cDNA Sequences and Molecular Evolution of Calmodulin Genes of Chicken and Eel

Yôichi IIDA
Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060
(Received April 18, 1984)

Synopsis. c DNA sequences of calmodulin genes taken from chicken and eel were compared to investigate the molecular evolution of its protein. One notable feature of the amino acid coding region of the gene is that the second positions of codons are more conservative than the first positions, while the third positions are highly diverged to exhibit almost saturation. This kind of behaviour is reminiscent of that of the histone gene.

Origin and evolution of protein is one of the most important problems in biology and biochemistry. Recently, primary nucleotide sequences have been determined with a number of genes which code for proteins. Studies on molecular evolution of proteins in terms of such DNA sequences are very interesting. Calmodulin (CaM) is one of the calcium-binding proteins.[1-3] It has four calcium-binding sites, and interacts reversibly with Ca^{2+} to form a calmodulin-Ca^{2+} complex. It contains 148 amino acids and has a trimethylated lysine at position 115. One notable feature of CaM is that it is widely distributed throughout eukaryotes and that the primary amino acid sequence is highly conservative against evolution. In previous papers,[4,5] we examined the origin and evolution of CaMs in terms of their primary amino acid sequences. So for, their amino acid sequences have been determined with a number of representative species.[1-5] However, little information was known with the m RNAs that code for CaM. Recently, Lagacé *et al.* and Putkey *et al.* isolated m RNAs of CaM from the electroplax of electric eel (Electrophorus electricus) and from chicken brain, respectively,[6,7] and their c DNAs (DNA complementary to m RNA) have been cloned and sequenced. Therefore, it is very interesting to compare the DNA sequences between both biological species. We pay attention to how the highly conservative amino acid sequence of CaM is related to the coding region of c DNA.

Figure 1 shows the amino acid coding regions of the nucleotide sequences of CaM c DNAs for chicken and eel. The numbers indicate amino acid positions. The initiation (ATG) and termination (TGA) codons are common to the two species. If we exclude these two codons, the coding region of the c DNA is composed of 148 codons. For the two species, the corresponding codons are aligned and compared in Fig. 1. Those nucleotides which are different between the two species are indicated by asterisk (*). The other nucleotides are found to be identical between the two. As is expected from the highly conservative amino acid sequences of CaM, the primary nucleotide sequences in the amino acid-coding regions of the respective c DNA are highly homologous, with an overall homology of 79%. Particularly, the first and second positions (replacement sites) of codons show markedly higher degree of homology than does the overall homology. In contrast to these, the third positions (silent sites) of codons are very much diverged.

In order to treat these situations more quantitatively, we consider one-parameter and two-parameter models for evaluating the degree of nucleotide changes proposed by Kimura *et al.*[8] In estimating the evolutionary distances between homologous sequences in terms of the number of nucleotide substitutions, corrections for multiple and revertant changes at homologous sites are important. The total number of nucleotide substitutions per site, K, which separate two sequences and therefore involve two branches each with time length T, is expressed by $K=2Tk$, where k is the total rate of substitutions per site per year. In the one-parameter model where probability of every nucleotide substitution is assumed to be common, K is given by[8]

$$K = -\frac{3}{4}\ln\left(1-\frac{4\lambda}{3}\right), \tag{1}$$

where λ is the fraction of sites for which two sequences differ from each other. From the data of Fig. 1 of the chicken and eel CaM genes, the λ value for the first positions of codons was calculated as 7/148=0.0473, while the values for the second and third positions, as 1/148=0.00676 and 84/148=0.568, respectively. As can be seen from the highly conservative amino acid sequences of CaM, the magnitudes of λ for the replacement sites are extremely small, while that for the silent site is found to be larger than 1/2. If nucleotide substitutions occur randomly, the λ value will show a saturated value of 3/4=0.75. In view of this, the value for the third position appears to approach such a saturated value. Using Eq. 1, the K values for the first, second and third positions of codons were calculated as 0.049 (±0.019), 0.0068 (±0.0068) and 1.06 (±0.17), respectively, where the figures in parentheses show standard error.[8]

In the two-parameter model, the probabilities of transition type nucleotide substitutions and transversion type substitutions are discriminated. The transitions are substitutions of purines by purines (A↔G), and pyrimidines by pyrimidines (T↔C), while the transversions are substitutions in which a purine is replaced by a pyrimidine, or a pyrimidine is replaced by a purine (A or G↔T or C). In this model, the K value is expressed by[9]

$$K = -\frac{1}{2}\ln\{(1-2P-Q)\sqrt{1-2Q}\}. \tag{2}$$

Here, P and Q are given by the fraction of sites for which two sequences differ from each other with respect to transition type and transversion type substitutions, respectively. In the chicken and eel CaM genes, P and Q for the first positions of codons were calculated as 4/148=0.0270 and 3/148=0.0203, respectively. These values lead to K=0.049 (±0.019) for the first positions, where the figure in parentheses shows stan-

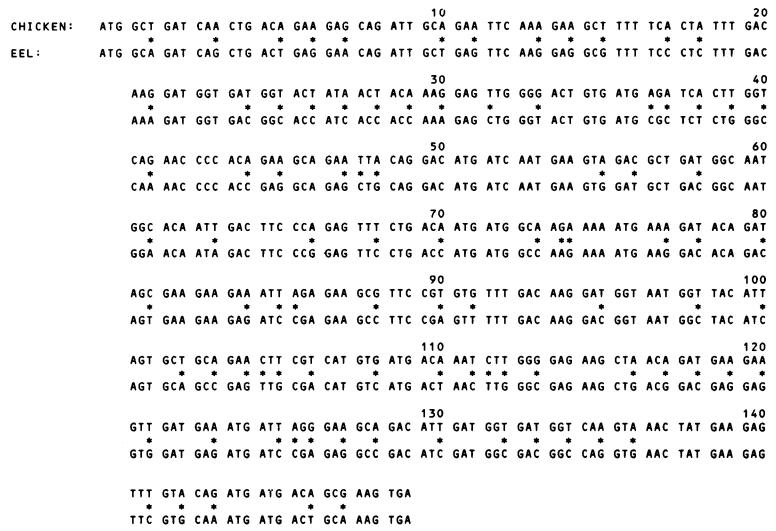

Fig. 1. Comparison and homology of the amino acid coding regions of the c DNA sequences taken from chicken and electric eel (Electrophorus electricus).[6,7] The initiation (ATG) and termination (TGA) codons are common between the two species. The numbers indicate amino acid positions. Those nucleotides which are different between the two species are shown by asterisk (*).

dard error.[9] Similarly, for the second positions, $P=1/148=0.00676$, $Q=0.00000$ and $K=0.0068$ (± 0.0068), and for the third positions, $P=53/148=0.358$, $Q=31/148=0.209$ and $K=1.44$ (± 0.49). In all three positions, P is greater than $(1/2)Q$, indicating that the transitions have occurred more frequently than have the transversions. In this respect, the two-parameter model is more realistic than the one-parameter model, and the following discussion will be made with the two-parameter model.

The most notable feature of the CaM genes is that the K value of the third positions (K_3) is much greater than those of the first (K_1) and second (K_2) positions. This means that the pressure by natural selection has been strongly applied to amino acid sequences of CaM, while the third positions of codons are almost neutral against evolution; most of the third positions are synonymous with respect to changes of amino acid residues. It is also found that the magnitude of K_2 is less than that of K_1. The order of $K_2<K_1<K_3$ holds generally with various protein genes, such as β-globin gene.[10]

Using the equation of $K=2Tk$, we estimate the values of k, the rate of nucleotide substitutions per site per year. We assume that the common ancestor of eel and chicken CaMs diverged $T=3.5\times10^8$ years ago (the time of divergence into osteichthyes and amphibians).[11] Then, the values of k for the first, second and third positions of codons of CaM are calculated as $k_1=0.070\times10^{-9}$, $k_2=0.0097\times10^{-9}$ and $k_3=2.1\times10^{-9}$, respectively. These results may be compared with histone H4 which is known to be the most highly conservative protein against evolution. The rate of amino acid substitutions per site per year of histone H4 is known as $k_{aa}=0.006\times10^{-9}$.[12] The value of k_2 of CaM is of comparable order with that of k_{aa}, while the value of k_1 of CaM is about ten times larger than that of k_{aa}. Therefore, although the first positions of codons of CaM are less conservative than those of histone H4, the conservative properties of these two proteins are comparable with each other. In contrast to these, the rate of $k_3=2.1\times10^{-9}$ of the third positions of CaM is as high as 2.3×10^{-9} of fibrinopeptide or $(3.7\pm1.4)\times10^{-9}$ of the third nucleotide positions of histone H4 gene.[12] The latter two values are known as the highest rate of nucleotide substitutions. It is remarkable that, in CaM gene, synonymous mutant substitutions have occurred with the highest rate, while the amino acid substitutions have occurred with as low rate as that of histone H4.

References

1) W. Y. Cheung, *Science*, **207**, 19 (1980).
2) T. Takagi, T. Nemoto, K. Konishi, M. Yazawa, and K. Yagi, *Biochem. Biophys. Res. Commun.*, **96**, 377 (1980).
3) 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).
4) Y. Iida, *Bull. Chem. Soc. Jpn.*, **55**, 2683 (1982).
5) Y. Iida, *J. Mol. Biology*, **159**, 167 (1982).
6) L. Lagacé, T. Chandra, S. L. C. Woo and A. R. Means, *J. Biol. Chem.*, **258**, 1684 (1983).
7) J. A. Putkey, K. F. Ts'ui, T. Tanaka, L. Lagacé, J. P. Stein, E. C. Lai and A. R. Means, *J. Biol. Chem.*, **258**, 11864 (1983).
8) M. Kimura and T. Ohta, *J. Mol. Evol.*, **2**, 87 (1972).
9) M. Kimura, *J. Mol. Evol.*, **16**, 111 (1980).
10) "Molecular Evolution, Protein Polymorphism and The Neutral Theory," ed by M. Kimura, Springer-Verlag, Berlin, Heidelberg, New York, p. 31 (1982).
11) A. S. Romer, "Vertebrate Paleontology," Third Edition, The University of Chicago Press, Chicago and London (1966).
12) M. Kimura, *Nature*, **267**, 275 (1977).