

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

Yōichi 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 far, 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 by⁸

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

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 $0.049 (\pm 0.019)$, $0.0068 (\pm 0.0068)$ and $1.06 (\pm 0.17)$, respectively, where the figures in parentheses show standard error.⁹

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 ($\text{A} \leftrightarrow \text{G}$), and pyrimidines by pyrimidines ($\text{T} \leftrightarrow \text{C}$), while the transversions are substitutions in which a purine is replaced by a pyrimidine, or a pyrimidine is replaced by a purine ($\text{A or G} \leftrightarrow \text{T or C}$). In this model, the K value is expressed by⁹

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

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 $K=0.049 (\pm 0.019)$ for the first positions, where the figure in parentheses shows stan-

CHICKEN: ATG GCT GAT CAA CTG ACA GAA GAG CAG ATT GCA GAA TTC AAA GAA GCT TTT TCA CTA TTT GAG
EEL: ATG GCA GAT CAG CTG ACT GAG GAA CAG ATT GCT GAG TTC AAG GAG GCG TTT TCC CTC TTT GAG

10
30
40
50
60
70
80
90
100
110
120
130
140

150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940
950
960
970
980
990
1000

1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
1440
1450
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720
1730
1740
1750
1760
1770
1780
1790
1800
1810
1820
1830
1840
1850
1860
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100
2110
2120
2130
2140
2150
2160
2170
2180
2190
2200
2210
2220
2230
2240
2250
2260
2270
2280
2290
2300
2310
2320
2330
2340
2350
2360
2370
2380
2390
2400
2410
2420
2430
2440
2450
2460
2470
2480
2490
2500
2510
2520
2530
2540
2550
2560
2570
2580
2590
2600
2610
2620
2630
2640
2650
2660
2670
2680
2690
2700
2710
2720
2730
2740
2750
2760
2770
2780
2790
2800
2810
2820
2830
2840
2850
2860
2870
2880
2890
2900
2910
2920
2930
2940
2950
2960
2970
2980
2990
3000
3010
3020
3030
3040
3050
3060
3070
3080
3090
3100
3110
3120
3130
3140
3150
3160
3170
3180
3190
3200
3210
3220
3230
3240
3250
3260
3270
3280
3290
3300
3310
3320
3330
3340
3350
3360
3370
3380
3390
3400
3410
3420
3430
3440
3450
3460
3470
3480
3490
3500
3510
3520
3530
3540
3550
3560
3570
3580
3590
3600
3610
3620
3630
3640
3650
3660
3670
3680
3690
3700
3710
3720
3730
3740
3750
3760
3770
3780
3790
3800
3810
3820
3830
3840
3850
3860
3870
3880
3890
3900
3910
3920
3930
3940
3950
3960
3970
3980
3990
4000
4010
4020
4030
4040
4050
4060
4070
4080
4090
4100
4110
4120
4130
4140
4150
4160
4170
4180
4190
4200
4210
4220
4230
4240
4250
4260
4270
4280
4290
4300
4310
4320
4330
4340
4350
4360
4370
4380
4390
4400
4410
4420
4430
4440
4450
4460
4470
4480
4490
4500
4510
4520
4530
4540
4550
4560
4570
4580
4590
4600
4610
4620
4630
4640
4650
4660
4670
4680
4690
4700
4710
4720
4730
4740
4750
4760
4770
4780
4790
4800
4810
4820
4830
4840
4850
4860
4870
4880
4890
4900
4910
4920
4930
4940
4950
4960
4970
4980
4990
5000
5010
5020
5030
5040
5050
5060
5070
5080
5090
5100
5110
5120
5130
5140
5150
5160
5170
5180
5190
5200
5210
5220
5230
5240
5250
5260
5270
5280
5290
5300
5310
5320
5330
5340
5350
5360
5370
5380
5390
5400
5410
5420
5430
5440
5450
5460
5470
5480
5490
5500
5510
5520
5530
5540
5550
5560
5570
5580
5590
5600
5610
5620
5630
5640
5650
5660
5670
5680
5690
5700
5710
5720
5730
5740
5750
5760
5770
5780
5790
5800
5810
5820
5830
5840
5850
5860
5870
5880
5890
5900
5910
5920
5930
5940
5950
5960
5970
5980
5990
6000
6010
6020
6030
6040
6050
6060
6070
6080
6090
6100
6110
6120
6130
6140
6150
6160
6170
6180
6190
6200
6210
6220
6230
6240
6250
6260
6270
6280
6290
6300
6310
6320
6330
6340
6350
6360
6370
6380
6390
6400
6410
6420
6430
6440
6450
6460
6470
6480
6490
6500
6510
6520
6530
6540
6550
6560
6570
6580
6590
6600
6610
6620
6630
6640
6650
6660
6670
6680
6690
6700
6710
6720
6730
6740
6750
6760
6770
6780
6790
6800
6810
6820
6830
6840
6850

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 $K=0.0068 (\pm 0.0068)$, and for the third positions, $P=53/148=0.358$, $Q=31/148=0.209$ and $K=1.44 (\pm 0.49)$. 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 $K=2T�$, 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 $k_3=2.1\times 10^{-9}$ of the third positions of CaM is as high as 2.3×10^{-9} of fibrinopeptide or $(3.7\pm 1.4)\times 10^{-9}$ 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).