Information Chemistry and Molecular Evolution of Protein. Evolutionary Tree of Calmodulin

Yôichi IIDA

Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060

(Received December 10, 1981)

Synopsis. The amino acid sequences of calmodulin, one of calcium-binding proteins, taken from bovine brain, scallop, sea anemone, and tetrahymena were compared to examine its molecular evolution. By the use of maximum parsimony method, its evolutionary tree was constructed, It was found that this protein has changed much more slowly than have such proteins as cytochrome c.

Origin and evolution of life is one of the most important problems in biology and biochemistry. One of the best ways to approach this problem is to study molecular evolution of proteins.[1] In the course of evolution, natural selection improved the functions of protein molecules and preserved the improvements. However, it has been known that there are rapidly changing and slowly changing proteins against evolution. An advantage of studying the molecular evolution arises from this phenomenon. If the broad outlines of the history of life are wanted, then a widely distributed, slowly evolving protein must be studied. On the other hand rapidly changing proteins such as fibrinopeptides are most useful to study the details of one branch of the evolutionary tree of living organisms.

So far, cytochrome c has been known to be a typical example of widely distributed, slowly evolving proteins, and its evolutionary tree has been constructed.[1] In the present paper, we examine the molecular evolution of calmodulin (CM). It is also widely distributed throughout eukaryotes, and changes much more slowly than does cytochrome c. This implies that the evolution of CM is useful to measure the history of life with much longer time scale than that of cytochrome c. By applying maximum parsimony method (a technique of information chemistry) to the primary amino acid sequence, we construct an evolutionary tree of CM taken from various biological species.

CM is one of the calcium-binding proteins.[2,3] It has four calcium-binding sites, and interacts reversibly with Ca^{2+} to form a calmodulin–Ca^{2+} complex, whose activity is controlled by the cellular flux of Ca^{2+}. Once bound to Ca^{2+}, CM assumes a more helical conformation to become the active form. Thus, CM plays an important role in regulating other enzymatic reactions or cellular processes. The complete amino acid sequence of bovine brain CM was determined by Watterson *et al.*[4] and by Kasai *et al.*[5] It contains 148 amino acids and has a trimethylated lysine at position 115. The primary structure of CM appears to be highly conservative throughout the phylogenetic scale. For example, the amino acid sequence of CM from bovine brain is virtually identical with the partially completed sequences of CM from rat testis and bovine uterus.[3] Recently, in order to examine such a highly conservative structure, Toda *et al.*,[6] Takagi *et al.*,[7] and Yazawa *et al.*[8] determined the amino acid sequences of CM from phylogenetically more distant species, such as scallop (*Patinopecten*), sea anemone (*Metridium senile*), and tetrahymena (*Tetrahymena pyriformis*).

For the complete amino acid sequences of CM from (A) bovine brain, (B) scallop, (C) sea anemone, and (D) tetrahymena, refer to those references. In the following, we discuss only the positions and the species of amino acid residues which are different from each other. For this purpose, we take an expression of {position, (a), (b), (c), (d)}, where (a), (b), (c), or (d) indicates the amino acid residue of CM from the biological species of (A), (B), (C), or (D) at the corresponding position, respectively. Using this expression, we can only find the following differences; {70, Thr, Thr, Thr, Ser}, {71, Met, Met, Met, Leu}, {85, Ile, Ile, Ile, Leu}, {86, Arg, Arg, Arg, Ile}, {90, Arg, Arg, Arg, Lys}, {94, Lys, Lys, Lys, Arg}, {99, Tyr, Phe, Phe, Leu}, {101, Ser, Ser, Ser, Thr}, {135, Gln, Gln, Gln, His}, {136, Val, Val, Val, Ile}, {143, Gln, Thr, Lys, Arg}, {146, Thr, Thr, Thr, Deletion}, and {147, Ala, Ser, Ser, Ala}. The amino acids at the other positions are common with the four biological species. Highly conservative property of the amino acid sequence of CM is well demonstrated, if we compare the difference of the amino acid residues at the same position. For example, only three amino acid differences are found between bovine brain and sea anemone. Concerning the sequences of bovine brain and tetrahymena, only eleven amino acid changes and one lack of amino acid residue at position 146 are found in tetrahymena.

It is generally known that codons of m RNA are translated into amino acids of protein.[1] Any codon is composed of a series of three nucleotides and is expressed by "the genetic code." Therefore, the observed difference of amino acid residues of CM between two biological species is attributable to the difference of nucleotides between the corresponding codons. Referring to the genetic code, we assume that the difference of amino acids of CM between the two species is given

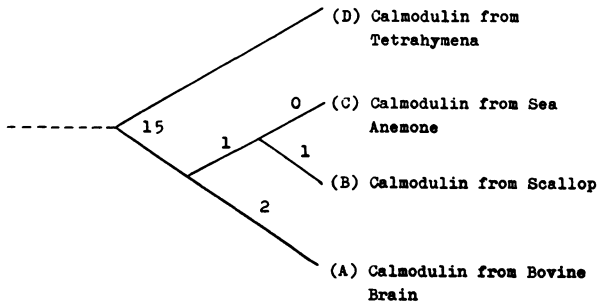

Fig. 1. The evolutionary tree estimated by applying maximum parsimony method to the difference in the amino acid sequences of calmodulin taken from (A) bovine brain, (B) scallop (*Patinopecten*), (C) sea anemone (*Metridium senile*), and (D) tetrahymena (*Tetrahymena pyriformis*). Numbers indicate nucleotide replacements.

by the least number of nucleotide replacements of the codons. This approach is called "maximum parsimony method." It is considered that the distance between two biological species may be also expressed by such a number of replacements. According to this method, the number of nucleotide replacements in CM between two biological species of (A) and (B) is 4, that between (A) and (C) is 3, that between (A) and (D) is 17, that between (B) and (C) is 1, that between (B) and (D) is 17, and that between (C) and (D) is 16. Here, the lack of amino acid residue at position 146 of (D) is counted as three nucleotide replacements. From these data, an evolutionary tree of CM is constructed as is shown in Fig. 1, where numbers indicate nucleotide replacements. Using this evolutionary tree together with the data on the difference of amino acid residues given previously, we can reconstruct the amino acid sequence of the common ancestor of (A), (B), and (C) as ···-$\overset{99}{\text{Phe}}$-···-$\overset{143}{\text{Lys}}$-···-$\overset{147}{\text{Ala}}$-···, where the remaining amino acid residues are common to those of the three species. In order to reconstruct the amino acid sequence of the common ancestor of (A), (B), (C), and (D), it is necessary to know the amino acid sequences of CM of fungi and plants, which had diverged earlier than the common ancestor of (A), (B), (C), and (D) diverged. However, no such data are available at present.

Next, we compare evolutionary rate of CM with that of cytochrome c, which contains 104 amino acid residues. In cytochrome c, the primates differ from other mammals by 8—12 residues; mammals and birds differ by an average of 9.9; mammals and reptiles and amphibia by 14; the vertebrates of all kinds differ from insects by 26.[1] On the other hand, as was mentioned, only three amino acid differences were found in CM between bovine brain and sea anemone. Biologically speaking, these two biological species are phylogenetically more distant from each other than those species given in cytochrome c. Therefore, CM appears to be much more conservative against evolution than cytochrome c. Why CM is so conservative? One of the reasons for this is that CM is a finely tooled molecule which must mesh with a number of other macromolecules to regulate various enzymatic reactions or cellular processes.[2,3] This means that a slight structural change in CM will have great influence upon such reactions and processes and that the vast majority of chance alternations are intolerable. On the other hand, cytochrome c, although it meshes with some other proteins, has less influence upon other reactions, and a chance mutation which is tolerable will occur more frequently. In this respect, if the primary amino acid sequences of CM are determined with more variety of biological species, we can measure the history of life with considerably longer time scale.

The author would like to express his appreciation to Professor Koichi Yagi for his kind advice and stimulating discussion regarding this work.

References

1) See, for example, M. O. Dayhoff, "Atlas of Protein Sequence and Structure," National Biomedical Research Foundation, Washington, D. C. (1972), Vol. 5.

2) See, for example, "Calcium Binding Proteins and Calcium Function," ed by R. H. Wasserman *et al.*, North Holland, New York (1977).

3) W. Y. Cheung, *Science*, **207**, 19 (1980).

4) D. M. Watterson, F. Sharief, and T. C. Vanaman, *J. Biol. Chem.*, **255**, 962 (1980).

5) H. Kasai, Y. Kato, T. Isobe, H. Kawasaki, and T. Okuyama, *Biomed. Res.*, **1**, 248 (1980).

6) H. Toda, M. Yazawa, K. Kondo, T. Honma, K. Narita, and K. Yagi, *J. Biochem.*, **90**, 1493(1981).

7) T. Takagi, T. Nemoto, K. Konishi, M. Yazawa, and K. Yagi, *Biochem. Biophys. Res. Commun.*, **96**, 377 (1980)

8) M. Yazawa, K. Yagi, H. Toda, K. Kondo, K. Narita, R. Yamazaki, K. Sobue, S. Kakiuchi, S. Nagano, and Y. Nozawa, *Biochem. Biophys. Res. Commun.*, **99**, 1051 (1981).