of divergent domains on their own

(acetylcholine receptor/domain exchange/oligomeric repeats/rhodopsin family)

SUSUMU OHNO

Beckman Research Institute of the City of Hope, Duarte, CA 91010

Contributed by Susumu Ohno, June 1, 1987

ABSTRACT **One of the more popular concepts to emerge in recent years is that new proteins evolved by domain exchanges between preexisting proteins. The presence of introns within eukaryotic genes is thought to enhance such exchanges. Yet domain exchanges must necessarily be the secondarily developed process in evolution, for they would have been effective only after multitudes of domains came into being. Many of the proteins with functionally divergent domains were established before the division of prokaryotes from eukaryotes; i.e., soon after the creation of life on this earth. I attribute the extreme innovativeness of early coding sequences to their construction; i.e., being repeats of oligomeric units. The rhodopsin family of proteins—with seven hydrophobic, α-helical transmembrane domains, four extracellular domains, and four intracytoplasmic domains—indeed arose before the division of prokaryotes from eukaryotes and later gave rise to muscarinic acetylcholine receptor and β-adrenergic receptor among others. In this paper, I show that the entire coding sequence for porcine muscarinic acetylcholine receptor is still replete with copies of three heptameric units that are very closely related to each other. Original heptameric units are more stringently conserved in parts encoding the seven transmembrane domains, whereas new repeating units are comingled with the old in parts encoding extracellular and intracytoplasmic domains.**

A popular concept in molecular evolution is that of evolution by domain exchanges. Intervening sequences that interrupt eukaryotic coding sequences are thought of as promoters of such domain exchanges between functionally unrelated genes (1). Indeed, a good case for evolution by domain exchange can be made for calcium-dependent protease (2). The chicken calcium-dependent protease is 705 residues long. Its amino-terminal portion is a typical thiol protease with two active sites, the first composed of Asp-Cys-Trp (residues 107–109) and the second of Gly-His-Ala (residues 224–226). Its carboxyl-terminal 150 residues or so, on the other hand, show significant homology with other calcium-binding proteins such as calmodulin and troponin C. This region contains at least three and possibly four calcium-binding regions, each 12 residues long. Thus, it appears that the calcium-dependent protease was a simple thiol protease that became calcium-dependent by a domain exchange with a calcium-binding protein.

At the extreme, one of the coated-pit receptors for cholesterol-transporting low density lipoproteins (LDL) appears to have no ancestor as such, for the suggestion was made that it has evolved exclusively by assembling borrowed domains (3). The ligand (LDL) binding site in the amino-terminal region of this 839-residue receptor comprises seven tandem repeats of a 40-residue unit in which six invariant cysteine residues are very prominent. This unit is said to have been derived from residues 77–113 of complement factor C9, which is 537 residues long. The remainder of the LDL receptor is said to be homologous with epidermal growth factor (EGF) precursor. Implicit in the "borrowed-domain" explanation of the origin of LDL receptor is the assumption that within C9, the region represented by residues 77–113 is a unique domain unlike any other parts of C9. Indeed, the concentration of six cysteine residues within so short a segment is not seen elsewhere in C9 (4). It would thus appear that whereas the LDL receptor had to borrow this domain from C9, C9 itself managed to create this domain on its own. The inevitable conclusion to be drawn is that coding sequences were more innovative at a certain time in evolution than at other times.

I have argued (5) that the first set of coding sequences that arose in the prebiotic world were repeats of oligomers, the numbers of bases in these oligomeric units not being multiples of 3. There are two main reasons for the above argument. First, elongation after each round of nucleic acid replication is an inherent property of oligomeric repeats, so that only oligomeric repeats could have attained a size worthy of coding sequences in the prebiotic world. Second, these oligomeric repeats are endowed with three open reading frames, all encoding polypeptide chains of the identical periodicity. Thus, they are not readily silenced by a high error rate of nonenzymatic nucleic acid replication; in the presence of Zn^{2+}, the error rate is estimated as 10^{-2} per base pair replicated (6). It seems as though coding sequences were at their innovative best at the very beginning of life before the division of prokaryotes from eukaryotes, for many of the prototype proteins were already established at the time of that division. This is not only true of all the sugar-metabolizing enzymes but also of hemoglobins; note the presence of hemoglobins in animals and plants as well as in bacteria. I shall now show that one kind of oligomeric repeat has evolved to generate a number of divergent domains.

One Protein with Seven Transmembrane Domains That Gave Rise to Bacterial Rhodopsin, Retinal Opsin, β_2-Adrenergic Receptor, and Muscarinic Acetylcholine Receptor Was Originally Encoded by Repeats of a Base Heptamer or Heptamers. One of the prototype proteins that subsequently has shown remarkable functional versatility was a protein with seven transmembrane domains, for among its descendants are bacterial rhodopsin, bakers' yeast mating factor receptor, vertebrate retinal opsin, mammalian β_2-adrenergic receptor, and mammalian muscarinic acetylcholine receptor (reviewed in ref. 7). As the basic construction is the same in all these proteins, I shall now concentrate on the porcine muscarinic acetylcholine receptor (7). Seven transmembrane domains, each about 23 residues long, numbered I–VII in Fig. 1, can be considered as one kind of domain, although they need not have arisen by tandem duplication, as we shall see

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: EGF, epidermal growth factor; LDL, low density lipoprotein.

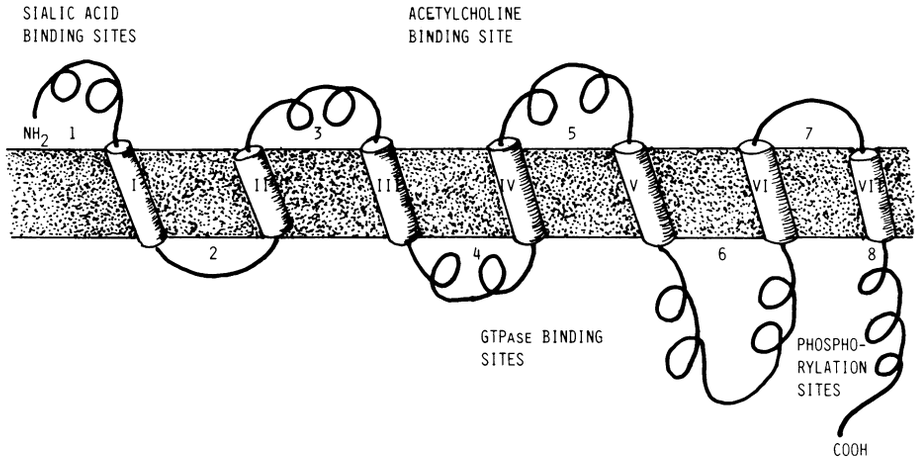

FIG. 1. Two-dimensional schematic demonstration of the functionally very versatile family of proteins that includes the muscarinic acetylcholine receptor as well as bacterial rhodopsin, yeast mating factor receptor, retinal rhodopsin, and β_2-adrenergic receptor. Seven hydrophobic α-helical transmembrane segments are numbered I–VII, while the amino-terminal, interconnecting, and carboxyl-terminal segments are numbered 1–8. Domains 1, 3, 5, and 7 protrude outside the plasma membrane, whereas 2, 4, 6, and 8 remain inside the cytoplasm. This numbering system is utilized throughout the text as well as in Figs. 2–6.

shortly. But what of the six interconnecting domains and the amino-terminal and carboxyl-terminal domains? Four of these domains, labeled 1, 3, 5, and 7 in Fig. 1, protrude outside the plasma membrane, while the other four (domains 2, 4, 6, and 8) remain in the cytoplasm. Of the four domains that protrude outside, the amino-terminal domain (domain 1) must necessarily have N-glycosylation sites for sialic acids and other sugars, and indeed there are two asparagine residues in this 23-residue domain. One or more of the remaining three external domains must supply the binding site for the ligand, acetylcholine. Of the four cytoplasmic domains, domains 4 and 6 are thought to be involved in binding to GTPases, while the carboxyl-terminal domain (domain 8) is thought to be phosphorylated via serine and/or threonine residues. It thus appears that the muscarinic acetylcholine receptor and its allies are composed of several different types of domains. However, the coding sequence for this 460-residue porcine muscarinic acetylcholine receptor is replete with recurring base oligomers (Fig. 2).

The 92-codon segment of the porcine muscarinic acetylcholine receptor coding sequence (8) shown in Fig. 2 starts from transmembrane domain II followed by interconnecting domain 3, which protrudes outside the plasma membrane. Therefore, this domain might be endowed with a portion of the acetylcholine binding site. Transmembrane segment III succeeds the above and is followed by interconnecting domain 4, which remains inside the cytoplasm and is thought to contain a portion of the GTPase binding site. The remainder represents roughly half of transmembrane segment IV. Recognizable copies of three primordial base heptamers that are very closely related to each other comprise 70% of this 276-base coding segment. Two invariant copies of primordial heptamer CCTGCTG are underlined by the thickest open bars in Fig. 2 (first row and fourth row), while its three single-base-substituted copies, one doubly substituted copy, and two triply substituted copies are identified by underlining with progressively thinner open bars. Invariant and singly to triply substituted copies of two other primordial heptamers—GCTGGCC and CCTGGCC—are identified similarly. Two invariant copies of GCTGGCC, both seen in the third row of

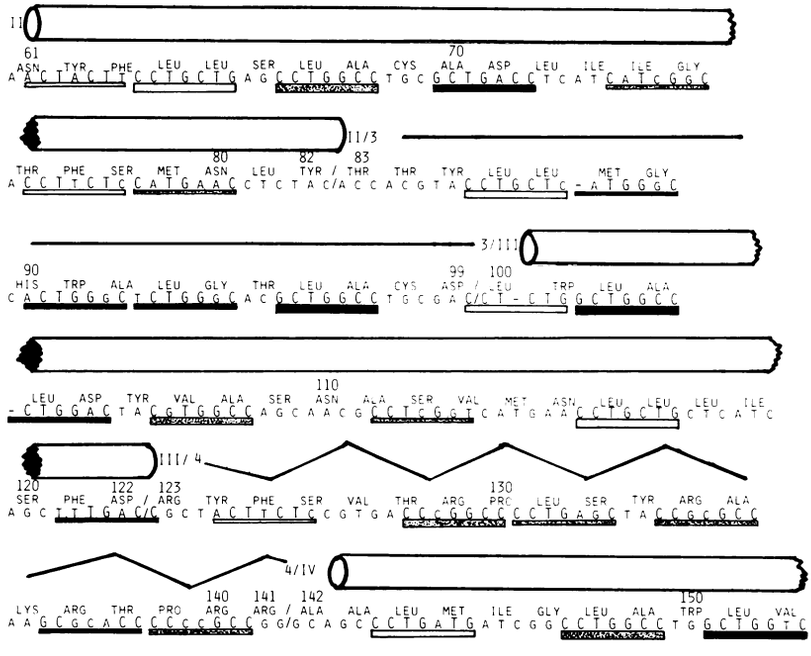

FIG. 2. Nucleotide sequence encoding domains II, 3, III, and 4 and a portion of domain IV of porcine muscarinic acetylcholine receptor; corresponding amino acid residues are shown above the coding sequence. Transmembrane domains are identified by cylinders, the extracellular interconnecting domain by a straight line, and the intracellular interconnecting domain by a zigzag line. The three primordial heptamers CCTGCTG, CCTGGCC, and GCTGGCC, each occurring twice in this 276-base-long segment, are shown in large uppercase letters underlined by the thickest open, stippled, and solid bars, respectively. The corresponding singly, doubly, and triply substituted copies are identified by progressively thinner bars of proper shading.

Fig. 2, are underlined by the thickest solid bars, whereas two invariant copies of the heptamer CCTGGCC, seen in the first row and last row of Fig. 2, are underlined by the thickest stippled bars. Thus, it would appear that the coding segment that originally consisted of repeats of the primordial base heptamer or heptamers managed to generate at least three different kinds of domains: the α-helical transmembrane domain and the extracellular and intracytoplasmic interconnecting domains. The three primordial heptamers identified in Fig. 2 are clearly related to each other. It follows then that one of the three must have been the ultimate ancestor of the other two, but the decision on the ultimate ancestor should be postponed until more sequences of this functionally divergent family from bacteria to man become known.

While 70% of the coding sequence shown in Fig. 2 is readily identified as copies of the three primordial heptamers, what of the remaining 30% that is shown as gaps between the heptamer copies? These gaps are essentially of four different kinds (Fig. 3). The 7-base-long gap seen at the junction between interconnecting segment 4 and transmembrane segment IV (Fig. 2, last row) exemplifies a very frequently encountered situation as shown in Fig. 3*a*. The GGGCA portion of the gap actually is a portion of a doubly substituted copy (CCGGGCA) of the stippled primordial heptamer (CCTGGCC), but the first two cytosines were already incorporated as members of the preceding copy. Thus, the bulk of the gap very often represents a truncated portion of a hexamer copy. As often, gaps are occupied by very degenerate copies of one of the three primordial heptamers that became independent recurring units. As shown in Fig. 3*b*,the sequence Leu-Ile occupying two gaps in Fig. 2 is encoded by the recurring heptamer CTCATCA. Its CTCATC portion must have been derived from CTCCTC, a doubly substituted deviant of the CTGCTG portion of the open primordial heptamer (CCTGCTC). The situation depicted in Fig. 3*c* is also commonly encountered. Often primordial heptamers—solid and stippled heptamers in this instance—became parts of the longer recurring unit CTGGCCTGCG. It can be seen that its tail-end TGCG portion supplied the bulk of two gaps, and its truncated copy, yet a third gap. The situation represented in Fig. 3*d* is actually a composite of the three situations shown in *a*–*c*. In addition, however, the situation in *d* also justifies the identification of three-base-substituted copies of each primordial heptamer as done in Fig. 2. CATGAAC is a triply substituted copy of the stippled heptamer. Nevertheless, it appears that it became a part of the longer, nonameric recurring unit CATGAACCT, its CATGAA portion supplying a bulk of one gap merely because its CCT portion was incorporated as a part of the open primordial heptamer. CTTCTCCATGAAC also became a new, tridecameric recurring unit, with the CGTGA portion of its doubly substituted copy becoming one gap. All in all, it would appear that gaps are no indication of intrusions by alien sequences but merely represent secondary derivatives of the primordial repeating units.

The amino acid sequences of the eight interconnecting domains are not as conserved as those of the seven transmembrane domains among members of this family, the maximization of sequence homology requiring the introduction of considerable stretches of deletions and insertions (as reviewed in ref. 7). It appears that functional diversity among members of this family is more evident in interconnecting domains than in transmembrane domains. Indeed, interconnecting domain 6 of the porcine acetylcholine receptor is roughly 100 residues longer than the corresponding domain of other members of the family; roughly, residues 230–326 of the acetylcholine receptor have no counterpart. At first glance, this 100-residue segment, rich in proline, arginine, glutamic acid, and cysteine, appears to be an insert conceivably borrowed by a domain exchange from the sodium-channel domain of another protein. However, Fig. 4*a* shows that this apparent insert is not an insert at all, but merely an extension of the adjacent coding segments, as this region too is composed essentially of copies of the same three primordial heptamers. Particularly noteworthy is the tandem recurrence of GCTGCTG, which is a single-base-substituted copy of the open primordial heptamer. This tandem recurrence is translated to yield Arg-Cys-Cys-Arg-Cys-Cys, as shown in the first row of Fig. 4*a*. The only truly "alien" segment is four successive copies of the base trimer GAG, encoding Glu-Glu-Glu-Glu. The carboxyl-terminal cytoplasmic domain 8 contains two serine residues and one threonine residue that are thought to be phosphorylatable. Inasmuch as phosphorylation of these residues is thought to be involved in the

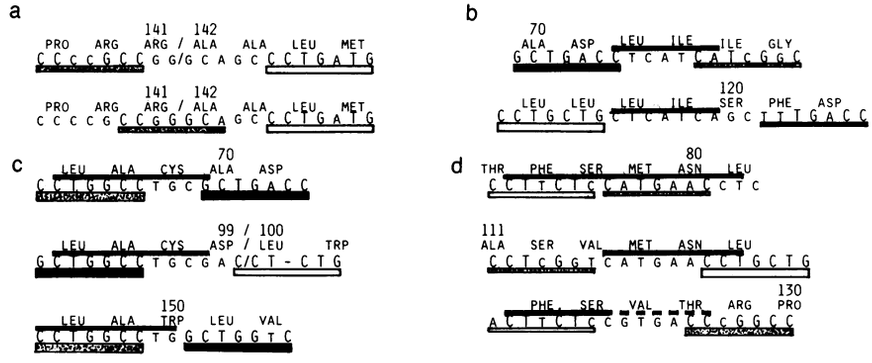

FIG. 3. Seventeen gaps are seen between neighboring copies of primordial heptamers in Fig. 2; these gaps account for ≈30% of the 92-codon-long coding segment. While six of these gaps are only 2 bases long, the longest gap, at the junction between domains II and 3, comprises 14 bases. These gaps are essentially of four different kinds as indicated in *a*–*d*. (*a*) GGGCA portion of the 7-base gap at the 4/IV junction actually is a part of a doubly substituted copy (CCGGGCA) of the stippled primordial heptamer, but the first two cytosines are already incorporated into the preceding copy. (*b*) During evolution, some of the very degenerate copies of the primordial heptamers established themselves as new recurring units. Gaps are often represented by such new recurring units. The bulk of two gaps shown here is represented by the new recurring heptamer CTCATCA, encoding Leu-Ile, which is a very degenerate derivative of the open primordial heptamer. (*c*) The primordial heptamer often became a part of a longer repeating unit. In such an instance, the tail end of the longer repeating unit becomes a gap. The last 4 bases of the decameric recurring unit CTGGCCTGCG are seen in two gaps, while the last two bases of the related nonameric recurring unit CCTGGCCTG comprise yet a third gap. (*d*) The situation here is actually a composite of *a*–*c*. The heptamer CATGAAC is a triply substituted copy of the stippled primordial heptamer. In the first line, it is identified as such. In the second line, however, 6 bases of it are regarded as a gap merely because the last cytosine is incorporated as the first cytosine of the subsequent open primordial heptamer. Meanwhile, CATGAAC became a part of the tridecameric recurring unit CTTCTCCATGAAC. The CGTGA portion of its doubly substituted copy CTTCTCCGTGACC is regarded as a gap in line 3.

down-regulation of the receptor, domain 8 might be considered as a functionally independent domain. Yet, as shown in Fig. 4*b*, domain 8 too is encoded by a sequence mostly composed of substituted variants of the three primordial heptamers. All in all, there remains little doubt that the ancestral coding sequence that comprised repeats of the three primordial heptamers has managed to generate not only seven α-helical transmembrane domains but also eight interconnecting domains that subsequently acquired independent and diverse functions.

Domain Differentiation Has Been Accomplished by Very Subtle Local Changes in the Coding Sequence. Observing Figs. 2 and 4, note that copies of the three primordial heptamers are more or less evenly distributed throughout the entire coding sequence for the porcine acetylcholine receptor. It follows that sequence differences between functionally divergent segments must necessarily be rather subtle. Yet there is one obvious difference that distinguishes the seven transmembrane domains from the eight interconnecting domains. Invariant and single-base-substituted copies of the three primordial heptamers recur far more frequently in the former than in the latter. Altogether there are 6 invariant and 13 single-base-substituted copies of the three primordial heptamers in seven transmembrane domain-encoding segments comprised of 155 codons, thus accounting for 29% of the 465 bases, whereas there are only 5 invariant and 17 single-base-substituted copies in the remainder representing 305 codons. Doubly and triply substituted copies of the primordial heptamers, on the other hand, are more evenly distributed, being roughly twice as numerous in the sequence encoding the eight interconnecting domains. It thus appears that the original coding sequence was more conserved in segments encoding transmembrane domains.

The manner in which evolutionarily invariant residues have been encoded is also very instructive. CCTGGAC at the left of the fourth row of Fig. 2 is a single-base-substituted copy of the stippled primordial heptamer, and at this position, it is translated in the most frequently used reading frame to yield Leu-Asp in transmembrane domain III. Yet, the same heptamer in the segment encoding transmembrane domain VI is translated in a different reading frame to generate Thr-Trp-Thr, as shown in Fig. 5*a*. In fact, this tryptophan is one of the four tryptophan residues that remain invariant in the porcine muscarinic acetylcholine receptor (7), hamster β-adrenergic receptor (8), and bovine retinal opsin (9). Fig. 5*a* shows that all the four evolutionarily invariant tryptophan residues are encoded by singly or doubly substituted copies of the stippled primordial heptamer translated in the second-choice reading frame. A reading-frame choice in translating copies of the primordial heptamers no doubt contributed to local differentiation of the coding sequence.

On occasion, the emergence of new repeating units derived from degenerate copies of the primordial repeating units appears to have contributed to local differentiation of coding sequences. Two examples given in Fig. 3 *b* and *d*, CTCATCA and CATGAACCT, recurred only in segments encoding transmembrane domains. The undecamer CCGGGGAAGGG, on the other hand, recurred only in interconnecting segments encoding Pro-Gly-Lys-Gly of domains 1 and 6. Domain 1, however, protrudes outside the plasma membrane, whereas domain 6 is intracytoplasmic (Fig. 5*b*). The octamer GGTCAACA recurred three times (two invariant copies and one doubly substituted copy) only within the 12-codon segment encoding interconnecting domain 2 (Fig. 5*c*). In this instance, there is little doubt that the uniqueness of intracytoplasmic domain 2 is supplied by recurrence of this new octameric unit, which is but a very degenerate copy of the solid primordial heptamer GCTGGCC.

Sequence Homology Among Functionally Diversified Domains. Inasmuch as the entire coding sequence for the porcine muscarinic acetylcholine receptor has apparently descended from repeats of the three primordial heptamers, and since the true evidence of local sequence differentiation was scarce as just noted, amino acid sequences of functionally diversified domains should not be so different as commonly imagined. The recurrence of Pro-Gly-Lys-Gly in interconnecting domains 1 and 6 has already been recorded in Fig. 5. I shall end this paper with further evidence of amino acid sequence homology between functionally divergent domains. Amino acid sequence comparison between porcine acetylcholine receptor (7) and hamster β-adrenergic receptor (8) revealed the longest stretch of conserved residues to reside in the amino-terminal half of transmembrane domain II. This Ser-Leu-Ala-Cys-Ala-Asp-Leu heptapeptide sequence is also conserved in the corresponding region of bovine retinal opsin, except for two substitutions: asparagine

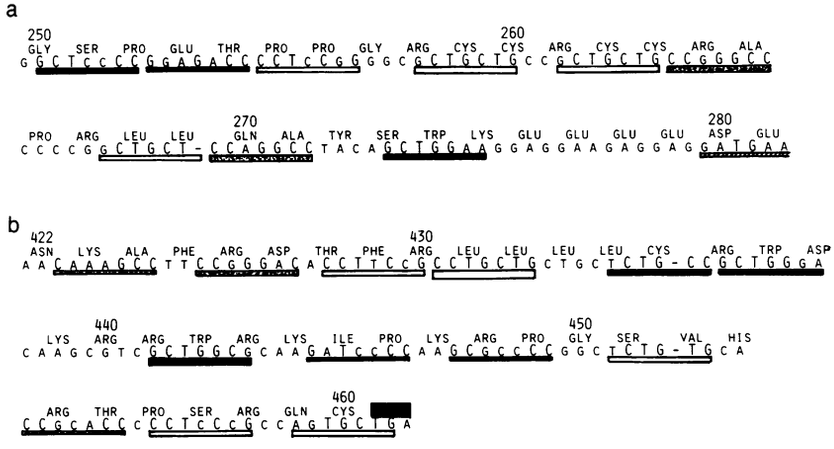

FIG. 4. The coding segment chosen in Fig. 2 is not an exception but quite representative of the entire coding sequence for porcine muscarinic acetylcholine receptor. This is shown in two additional, rather unusual coding segments. (*a*) Interconnecting cytoplasmic domain 6 is ≈100 residues longer than domain 6 of other members of this family. Therefore, residues 230–329 of this receptor have no counterpart in either hamster β-adrenergic receptor or in bovine opsin (reviewed in ref. 7), and this region is rich in proline, arginine, glutamic acid, and cysteine. At first glance, this suggests the intrusive insertion of an alien element via a domain exchange. Yet as shown in the first two rows, this coding segment too is very rich in copies of the three primordial heptamers. The only unusual feature here is a tandem duplication of a GAG trimer, thus resulting in four successive glutamic acid residues. (*b*) The carboxyl-terminal domain 8 is functionally independent, for phosphorylation of its threonine and serine residues is thought to down-regulate the receptor. Yet the coding segment for it is as rich in copies of the three primordial heptamers as other parts of the coding sequence, as shown in the last three rows.

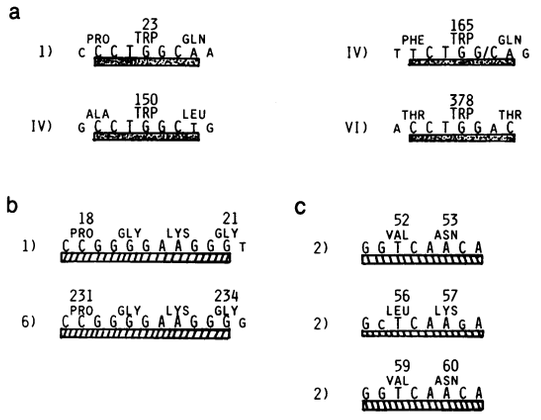

FIG. 5. (*a*) Four evolutionarily invariant tryptophan residues in the acetylcholine receptor. These residues have been conserved by all members of the family. In every instance, the tryptophan residue is encoded by a part of the stippled primordial heptamer—singly substituted copies in three instances, and a doubly substituted copy in one. More often, these heptameric units are translated to yield Leu-Ala and Leu-Asp, but in these instances, a different reading frame is utilized to encode tryptophan residues. Reading-frame choice in translation of heptameric units no doubt contributed to domain differentiation. The emergence of new recurring units derived from very degenerate copies of the primordial heptamers also contributed to domain differentiation. Since two examples are already given in Fig. 3, only two more are given. (*b*) The undecameric sequence shown was derived from the stippled or solid primordial heptamer. This new recurring sequence gives Pro-Gly-Lys-Gly tetrapeptide to interconnecting domains 1 and 6, thus setting them apart from all other domains. (*c*) The uniqueness of the shortest interconnecting domain (domain 2), on the other hand, is found in the tandem recurrence of a new octameric unit. Two invariant copies and one doubly substituted copy of GGTCAACA encode 6 of the 12 amino acid residues of domain 2.

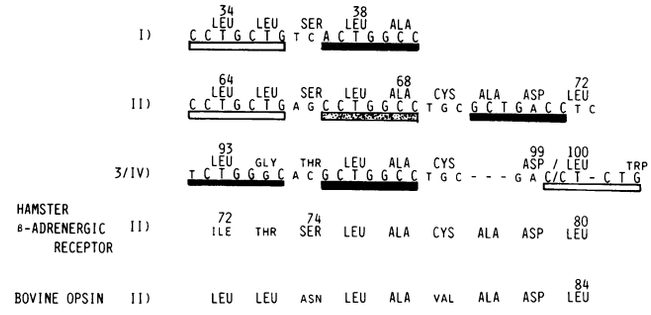

FIG. 6. Wherever invariant and singly substituted copies of the primordial heptamers congregate, similar or identical amino acid sequences are expected to occur totally independent of domain differentiation. One such instance involving the most evolutionarily conserved amino acid sequence segment is shown here. Residues 64–72 in transmembrane domain II represent one of the more evolutionarily conserved segments, the heptapeptide sequence Ser-Leu-Ala-Cys-Ala-Asp-Leu (residues 66–72) being the longest stretch of invariant residues shared between the porcine muscarinic acetylcholine receptor and the hamster β-adrenergic receptor. Yet, residues 95–100 at the end of interconnecting domain 3 of the acetylcholine receptor differ from the above heptapeptide only by a single substitution and a single deletion, as shown. Residues 64–68 in domain II have also been conserved, the corresponding pentapeptide in bovine opsin differing from the Leu-Leu-Ser-Leu-Ala pentapeptide only by a single substitution, as shown. The above pentapeptide sequence, however, is also found in transmembrane domain I.

for serine and valine for cysteine (ref. 9 and Fig. 6). The segment coding for these evolutionarily conserved residues, already shown in Fig. 2, is recapitulated in Fig. 6. Note that the stippled primordial heptamer CCTGGCC and a singly substituted copy (GCTGACC) of the solid primordial heptamer contribute four of the seven codons to this segment. Not surprisingly, a sequence very similar to the above is found near the end of the sequence encoding interconnecting domain 3, which protrudes outside the plasma membrane. This sequence, also already shown in Fig. 2, is again recapitulated in Fig. 6 for comparison. The hexapeptide Thr-Leu-Ala-Cys-Asp-Leu (residues 95–100) differs from the above noted conserved heptapeptide sequence only by a substitution and a deletion: serine is replaced by threonine, and the alanine following cysteine has been deleted. In addition, in transmembrane domain II of the acetylcholine receptor, the conserved heptapeptide is preceded by Leu-Leu; the resulting pentapeptide Leu-Leu-Ser-Leu-Ala is conserved in the corresponding position of bovine opsin, except for a substitution of asparagine for serine, but not in hamster β-adrenergic receptor (Fig. 6). This pentapeptide sequence recurs in positions 34–39 of transmembrane segment I (Fig. 6). The situation depicted in Fig. 6 is but one of a number of examples. Wherever relatively unadulterated copies of the three primordial heptamers congregate, they tend to encode homologous oligopeptide sequences regardless of to which domain they belong. Because members of this family of proteins are uniformly endowed with seven hydrophobic α-helical transmembrane segments of similar size (22 residues on the average), it has been the common assumption that the ancestral gene for this family arose by tandem seven-times duplication of a small primordial unit roughly 50 codons long. The present analysis shows that there is no need to evoke such a large internal duplication. A single primordial coding sequence about 400 codons long, composed of repeats of the primordial heptamer (one of the three presently identified) apparently has evolved to encode all seven transmembrane segments as well as the four extracellular and four intracytoplasmic domains that characterize this family of proteins. Indeed, the recurrence of the Leu-Leu-Ser-Leu-Ala pentapeptide in transmembrane domains I and II can not be taken as evidence of domain I being the ancestor of domain II, for this pentapeptide occupies positions 10–14 in the 24-residue domain I, whereas it occupies positions 4–8 in the 22-residue domain II. In view of the demonstrated versatility of oligomeric repeats as the coding sequence, convergent evolution of similar domains appears to have been quite probable at the early stage of evolution when coding sequences were at their innovative best. The seven transmembrane domains of this family appear to be the products of convergent evolution within one coding sequence.

1. Gilbert, W. (1978) *Nature (London)* **271,** 501.
2. Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M. & Suzuki, K. (1984) *Nature (London)* **312,** 566–570.
3. Sudhof, T. C., Goldstein, J. L., Brown, M. S. & Russell, D. W. (1985) *Science* **228,** 815–824.
4. DiScipio, R. G., Gehring, M. R., Podack, E. R., Kan, C. C., Hugli, T. E. & Fey, G. H. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7298–7302.
5. Ohno, S. (1985) *Trends Genet.* **1,** 196–200.
6. Bridson, P. K. & Orgel, L. E. (1980) *J. Mol. Biol.* **144,** 567–577.
7. Kubo, T., Fukuda, K., Mikami, A., Maeda, A., Takahashi, H., Mishina, M., Haga, T., Haga, K., Ichiyama, A., Kojima, M., Matsuo, H., Hirose, T. & Numa, S. (1986) *Nature (London)* **323,** 411–416.
8. Dixon, R. A. F., Kobilka, B. K., Strader, D. J., Benovic, J. L., Dohlman, H. G., Frielle, T., Bolanowski, M. A., Bennett, C. D., Rands, E., Diehl, R. E., Mumford, R. A., Slater, E. E., Sigal, I. S., Caron, M. G., Lefkowitz, R. J. & Strader, C. D. (1986) *Nature (London)* **321,** 75–79.
9. Ovchinnikov, Y. A. (1982) *FEBS Lett.* **148,** 179–191.