Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 1605–1608, March 1987
Evolution

Possible origin of a calmodulin gene that lacks intervening sequences

KAREN D. GRUSKIN*, TEMPLE F. SMITH†, AND MORRIS GOODMAN‡

*Dana-Farber Cancer Institute, Boston, MA 02115; †Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115; and ‡Department of Anatomy, Wayne State University, Detroit, MI 48202

Communicated by Roy J. Britten, September 8, 1986 (received for review July 8, 1986)

ABSTRACT **The divergent, muscle-specific allele of the chicken calmodulin gene contains no intervening sequences and apparently was produced by a reverse transcriptase-mediated event. The nucleotide and deduced amino acid sequences of this gene were compared with nucleotide and amino acid sequence data of other known calmodulin genes in order to investigate its evolutionary history. These comparisons, as well as the CpG dinucleotide content, support the conclusion that this highly divergent chicken calmodulin gene did not exist for any significant period of time as a pseudogene and suggest plausible alternative genetic histories. The most parsimonious history involves the viral import of a very old foreign gene of high CpG content.**

The calmodulins are four-domain calcium-binding proteins that appear to have arisen from two gene-duplication events (1). They are among the most conserved proteins known (2). The amino acid sequences, with few exceptions, differ among the vertebrates by only one or two amino acids. The tissue-specific calmodulin gene of chicken muscle cells, *CM1* (3), which apparently was produced by the action of reverse transcriptase (RNA-directed DNA polymerase), is an important exception. *CM1* contains no intervening sequences and encodes a number of amino acid differences. Most eukaryotic genes occur with the coding sequences (exons) split up by noncoding intervening sequences (introns). Some gene families contain genetic loci that lack introns and arose, as postulated for *CM1* of chicken, from genomic insertions via reverse transcription. However, such intronless loci are typically found to be nonfunctional pseudogenes and usually are not chromosomally linked to functional genes of their gene family, in contrast to those pseudogenes that have undergone *in situ* inactivation without the singular removal of all introns. The existence of an active chicken calmodulin gene that lacks introns raises the question of whether it ever existed as an inactive pseudogene before being recruited for a tissue-specific function.

The *CM1* gene encodes a product differing by 19 amino acids from the second chicken calmodulin allele, *CL1* (3, 4). This difference is greater than the maximum 14 amino acid differences found between the calmodulin of the protozoan *Tetrahymena pyriformis* (5) and the mammalian calmodulins (Fig. 1). In spite of the amino acid divergence, the two apparently divergent chicken genes are equidistant, at the DNA level, from the calmodulin sequence of an eel, *Electrophorus electricus*. The chicken *CM1* gene also shares a high CpG content, particularly at the codon–codon boundaries, with the eel calmodulin gene (6) and one of the calmodulin genes of the African aquatic toad *Xenopus laevis* (7) (Fig. 2*a*). The data from comparisons between the chicken *CM1* gene and the known DNA and amino acid sequences of other calmodulin genes contain clues about the origin of this intronless gene.

OBSERVATIONS

It has been suggested (3) that the main hint as to the history of *CM1* is its unusually great divergence in comparison to other calmodulins at the amino acid level (Fig. 2*a*). Upon initial inspection of the number of amino acid replacements, the 19 amino acid differences between *CM1* and other calmodulins seem to indicate that the *CM1* is highly divergent. In Fig. 1, the four repeat units of the chicken calmodulins encoded by *CM1* and *CL1* and the calmodulin of *T. pyriformis* are aligned in order to identify shared amino acids and conservative changes. An analysis of these data shows that the *CM1* divergence is not as great as might be inferred from simply counting amino acid identities. First, 11 of the amino acid changes between *CM1* and *CL1* are conservative, maintaining charge, hydrophobicity, and/or functional group. Note that methionine, arginine, and arginine, at positions 31, 4, and 144, respectively, are not conservative changes between species at those positions but actually increase the inter-repeat unit similarity.

The three amino acids cysteine, asparagine, and asparagine at positions 131–133 initially appear to destroy the most conserved region [the purported Ca^{2+} binding site (2)] in the fourth repeat unit; however, there is evidence that these changes are structurally and functionally conservative. These differences are compatible with the strongly predicted β-turn (8) in this region for all four repeat domains (data not shown). The two asparagines appear at equivalent positions in the active site of troponin C, a Ca^{2+}-binding protein (9). It should also be noted that although *CM1* is the only known vertebrate calmodulin gene to encode cysteine, spinach calmodulin has one cysteine in a position equivalent to the first cysteine of *CM1* (10).

Other amino acid differences between the chicken calmodulins also may be considered conservative when compared to other known calmodulin genes. Three of the amino acid differences are common to the protozoan calmodulin gene; these three amino acids appear in the same area of the third and fourth repeats, with two of the amino acids in exactly the same position, 144 and 71 (see Fig. 1). Six of the amino acid changes introduce the rare CpG dinucleotide (the significance of which is discussed below). The conservative nature of the amino acid differences between the two chicken calmodulin genes is verified by their functional similarity as determined by Putkey *et al.* (11). Thus, the observed amino acid differences (Fig. 1) do not appear to require the prolonged period of relaxed selective pressure expected for a pseudogene.

At the DNA level, the *CM1* gene's divergence is less unusual than at the amino acid level. For example, the eel

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: *CM1*, muscle-specific chicken calmodulin gene (divergent, no introns); *CL1*, second chicken calmodulin gene.

	CM1	*T. pyr.*	*CL1*			*CM1*	*T. pyr.*	*CL1*			*CM1*	*T. pyr.*	*CL1*			*CM1*	*T. pyr.*	*CL1*
1	MET(hph)	MET(hph)	MET(hph)	——	77	MET(hph)	MET(hph)	MET(hph)	- - - -	40	LEU(hph)	LEU(hph)	LEU(hph)	——	113	LEU(hph)	LEU(hph)	LEU(hph)
2	ALA(hph)	ALA(hph)	ALA(hph)		78	ARG(+)	LYS(+)	LYS(+)		41	GLY( )	GLY( )	GLY( )	——	114	GLY( )	GLY( )	GLY( )
3	GLU(-)	ASP(-)	ASP(-)	——	79	ASP(-)	ASP(-)	ASP(-)	- - - -	42	GLN(NH2)	GLN(NH2)	GLN(NH2)		115	GLU(-)	GLU(-)	GLU(-)
4	ARG(+)	GLN(NH2)	GLN(NH2)		80	SER(-OH)	THR(-OH)	THR(-OH)		43	ASN(NH2)	ASN(NH2)	ASN(NH2)		116	LYS(+)	LYS(+)	LYS(+)
5	LEU(hph)	LEU(hph)	LEU(hph)		81	ASP(-)	ASP(-)	ASP(-)		44	PRO(hph)	PRO(hph)	PRO(hph)	- - - -	117	LEU(hph)	LEU(hph)	LEU(hph)
6	SER(-OH)	THR(-OH)	THR(-OH)	- - - -	82	SER(-OH)	SER(-OH)	SER(-OH)	- - - -	45	THR(-OH)	THR(-OH)	THR(-OH)	——	118	THR(-OH)	THR(-OH)	THR(-OH)
7	GLU(-)	GLU(-)	GLU(-)	——	83	GLU(-)	GLU(-)	GLU(-)	——	46	GLU(-)	GLU(-)	GLU(-)	- - - -	119	ASP(-)	ASP(-)	ASP(-)
8	GLU(-)	GLU(-)	GLU(-)	——	84	GLU(-)	GLU(-)	GLU(-)		47	ALA(hph)	ALA(hph)	ALA(hph)		120	GLU(-)	GLU(-)	GLU(-)
9	GLN(NH2)	GLN(NH2)	GLN(NH2)		85	GLU(-)	GLU(-)	GLU(-)		48	GLU(-)	GLU(-)	GLU(-)	——	121	GLU(-)	GLU(-)	GLU(-)
10	ILU(hph)	ILU(hph)	ILU(hph)	——	86	ILU(hph)	LEU(hph)	ILU(hph)	- - - -	49	LEU(hph)	LEU(hph)	LEU(hph)	- - - -	122	VAL(hph)	VAL(hph)	VAL(hph)
11	ALA(hph)	ALA(hph)	ALA(hph)															
12	GLU(-)	GLU(-)	GLU(-)															
13	PHE(hph)	PHE(hph)	PHE(hph)															
14	LYS(+)	LYS(+)	LYS(+)	- - - -	87	ARG(+)	ILU(hph)	ARG(+)		50	GLN(NH2)	GLN(NH2)	GLN(NH2)		123	ASP(-)	ASP(-)	ASP(-)
15	GLU(-)	GLU(-)	GLU(-)	——	88	GLU(-)	GLU(-)	GLU(-)	- - - -	51	ASP(-)	ASP(-)	ASP(-)	- - - -	124	GLU(-)	GLU(-)	GLU(-)
16	ALA(hph)	ALA(hph)	ALA(hph)	——	89	ALA(hph)	ALA(hph)	ALA(hph)	- - - -	52	MET(hph)	MET(hph)	MET(hph)	——	125	MET(hph)	MET(hph)	MET(hph)
17	PHE(hph)	PHE(hph)	PHE(hph)	——	90	PHE(hph)	PHE(hph)	PHE(hph)	- - - -	53	VAL(hph)	ILU(hph)	ILU(hph)	——	126	ILU(hph)	ILU(hph)	ILU(hph)
18	SER(-OH)	SER(-OH)	SER(-OH)		91	ARG(+)	LYS(+)	ARG(+)		54	GLY( )	ASN(NH2)	ASN(NH2)		127	LYS(+)	ARG(+)	ARG(+)
19	LEU(hph)	LEU(hph)	LEU(hph)	- - - -	92	VAL(hph)	VAL(hph)	VAL(hph)		55	GLU(-)	GLU(-)	GLU(-)	——	128	GLU(-)	GLU(-)	GLU(-)
20	PHE(hph)	PHE(hph)	PHE(hph)	——	93	PHE(hph)	PHE(hph)	PHE(hph)	- - - -	56	VAL(hph)	VAL(hph)	VAL(hph)	- - - -	129	ALA(hph)	ALA(hph)	ALA(hph)
21	ASP(-)	ASP(-)	ASP(-)	——	94	ASP(-)	ASP(-)	ASP(-)	——	57	ASP(-)	ASP(-)	ASP(-)	——	130	ASP(-)	ASP(-)	ASP(-)
22	ARG(+)	LYS(+)	LYS(+)	- - - -	95	LYS(+)	ARG(+)	LYS(+)		58	ALA(hph)	ALA(hph)	ALA(hph)	- - - -	131	CYS(-s-)	ILU(hph)	ILU(hph)
23	ASP(-)	ASP(-)	ASP(-)	——	96	ASP(-)	ASP(-)	ASP(-)	——	59	ASP(-)	ASP(-)	ASP(-)	——	132	ASN(NH2)	ASP(-)	ASP(-)
24	GLY( )	GLY( )	GLY( )	——	97	GLY( )	GLY( )	GLY( )	——	60	GLY( )	GLY( )	GLY( )	——	133	ASN(NH2)	GLY( )	GLY( )
25	ASP(-)	ASP(-)	ASP(-)	- - - -	98	ASN(NH2)	ASP(-)	ASN(NH2)	- - - -	61	SER(-OH)	ASP(-)	ASN(NH2)	- - - -	134	ASP(-)	ASP(-)	ASP(-)
26	GLY( )	GLY( )	GLY( )	——	99	GLY( )	GLY( )	GLY( )	——	62	GLY( )	GLY( )	GLY( )	——	135	GLY( )	GLY( )	GLY( )
27	CYS(-s-)	THR(-OH)	THR(-OH)	- - - -	100	TYR(-OH)	LEU(hph)	TYR(-OH)	- - - -	63	THR(-OH)	THR(-OH)	THR(-OH)		136	GLN(NH2)	HIS(N+)	GLN(NH2)
28	ILU(hph)	ILU(hph)	ILU(hph)	——	101	ILU(hph)	ILU(hph)	ILU(hph)	——	64	ILU(hph)	ILU(hph)	ILU(hph)	- - - -	137	VAL(hph)	ILU(hph)	VAL(hph)
29	THR(-OH)	THR(-OH)	THR(-OH)	- - - -	102	SER(-OH)	THR(-OH)	SER(-OH)		65	ASP(-)	ASP(-)	ASP(-)		138	ASN(NH2)	ASN(NH2)	ASN(NH2)
30	THR(-OH)	THR(-OH)	THR(-OH)		103	ALA(hph)	ALA(hph)	ALA(hph)	- - - -	66	PHE(hph)	PHE(hph)	PHE(hph)		139	TYR(-OH)	TYR(-OH)	TYR(-OH)
31	MET(hph)	LYS(+)	LYS(+)		104	ALA(hph)	ALA(hph)	ALA(hph)	- - - -	67	PRO(hph)	PRO(hph)	PRO(hph)		140	GLU(-)	GLU(-)	GLU(-)
32	GLU(-)	GLU(-)	GLU(-)	——	105	GLU(-)	GLU(-)	GLU(-)	——	68	GLU(-)	GLU(-)	GLU(-)	——	141	GLU(-)	GLU(-)	GLU(-)
33	LEU(hph)	LEU(hph)	LEU(hph)	——	106	LEU(hph)	LEU(hph)	LEU(hph)	- - - -	69	PHE(hph)	PHE(hph)	PHE(hph)	——	142	PHE(hph)	PHE(hph)	PHE(hph)
34	GLY( )	GLY( )	GLY( )		107	ARG(+)	ARG(+)	ARG(+)		70	LEU(hph)	LEU(hph)	LEU(hph)	- - - -	143	VAL(hph)	VAL(hph)	VAL(hph)
35	THR(-OH)	THR(-OH)	THR(-OH)		108	HIS(N+)	HIS(N+)	HIS(N+)		71	SER(-OH)	SER(-OH)	THR(-OH)		144	ARG(+)	ARG(+)	GLN(NH2)
36	VAL(hph)	VAL(hph)	VAL(hph)	——	109	VAL(hph)	VAL(hph)	VAL(hph)	- - - -	72	LEU(hph)	LEU(hph)	MET(hph)	——	145	MET(hph)	MET(hph)	MET(hph)
37	MET(hph)	MET(hph)	MET(hph)	——	110	MET(hph)	MET(hph)	MET(hph)	——	73	MET(hph)	MET(hph)	MET(hph)	——	146	MET(hph)	MET(hph)	MET(hph)
38	ARG(+)	ARG(+)	ARG(+)		111	THR(-OH)	THR(-OH)	THR(-OH)		74	ALA(hph)	ALA(hph)	ALA(hph)		147	THR(-OH)		THR(-OH)
39	SER(-OH)	SER(-OH)	SER(-OH)		112	ASN(NH2)	ASN(NH2)	ASN(NH2)		75	ARG(+)	ARG(+)	ARG(+)		148	GLU(-)	ALA(hph)	ALA(hph)
										76	LYS(+)	LYS(+)	LYS(+)	——	149	LYS(+)	LYS(+)	LYS(+)

FIG. 1. Amino acid repeat-domain composition encoded by the *CM1*, protozoan (*T. pyriformis*), and *CL1* genes. Chemical characteristics/functional groups of residues are given in parentheses: hph, hydrophobic; − or +, charge; -s-, sulfhydryl; NH_2, amido; OH, hydroxyl. Nonstandard amino acid abbreviation: ILU, isoleucine. The amino acid similarities (identities and chemical–functional similarities) clearly display the domain (repeat) structure indicative of two very old duplications (1). Solid lines indicate amino acid identities common to six repeats, and broken lines indicate chemical–functional inter-repeat similarities. The *Xenopus* amino acid sequences are identical to the chicken *CL1* sequence (see Fig. 2*a*). The eel calmodulin sequence is identical to the *CL1* sequence with one exception at position 75, where lysine replaces arginine. The subsequence hydrophobic-Asp-Xaa-Asp-Gly-Asp-Gly appears to be the most conserved and is reported to be the Ca^{2+} binding site (2). This differs significantly from the four bacterial subsequence repeats (19) Glu-Xaa-Xaa-Gly-hydrophobic-Asn-Asn-Xaa-hydrophobic-Ser-Ser-hydrophobic-Lys.

calmodulin gene is equidistant from the *CM1* gene and the *CL1* gene. The two chicken calmodulin genes do differ in CpG content (Fig. 2*b*) at least as dramatically as in intron content and amino acid sequence dissimilarity. Curiously, it is the intron-lacking *CM1* locus that is high in CpG, as is the eel locus and one of the *Xenopus* loci, and not the intron-containing chicken calmodulin gene *CL1*.

The CpG dinucleotide is generally found at a suppressed level (by a factor of 2 or 3) in most eukaryotic sequences (12). This suppression is partly due to the known methylation of such pairs and the resulting ease of C→T mutability (13). There is strong evidence that the state of such pairs has been recruited as a regulatory signal (14, 15). For example, the normal CpG suppression is seen in the mammalian β-globin genes but not in the α-globin genes, presumably because of differential regulation (16). It has been noted that the CpG dinucleotides in the α pseudogenes appear to "decay away" with time (16). Therefore, if the intronless chicken gene *CM1* had existed for any significant length of evolutionary time as an inactive pseudogene, the currently observed high CpG content would have to represent selective introduction or reintroduction of these nucleotide substitutions. As mentioned earlier, 6 of the 19 amino acid differences seen in the chicken *CM1* gene introduce seven of the CpG sites not found in the intron-containing chicken gene *CL1*. Thus, it is doubtful that these amino acid replacements were randomly introduced during a prolonged pseudogene existence.

The above CpG characterizations of the various calmodulin genes might be related to another property of most protein-encoding DNA sequences, a preference for pyrimidine/purine codon–codon boundaries over purine/pyrimidine codon–codon boundaries by a factor of 2 or more (12). This preference seems to be maintained with minimal constraint at the amino acid level by using the genetic code's third-base degeneracies; therefore, the percentage of CpG dinucleotides at codon–codon boundaries in the various calmodulins is of interest (Fig. 2*b*). Although the majority of the CpG dinucleotides in the three loci with high CpG content cross codon–codon boundaries, there is no obvious correlation between CpG content and boundary ratio (see Fig. 2 legend). This means that the high CpG content is not a result of a differential preference for pyrimidine/purine over purine/pyrimidine codon–codon boundaries.

One can conservatively estimate the number of CpG-convertible codon–codon boundaries in the eel and *CL1* sequences, assuming that the amino acid sequence of calmodulin is highly constrained. If one defines convertible boundaries as those between a codon starting with G and any 4-fold degenerate or 2-fold degenerate pyrimidine-ending codon sets [or the AGR (R = A or G) codons in the arginine case, which are convertible to CGN codons], there are 45 such convertible sites in these calmodulins. Twenty-eight of these sites are actual locations of the CpG dinucleotide in at least one of the *CL1*, eel, or *CM1* calmodulin sequences (Fig. 2*b*). The chicken *CM1* gene contains seven boundary CpG dinucleotides at common positions with those in the high-CpG eel gene. Thus there is little evidence that the common CpG dinucleotides are ancestral retentions. Rather, some appear to share common position by chance, as would be expected if their numbers rather than the particular positions were being dynamically maintained (16). Therefore, the CpG sequences alone do not explain why the eel calmodulin gene is equidistant from the two chicken calmodulin genes. This is supported by the fact that the eel sequence is roughly equidistant from the two *Xenopus* sequences, which differ greatly in CpG content.

The relationship among the known calmodulin DNA sequences was investigated further by constructing all possible (17) unrooted trees linking the five sequences from chicken, eel, and *Xenopus* with minimal total base substitution branch a

	CL1	*CM1*	Eel	*Xen1*	*Xen2*
CL1		117 (24)	92 (2)	65 (0)	50 (0)
CM1	19		92 (25)	121 (25)	129 (25)
Eel	1	20		88 (2)	87 (1)
Xen1	0	19	1		23 (0)
Xen2	0	19	1	0	
T. pyr.	14	26	15	14	14

b

	CL1	*CM1*	Eel	*Xen1*	*Xen2*
CL1	6/2				
CM1	4	35/19			
Eel	2	11	20/12		
Xen1	3	9	6	18/12	
Xen2	3	4	3	5	6/2

FIG. 2. (*a*) Summary of nucleotide and amino acid replacements among five calmodulin genes (chicken *CL1* and *CM1*, eel, and *X. laevis* genes here abbreviated *Xen1* and *Xen2*). The numbers in the upper right indicate the absolute nucleotide differences between the alleles indexing that cell, with the number in parentheses indicating the number of silent substitutions. The numbers in the lower left indicate the absolute amino acid differences between the alleles indexing those cells. Note that only the amino acid sequence is available for the protozoan (*T. pyriformis*) calmodulin. (*b*) CpG comparison. The single values below the diagonal are the number of CpG dinucleotides shared in common between the alleles indexing those cells. The numbers on the diagonal are the total number of CpG dinucleotides in each allele over the number of CpG dinucleotides crossing the codon–codon boundaries (all of the latter are potential silent mutation sites for C→T transitions).

length. The three shortest trees are shown in Fig. 3. It is significant that none of the trees cluster the two chicken alleles together.

DISCUSSION

Given the above sequence characterizations of the five calmodulin-encoding DNA sequences, a number of genetic histories are plausible. Two of these histories are diagrammed in Fig. 4. In both it has been assumed that the lack of introns is completely indicative of a reverse transcriptase-mediated event. Estimation of the likelihood of the various histories is complicated by the fact that both the chicken and *Xenopus* have two loci, whereas the locus structure of the eel is in some doubt, with only one locus so far identified.

There are two basic schemes that could account for the *CM1* chicken sequence. The first is that an indigenous chicken gene, of either high or low CpG content, was reproduced by a reverse transcriptase event. It has been suggested (3) that accounting for the 19 amino acid differences between the *CM1* and *CL1* chicken gene would require such a reverse-transcribed copy to have spent a considerable length of time (in the evolutionary sense) as a pseudogene. This length of time would be sufficient for CpG decay and the introduction of a number of nonconservative amino acid changes. These genetic events would have to be selectively corrected after reactivation to produce the observed functional high-CpG gene. This history requires no extremely rare events; however, it does involve a large number of genetic events.

The second, alternative history involves the reverse transcriptase viral import of a very distant/divergent calmodulin gene. The biological evidence supporting this scheme is that

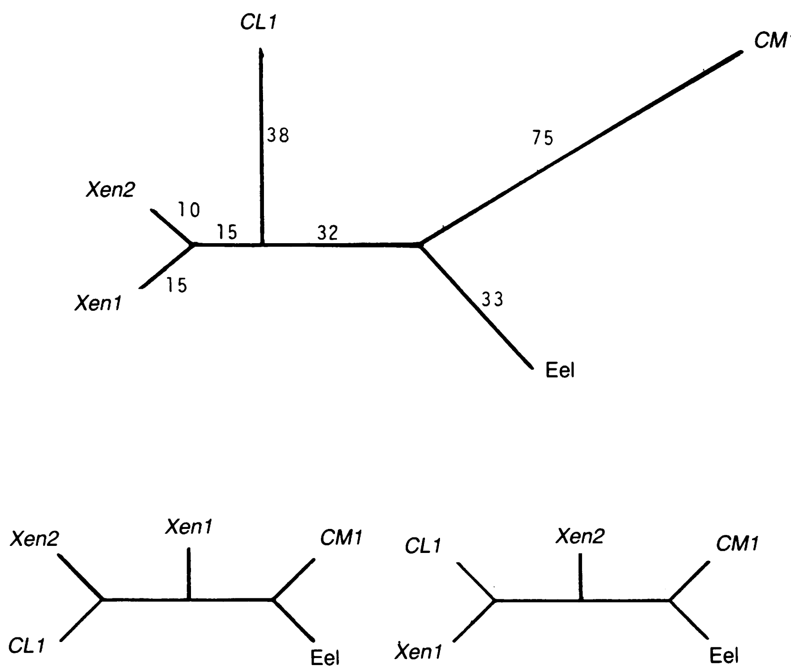

FIG. 3. Minimum-length replacement trees among the five calmodulin genes. (*Upper*) Absolute minimum-length tree (218 replacements) found among all 105 possible trees. (*Lower Left*) Second is a tree with 226 replacements. (*Lower Right*) Third is a tree with 231 replacements. Absolute minimum-length tree is drawn with proportional branch lengths; the others show only the difference in topology.

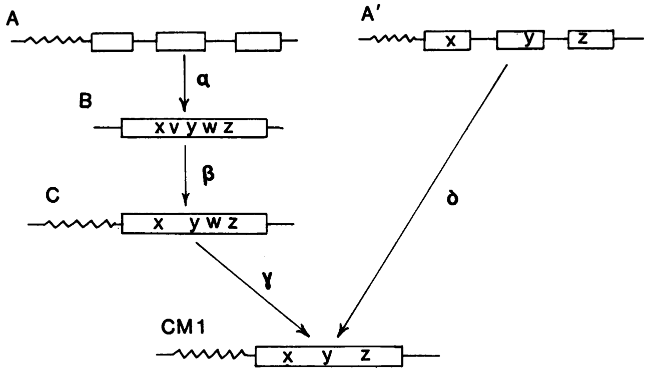

FIG. 4. Diagram of the alternative genetic events composing two possible histories of the intron-lacking chicken calmodulin gene *CM1*. A represents a chicken calmodulin gene, either the predecessor of the contemporary *CL1* or a high-CpG locus now "overwritten" or deleted. B represents an inactive pseudogene containing amino acid replacements x, v, y, w, and z and fewer CpG dinucleotides. A′ is a foreign calmodulin gene containing many of the current *CM1* amino acid differences. α represents a reverse transcriptase-mediated event producing a new pseudogene. δ represents a viral import of a reverse transcriptase mRNA copy. β is the selective activation of the pseudogene followed by (γ) the selective elimination of any deleterious mutations and the introduction or reintroduction of the CpG dinucleotides.

the majority of the observed amino acid differences are conservative and that three of the differences are common to a very distant protozoan calmodulin gene. In addition, if the imported gene was of high CpG content, no additional selection for tissue-specific activation, other than promoter association, might have been necessary. One might even speculate on a viral reverse transcriptase-mediated transfer from a parasitic protozoan. The viral import of a foreign, high-CpG, very old calmodulin gene seems to be compatible with most of the chicken *CM1* data and clearly requires fewer total genetic events. The probability of horizontal gene transfers is obviously small, but there is some evidence pointing to such events during molecular evolution (17, 18). The viral-import scheme is clearly the simplest history, based on the minimal number of required independent genetic events given the current comparative data. The assignment of probability to such an occurrence requires the assessment of the probabilities for each sequential event, and that currently is not possible. The basic problem is how to compare the likelihood of a large number of events of reasonable probability with a few very rare events. Finally, a scan by position of the amino acid differences in Fig. 1 reveals identical amino acids between residues 81 and 126 encoded by the two chicken genes. In addition, all of the common-position CpG dinucleotides between *CM1* and the high-CpG *Xenopus* locus are within this same region. The first observation suggests a possible gene-conversion event between two chicken genes. The second hints that if such a gene conversion had occurred, it may have been between *CM1* and another chicken calmodulin locus of high CpG content, one currently unobserved or deleted as unnecessary. The latter idea would only be supportable if future investigations were to show two and only two calmodulin genes to be the general rule. With the limited data available, the likelihoods of these events cannot be estimated.

The above alternative histories clearly point out the many problems encountered in these studies. In particular, they emphasize the need to identify functional or regulatory contraints as potentially represented in these data by the differing CpG contents. Such analyses need to be performed in addition to standard sequence comparisons and minimal-phylogenetic-tree reconstructions.

Note. Robbins *et al.* (20) also concluded that *CM1* was possibly of viral origin.

We thank John Czelusniak for running the unrooted tree analyses and the Molecular Biology Computer Research Resource for the use of their programs and data bases. We thank Maryellen Ruvolo, Walter Fitch, and Ethan Bier for their suggestions. This work was supported by National Institutes of Health Grants RR02275-01 and GM27055-04 and National Science Foundation Grant BSR83-073336.

1. Goodman, M., Pechere, J. R., Haiech, J. & Demaille, G. (1979) *J. Mol. Evol.* **13,** 331–352.
2. Klee, C. B. & Vanaman, T. C. (1982) *Adv. Protein Chem.* **35,** 213–321.
3. Stein, J. P., Munjaal, R. P., Lagace, L., Lai, E. C., O'Malley, B. W. & Means, A. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6485–6489.
4. Putkey, J. A., Ts'ui, K. F., Tanaka, T., Lagace, L., Stein, J. P., Lai, E. C. & Means, A. R. (1983) *J. Biol. Chem.* **258,** 11864.
5. Yazawa, M., Yagi, K., Toda, H., Kondo, K., Narita, K., Yamazaki, R., Sobue, K., Kakiuchi, S., Nagao, S. & Nozawa, Y. (1981) *Biochem. Biophys. Res. Commun.* **99,** 1051.
6. Lagace, L., Chandra, T., Woo, S. L. C. & Means, A. R. (1983) *J. Biol. Chem.* **258,** 1684.
7. Chien, Y. & David, I. B. (1984) *Mol. Cell. Biol.* **4,** 507.
8. Chou, P. Y. & Fasman, G. D. (1977) *J. Mol. Biol.* **115,** 135–175.
9. Watterson, D. M., Harrelson, W. G., Jr., Keller, P. M., Sharief, F. & Vanaman, T. C. (1976) *J. Biol. Chem.* **251,** 4501–4513.
10. Lukas, T. J., Iverson, D. B., Schleicher, M. & Watterson, D. M. (1984) *Plant Physiol.* **75,** 788–795.
11. Putkey, J. A., Staughter, G. R. & Means, A. R. (1985) *J. Biol. Chem.* **268,** 4704–4712.
12. Smith, T. F., Waterman, M. S. & Sadler, J. R. (1983) *Nucleic Acids Res.* **11,** 2205.
13. Bird, A. P. (1980) *Nucleic Acids Res.* **8,** 1499.
14. Wolf, S. F. & Migeon, B. R. (1985) *Nature (London)* **314,** 467–469.
15. Korba, B. E., Wilson, V. L. & Yoakum, G. H. (1985) *Science* **228,** 1103.
16. Smith, T. F., Ralph, W. W., Goodman, M. & Czelusniak, J. (1985) *Bio. Evol.* **3,** 390–398.
17. Scheller, R. H., Anderson, D. M., Posakony, J. W., McAllister, L. B. & Britten, R. J. (1981) *J. Mol. Biol.* **149,** 15.
18. Benveriste, R. E. & MacIntyre, R. J. (1985) *Molecular Evolutionary Genetics* (Plenum, New York).
19. Inouye, S., Francceschini, T. & Inouye, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6829–6833.
20. Robbins, J., Horan, T., Gulick, J. & Kropp, K. (1986) *J. Biol. Chem.* **261,** 6606–6612.