

*Hypothesis*

# Molecular evolution of the calmodulin gene

Hiroshi Nojima

*Department of Pharmacology, Jichi Medical School Minamikawachi-machi, Tochigi-ken 329-04, Japan*

Received 9 March 1987

On the basis of the intron/exon organization and the intramolecular homology of DNA sequences, I propose a novel model for genesis of the calmodulin gene. A primordial calmodulin gene consisting of 51 base pairs (17 amino acids) was subjected to three-fold duplication to create modern calmodulin with four calcium-binding subdomains. The model elucidates the seemingly enigmatic positions of splice junctions observed in calmodulin genes.

Calmodulin; Calmodulin gene; Nucleotide sequence; Sequence homology; Evolution

## 1. INTRODUCTION

Calmodulin, a monomeric protein of 148 amino acids with four calcium-binding subunits, is composed of a highly conserved amino acid sequence (100% homology among mammals, birds and amphibians). It is noteworthy that the molecule harbors four domains with homologous amino acid sequences, each domain containing one calcium-binding site. The level of homology is most apparent between domains 1 and 3 and domains 2 and 4. From these results, a model for the evolutionary history of calmodulin is proposed [1,2], where a primitive gene encoding a peptide with one calcium-binding helix-loop-helix structure is supposed to have been duplicated twice to create modern calmodulin with four calcium-binding subdomains.

Considering the split organization of eukaryotic genes, Gilbert [3] proposed that exons correspond to functional domains of proteins which can be

reshuffled by recombinations within introns to create novel proteins. This idea was refined by Blake [4] who suggested that exons might code for discrete, stable regions of protein. Go [5] has developed this idea by introducing the diagonal plotting method (Go plot) and successfully predicted the position of the third intron of the leghemoglobin gene.

To comply with both the standard exon-shuffling model and intramolecular homology, an expected calmodulin gene would consist of four exons in the coding region, each exon bearing one calcium-binding site. This situation is actually observed in a calcium-dependent protease with four calcium-binding domains [6]. However, the determined structure of chicken [7], rat [8,9] and *Drosophila* [10] calmodulin genes was found to be at variance with anticipation. Two splice junctions in *Drosophila* and three splice junctions in chicken and rat calmodulin genes separate the calcium-binding subdomains. These results indicate that a simple 'double intragenic duplication model' does not explain the evolutionary history of calmodulin. Here, I propose a novel evolutionary model for the calmodulin gene.

Correspondence address: H. Nojima, Department of Pharmacology, Jichi Medical School Minamikawachi-machi, Tochigi-ken 329-04, Japan

## 2. A THREE-FOLD DUPLICATION MODEL

The nucleotide sequences in the coding region of rat [8,9] and chicken [7,11] calmodulin genes are aligned based on the intron/exon organization of chicken gene for the convenience of discussion (fig.1a). Comparison of nucleotide sequences among individual exons detected eight segments that displayed internal homology (horizontal arrows in fig.1a). Nucleotide sequences of each seg-

ment were aligned (fig.1b) to display homologous regions. Segments 2-7 closely resemble each other. Segments 1 and 8 show comparatively less homology. It is notable that each segment harbors a region with the most homologous consensus sequence (core sequence) of 12 nucleotides. Conservation of sequences among segments and core sequences is more pronounced in chicken than in rats. These results indicate that the eight segments are somewhat related in sequence. No obvious

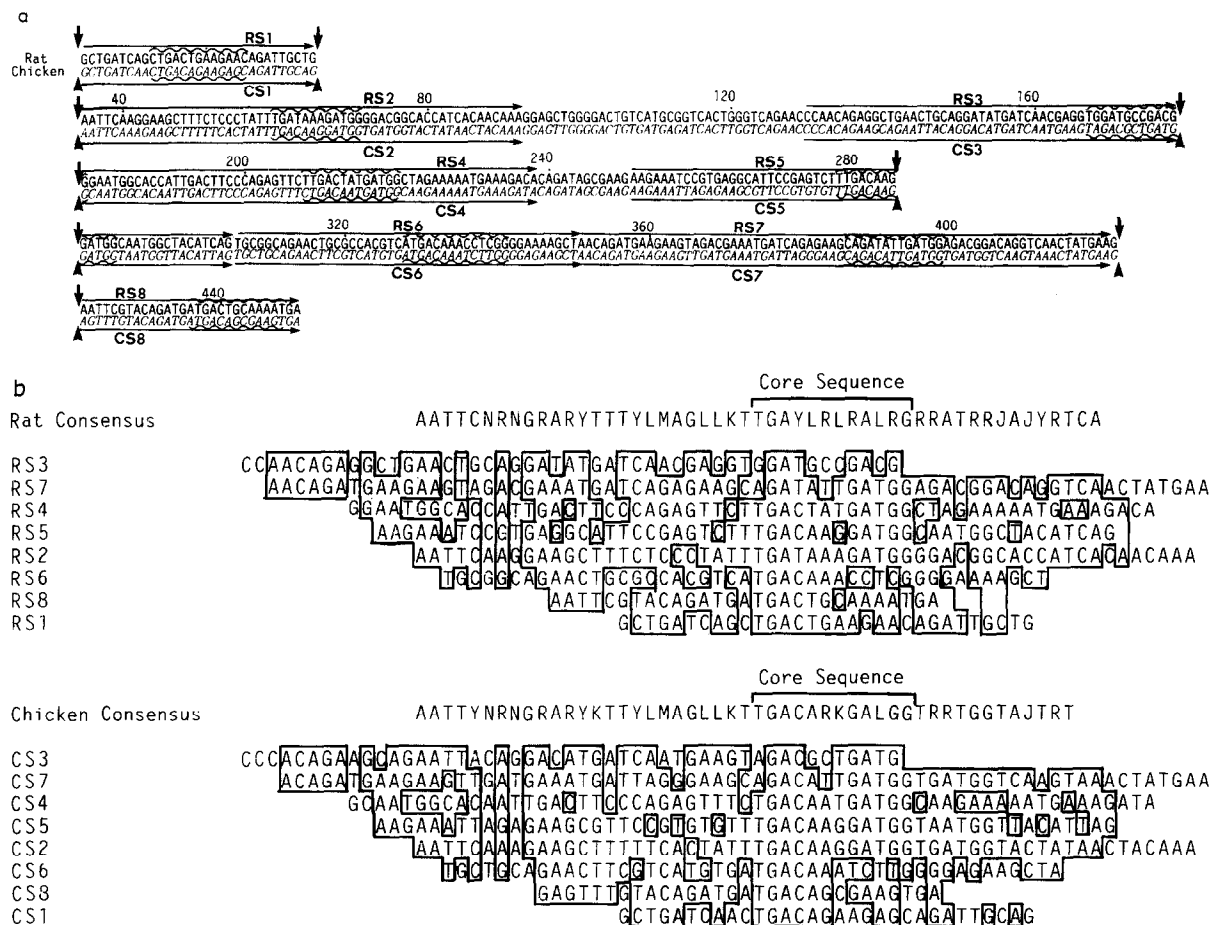


Fig.1. Alignment of nucleotide sequences in the coding region of rat and chicken calmodulin genes. (a) Locations of homologous segments for rat (RS1-RS8) and chicken (CS1-CS8) calmodulin genes are shown by horizontal arrows above and below the nucleotide sequences. Nucleotide sequences in the coding regions are blocked on the basis of the intron/exon organization of rat and chicken calmodulin genes. Splice junction for rat (↓) and chicken (▲) calmodulin genes are denoted;. Nucleotide numbering begins with the translational initiation codon (ATG). (b) Alignment of intramolecular homologous regions of nucleotide sequences among segments of rat (RS1-RS8) and chicken (CS1-CS8) calmodulin genes. Sequences of each segment identical to consensus sequences are encircled by lines. Most homologous regions are denoted 'Core Sequence'. Pairs of nucleotides: Y (T or C), R (A or G), K (G or T), M (A or C), S (G or C) and W (A or T).

homology was detected in the 5'- and 3'-non-coding regions.

To explain the unexpected intron/exon organization and the eight-fold intragenic homologies observed in rat and chicken calmodulin genes, a model for the genesis of the modern calmodulin gene is given in fig.2. A primordial calmodulin (PCM) gene (17 amino acids, 51 bp) with one inefficient calcium-binding activity was firstly duplicated twice. At the end of the second duplication, the middle two exons were fused to form a

double-sized exon. These processes (step 1) create an ancient calmodulin (ACM) with three exons and two calcium-binding subdomains. ACM was furthermore duplicated (third duplication) to produce a prototype of modern calmodulin (p-MCM) (step 2). Before the generation of the modern calmodulin (MCM) gene, the refinement procedure took place (step 3). At this step, the 5'- and 3'-noncoding segments that are unrelated to PCM were introduced into the gene. This step may be a comparatively recent event since the sizes and

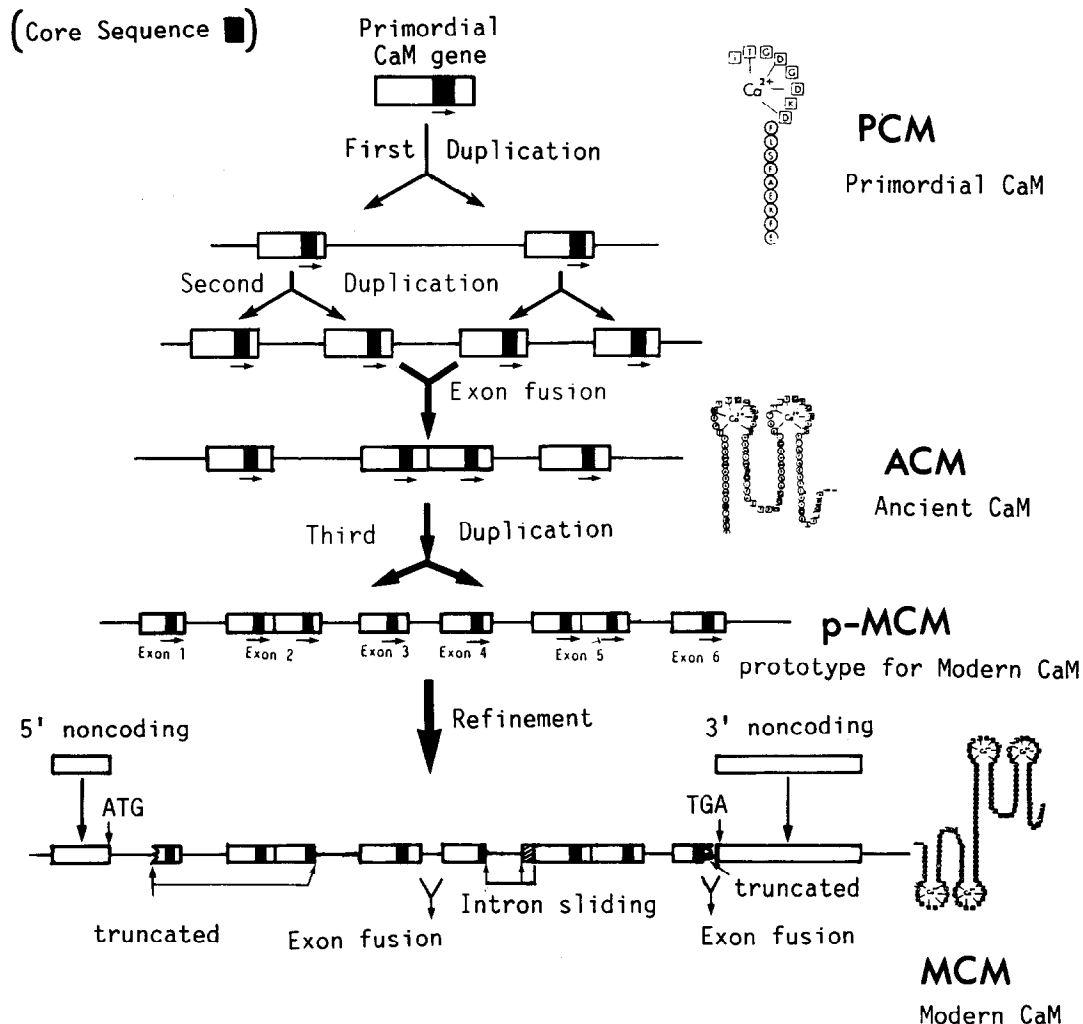


Fig.2. A possible model for the genesis of the calmodulin gene. Primordial calmodulin (PCM) gene with 17 amino acids (51 nucleotides) evolved sequentially to ancient calmodulin (ACM), prototype of modern calmodulin (p-MCM) and modern calmodulin (MCM) genes through the process including three-fold duplications of the gene. Horizontal arrows display the orientation of the consensus core sequence (■).

nucleotide sequences of these regions are quite distinct among calmodulin genes [8-14]. The 3'-noncoding region seems to have been fused to exon 6 of p-MCM. Exon 1 (segment 1) and exon 6 (segment 8) of p-MCM were truncated to become smaller exons and their nucleotide sequences diverged probably because they are situated at the ends of the molecule and are not responsible for the calcium-binding activity. Exon 2 (segment 3) was also truncated during evolution. From exon 4 (segment 5) to exon 5 (segment 6) of p-MCM, an intron-sliding procedure seems to have occurred resulting in elongated exon 5 and truncated exon 4 of MCM. Given the expected age of the calmodulin gene, it is feasible that the locations of introns within codons have shifted and that the size and the primary sequence homology have been eroded by mutations. Sliding of intron/exon junctions has been observed in a variety of genes and is considered to constitute one mechanism for generating the peptide length polymorphisms and divergent sequences found in protein families [15]. Intron sliding may have occurred because of comparatively inaccurate primitive splicing mechanisms as compared with the present splicing machinery. In addition, mutations near the splicing junctions may have served as a source of the creation of cryptic splice signals, thus leading to clipping and joining cryptic splice sites of ancient genes. Joining of exons has taken place twice, one between segments 2 and 3 and the other between segments 6 and 7. Exon fusion seems to be a common event among calmodulin genes. Three exon fusions between exons 1-2, and exons 4-5 of p-MCM seem to have occurred in the *Drosophila* calmodulin gene [10]. All exons in the coding region seem to have been fused in yeast [16,17]. The location of the first intron right after the initiation codon (ATG) is conserved among rat [8,9], chicken [7], *Drosophila* [10] and yeast [16,17] calmodulin genes, indicating that the addition of the 5'-non-

coding sequence is actually a recent event. If the above-stated model is correct, extra splice junctions, for example between Ile-27 and Thr-28 or Leu-116 and Thr-117 of MCM, created by a distinct exon-fusion pattern may be predicted in the calmodulin genes of other species. It is therefore interesting to investigate the intron/exon organization of calmodulin genes from other species with large evolutionary separation.

## REFERENCES

- [1] Watterson, D.M., Sharief, F. and Vanaman, T.C. (1980) *J. Biol. Chem.* 255, 962-975.
- [2] Iida, Y. (1982) *J. Mol. Biol.* 159, 167-17.
- [3] Gilbert, W. (1978) *Nature* 271, 501.
- [4] Blake, C.C.F. (1978) *Nature* 273, 267.
- [5] Go, M. (1981) *Nature* 291, 90-92.
- [6] Emori, Y., Ohno, S., Tobita, M. and Suzuki, K. (1985) *FEBS Lett.* 194, 249-252.
- [7] Epstein, P., Simmen, R.C., Tanaka, T. and Means, A. (1986) *Methods Enzymol.* 139, 217-229.
- [8] Nojima, H. and Sokabe, H. (1986) *J. Mol. Biol.* 190, 391-400.
- [9] Nojima, H. and Sokabe, H. (1987) *J. Mol. Biol.* 193, 439-445.
- [10] Yamanaka, M.K., Saugstad, J.A., Hanson-Painton, O., McCarthy, B.J. and Tobin, S.L. (1987) *Nuc. Acids Res.*, in press.
- [11] Putkey, J.A., Ts'ui, K.F., Tanaka, T., Lagace, L., Stein, J.P., Lai, E.C. and Means, A.R. (1983) *J. Biol. Chem.* 258, 11864-11870.
- [12] Lagace, L., Chandra, T., Woo, S.L.C. and Means, A.R. (1983) *J. Biol. Chem.* 258, 1684-1688.
- [13] Chien, Y.-H. and Dawid, I.B. (1984) *Mol. Cell. Biol.* 4, 507-513.
- [14] Goldhage, H. and Clarke, M. (1986) 6, 1851-1854.
- [15] Craik, C.S., Rutter, W.J. and Fletterick, R. (1983) *Science* 220, 1125-1129.
- [16] Davis, T.N., Urdea, M.S., Masiarz, F.R. and Thorner, J. (1986) *Cell* 47, 423-431.
- [17] Takeda, T. and Yamamoto, M. (1987) *Proc. Natl. Acad. Sci. USA*, in press.