Differential expression of the three independent CaM genes coding for an identical protein
Potential relevance of distinct mRNA stability by different codon usage
Munk, Mads; Villalobo, Eduardo; Villalobo, Antonio; Berchtold, Martin W.

Published in:
Cell Calcium

DOI:
10.1016/j.ceca.2022.102656

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Review

Differential expression of the three independent CaM genes coding for an identical protein: Potential relevance of distinct mRNA stability by different codon usage

Mads Munk\textsuperscript{a}, Eduardo Villalobo\textsuperscript{b}, Antonio Villalobo\textsuperscript{c,}* , Martin W. Berchtold\textsuperscript{a,}* 

\textsuperscript{a} Department of Biology, University of Copenhagen, 13 Universitetsparken, Copenhagen Ø DK-2100, Denmark
\textsuperscript{b} Facultad de Biología, Departamento de Microbiología, Universidad de Sevilla, Avenida Reina Mercedes 6, Sevilla E-41012, Spain
\textsuperscript{c} Cancer and Human Molecular Genetics Area – Oto-Neurosurgery Research Group, University Hospital La Paz Research Institute (IdiPAZ), Paseo de la Castellana 261, Madrid E-28046, Spain

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Calmodulin genes
Codon usage
Epigenetic control
mRNA control
Mutation-mediated diseases
Phylogeny
Pseudogenes

\textbf{ABSTRACT}

The Ca\textsuperscript{2+}-sensor protein calmodulin (CaM) is a major regulator of multiple cell functions. A unique and puzzling feature of human, and all so far investigated mammals, is the presence of three distinct CaM genes on different chromosomes, which code for identical proteins. How this case of apparent genetic redundancy evolved and why it could be to the advantage of the mammalian organisms is not well established. With a main focus on humans, this article aims to review existing literature addressing how the genes nonetheless differ in function. Clearly, the three CaM genes are differentially expressed in different tissues, during development, in response to different stimuli, and other factors including environmental conditions. As shown in hippocampal neurons, different mRNAs from the CAM genes may even localize differently within the same cell. Regulation of CaM gene expression is achieved by a variety of regulatory elements present in the three genes, including different pro-motor/insulator elements and 3′- and 5′-noncoding regions differing in length and sequence, as well as regulation by epigenetic factors and miRNAs. Here, we hypothesize that predicted differences in mRNA stability and translational efficiency due to divergent codon usage could play an additional regulatory role as the three genes differ markedly in their use of synonymous codons. CALM3, predicted to produce a relatively stable mRNA may be important where the transcription level is low or transiently absent, e.g. during spermatogenesis. In contrast, CALM2 with a predicted much shorter mRNA half-life, may provide better temporal control of CaM levels. Deciphering the underlying mechanisms responsible for all this complexity may help to understand why this unique multigenic arrangement may be an advantage for the optimal spatio-temporal expression of CaM in higher eukaryotes. Finally, we discuss the expression of the CaM genes in selected human pathologies, and how mutations in these genes are responsible for the appearance of serious congenital syndromes, mainly affecting the heart, and although less known, possibly also affecting the functionality of the central nervous system and other organs.

\textbf{Abbreviations}

3′/5′-UTR, 3′- and 5′-untranslated regions;
ALS, amyotrophic lateral sclerosis;
AP1, activator protein 1;
AP2, activator protein 2;
bp, base pairs;
CALM1,2,3/Calm1,2,3, calmodulin-1,2,3 genes;
CaM, calmodulin;

CaML, calmodulin-like protein;
CaMK-II, calmodulin-dependent protein kinase-II;
Ca\textsubscript{v}1.2, L-type voltage-gated Ca\textsuperscript{2+} channel 1.2;
CCR4-NOT, carbon catabolite repression negative on TATA-less;
cDNA, copy DNA;
CDS, coding sequence;
CPEB, cytoplasmic polyadenylation element binding protein;
CPVT, catecholaminergic polymorphic ventricular tachycardia;
CRE, cAMP-response element;

* Corresponding authors.
E-mail addresses: antonio.villalobo@idipaz.es (A. Villalobo), mabe@bio.ku.dk (M.W. Berchtold).

https://doi.org/10.1016/j.ceca.2022.102656
Received 1 July 2022; Received in revised form 1 September 2022; Accepted 25 September 2022
Available online 8 October 2022
0143-4160/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
CREB, cAMP-response element binding protein; 
CSC, codon stabilization coefficient; 
CTCF, CCCCCTC-binding protein; 
DACT3, disheveled binding antagonist of β-catenin 3 gene; 
DACT3-AS1, DACT antisense RNA 1; 
EPcam-DT, epithelial cell adhesion molecule divergent transcript 
lnRNA; 
GAPDHP38, glyceraldehyde 3-phosphate dehydrogenase pseudogene 
38; 
GNGB/GnG8, G protein subunit γ8 gene; 
HEK, human embryonic kidney; 
IP3β, inositol 1,4,5-trisphosphate; 
IVF, idiopathic ventricular tachycardia; 
LINC00642, lncRNA-642; 
LINC02317, lncRNA-2317; 
lnRCNA, long non-coding RNA; 
LQT5, long-QT syndrome; 
Mef2a, myocytes enhancer factor 2a; 
Mir-3558, microRNA-3558; 
MLCK, myosin light-chain kinase; 
NFAT, nuclear factor of activated T cells; 
NMDA, N-methyl-D-aspartate; 
NRDE2/Nrde2, nuclear RNA interference defective-2 gene; 
Pnma8a/b, paraneoplastic Ma antigen protein 8a/b gene; 
Pnnal2, paraneoplastic Ma antigen-like protein 2 gene; 
Ppp5d1p, Ser/Thr protein phosphatase tetraprotein repeat 
domain containing 1 pseudogene; 
PSMC1/PSMC1, 265 proteasome subunit - ATPase 1 gene; 
PTGIR/PTgir, prostaglandin I2 receptor gene; 
qPCR, quantitative polymerase chain reaction; 
RPL21P11, ribosomal protein L21 pseudogene 11; 
Rps6ka5, ribosomal protein S6 kinase a5 gene; 
RyR2, ryanoctide receptor 2; 
siRNA, small interfering RNA; 
SNP, single nucleotide polymorphism; 
SP1, specificity protein 1; 
STPG4/STpg4, sperm-tail PG-rich repeat containing 4 gene; 
TPM, transcripts per million; 
TTCA7A/Trtc7b, tetraprotein repeat domain 7A/7b gene. 
nTPM, normalized transcripts per million

1. Introduction

The evolutionary selection of Ca²⁺ ions as a key signaling element, based on its charge, ionic radius, coordinating and chemical properties, was accompanied by the appearance of proteins holding distinct Ca²⁺-binding motifs. This made it possible to transduce changes in the concentration of free Ca²⁺, first in the interior of protocells and early prokaryotes, and thereafter in the cytosol and lumen of the different intracellular compartments of more complex eukaryotic cells (reviewed in [11]). Among the proteins containing Ca²⁺-binding EF-hand motifs [2] in eukaryotes, calmodulin (CaM) stands out by its ubiquity, high phylogenetic conservation, a high number of targeted proteins under its control and the versatility of cellular functions, in which this protein is involved (reviewed in [3–7]). CaM has two pairs of EF-hands located at each of the N- and C-lobes connected by a flexible linker (reviewed in [8]). Furthermore, due to the ability of each lobe to act independently, CaM has the capacity to bridge different or identical target proteins, leading to dimerization, or interact with distinct regions of the same protein forming new functional domains [9] (reviewed in [8]). Upon Ca²⁺ binding to CaM, a series of hydrophobic patches containing methionine residues are exposed and its 3D-structure undergoes modifications allowing interaction with target proteins, thereby regulating their functions. Nevertheless, CaM is also able to bind to and control some proteins in its Ca²⁺-free form (reviewed in [10,11]).

CaM gene(s) are present in all so far examined eukaryotes, from human to single cell organisms such as yeast. CaM is not only a ubiquitous protein among eukaryotes, but is also extraordinary well conserved among them [12]. This is exemplified by the fact that the protein is identical in all studied vertebrates, even though the genes show remarkable diversity at the nucleotide level [12]. An alignment of CaM sequences from different invertebrates shows only few changes (2 or 3 amino acids), mostly located in the C-terminus, with respect to vertebrate CaM (Fig. 1a). Of interest, the presence, in most cases, of a threonine at position 144 and serine at position 148 in CaM of invertebrates, in contrast to vertebrates, having a glutamine and an alanine, respectively, at these positions, may suggest that these positions in invertebrate CaM could serve as specific phosphorylation targets for Ser/Thr-protein kinases, as it occurs in different Ser/Thr residues of CaM in many organisms (reviewed in [13,14]).

A single CaM gene is present in most non-metazoan eukaryotes, including yeast/fungi [15–20], protozoans [21,22], and green algae [23]. However, there are some exceptions, as for example in Trypanosoma cruzi and T. brucei, which have eight and three tandemly repeated CaM genes, respectively, either in two different loci or in a single one [24,25]. Fig. 1b shows an alignment of the sequences of CaM from different protists and yeast/fungi, as compared with vertebrate CaM. Interestingly, the highest number of substitutions is found in the C-terminal helix of EF-hand IV. Even though Sacharomyces cerevisiae shows the lowest identity (59% with vertebrate CaM), it can functionally be replaced by vertebrate calmodulin [26]. In plants, several CaM isoforms and multiple CaM-like proteins (CaML) encoded by distinct genes are expressed. For example, in rice 5 CaM genes and 32 CaML genes and in Arabidopsis 7 CaM genes and 50 CaML genes are present with differential structural features and specific target recognition properties. We are not considering protist, invertebrate or plant CaMs and CaMLs in this review, as extensive information is available on the properties, functions and evolution of both protein families, particularly in plants (reviewed in [27–29]).

It is well known that many organisms harbor multiple and distinct CaM genes giving rise to identical proteins [30], as for example the three CaM genes present in all so far investigated mammals [31]. The coding sequences for the three human CaM genes (CALM1,2,3) share only about 80% identity at the nucleotide level [12]. Comparing nucleotide coding sequences of CALM1, CALM2 and CALM3 of rat and human reveals similarities of 88, 90 and 90%, respectively. In chicken, 2 CaM genes (CALM1,2) are present with an identity of about 85%, and identities of about 92% with their human counterparts [32,33]. This redundancy of identical proteins being expressed by multiple independent genes is puzzling. Seemingly, this phenomenon may represent, at least in part, a safeguard for cell function and viability in case of mutational mishaps. However, something more is perhaps hidden behind this phenomenon. Evidence is mounting that there is differential expression of the three CaM genes in a given tissue, as for example in the white matter of the rat spinal cord [34], among many other tissues. Furthermore, it has been shown that natural occurring monogenic CaM mutations [35] provoke organ-specific dysfunctions, as shown for the heart. This underscores either that the expression of the single affected CaM gene is the only one expressed in specific regions of the myocardium (see Section 4.1), or alternatively, that the expression of the other CaM genes is not apt to rescue the dysfunction. This may occur either because: (i) the three CaM genes are not expressed in sufficient quantities; (ii) they present extemporal expression timing; (iii) the other expressed wild-type CaM is involved (reviewed in [34–38]); (iv) and/or probably most important, because the arrhythmogenic mutant CaM proteins may work in a dominant negative fashion on target proteins (e.g. RyR2, CaMK-II, SK2 channel, Ca, 1.2 channel) and therefore their function cannot be compensated by wild-type CaM [36–38].
2. From a single EF-hand Ca\(^{2+}\)-binding primordial protein to CaM with four EF-hands

EF-hand motifs, first described in the Ca\(^{2+}\) buffering protein parvalbumin by Kretsinger and Nockolds in 1973 [39], are structural elements found in a large number of proteins of roughly 30 amino acids, containing a loop of 12 residues, providing the Ca\(^{2+}\) coordinating amino residues, between two perpendicularly arranged alpha helices. Most EF-hands are able to bind Ca\(^{2+}\), though with greatly variable affinities. Calcium is bound in a pentagonal bipyramid configuration [40] with oxygen atoms of acidic amino acids and water. As EF-hand proteins have been found in prokaryotes (reviewed in [41]) it is possible that EF-hands have arisen from a single non-eukaryotic organism that gave rise to eukaryotes or that EF-hands have been evolved in early eukaryotes and then were transmitted to prokaryotes [42-45]. Alternatively, it cannot be excluded that EF-hands may have evolved independently in prokaryotes and eukaryotes.

In most cases, two EF-hands, generated by duplication, are placed together in a characteristic way providing enhanced stability and leading to increased Ca\(^{2+}\) binding affinity. A single protein may have several mostly paired EF-hands. Calmodulin, as well as some other proteins including troponin C and myosin light chains, contains 2 pairs of EF-hands distributed in 2 lobes. The fact that domains 1 and 3, as well as 2 and 4, are most similar points to duplication of a single domain with 2 EF-hands, which themselves have evolved from a single one EF-hand domain. Interestingly, the 4 EF-hands in CaM have different structural and functional properties, which is believed to contribute to the huge flexibility of CaM to interact with and regulate a large number (>300) of target proteins.

3. The three mammalian CaM genes

Mammalian CaM is special in that it is encoded by 3 different nonallelic genes, placed on three different chromosomes, all producing an identical protein sequence [30]. In human, for instance, extensive use of synonymous codons means only about 80% identity at the nucleotide level [12]. On the other hand, the individual gene orthologs among different mammals have higher identities. The three CaM genes have the same intron-exon boundaries, which are also conserved in some non-mammalian species. Chicken and frog lack the CALM3 gene, and only have two CALM genes, which are orthologs to mammalian CALM1 and CALM2, indicating that CALM3 was either developed at a later evolutionary step, or that it was present but lost again in these species. Having three different promotors and 5' non-coding RNA sequences, the three CaM genes of mammals are differentially regulated during development (reviewed in [46]), and their mRNAs were shown to be
differentially localized in neuronal cells.

3.1. CaM genes and pseudogenes. Comparison of their genomic organization: intron-exon organization, chromosomal localization

3.1.1. CaM genes

The chromosomal loci of the three human CaM genes were first reported in 1993 [31] to be on chromosomes 14, 2 and 19 for CALM1, CALM2 and CALM3, respectively (Table 1), as analyzed by in situ hybridization of metaphase spreads of human lymphocytes, and confirmed with the precise physical localization after the human genome was sequenced [47–49]. CALM1 and CALM3 are located on the forward DNA strand while CALM2 is placed on the reverse strand [47–50] (see Fig. 2). The CALM2 gene is the longest of the three CaM genes [51], and all human CALM genes (as well as all known mammalian CaM genes) have 6 exons and 5 introns (see Table 1). Of interest and supporting a common evolutionary history, the intron/exon boundaries with respect to the coding sequences are conserved, not only in human CALM1, CALM2, and CALM3 [52], but also in other mammals, as well as in the two chicken CaM genes. Table 2 summarizes the sequences of the 10 intron/exon boundaries in the CaM genes of human, rat, mouse and chicken, showing that all of them are conserved. The internal exons 2–5, containing only coding sequences, are identical in length, while exons 1 and 6, with mostly noncoding sequences, are different in length (see Table 1). Exon 1 includes the 5′-UTR as well as the ATG start codon at its 3′-end, while exon 6, includes the 10 C-terminal codons as well as the 3′-UTR sequence (see Table 1).

The location of the orthologous genes upstream and downstream of the three CaM genes in human, rat and mouse is similar (see Fig. 2). The genes PSMC1/Psmc1 and NRDE2/Nrde2, encoding the 26S proteasome subunit-ATPase 1 and the nuclear exosome regulator NRDE2 (necessary for RNA interference-2), are located upstream of CALM1/Calm1 in human and rat, respectively. The genes STPG4/Stpg4 and Ttc7b, coding for the sperm-tail PG-rich repeat containing 4 protein and the tetrapeptide repeat domain 7b protein are located downstream of CALM1 in human rat and mouse, respectively. The gene encoding the 7a isoform of the latter protein is located upstream of CALM2/Calm2 in human and mouse, while it is placed downstream in rat. Moreover, the genes PTGIR/Ptgir and GNG8/Gng8, coding the prostaglandin I2 receptor and the G protein subunit g8, respectively, are located downstream of CALM3 in human, and upstream of Calm3 in rat and mouse. The occurrence of synteny, in the chromosome regions where the CaM genes in the three species are located, reinforces the idea of the common evolutionary history, as stated above. Whether regulatory elements in the mentioned genes located in the genomic regions close to the CaM genes have an influence on the transcriptional activities of the 3 CaM genes is unknown.

Even though mammalian CaM genes encode identical proteins, as described above, the nucleotide sequence varies quite markedly [53]. As an example, Fig. 3a shows extensive differences in the nucleotide coding sequence of the three CaM genes in human, presenting more than one hundred nucleotide differences, almost exclusively from differences in the third codon position. Similar numbers of variations are also found in rat (Fig. 3a) and mouse (not shown) and mostly in the same positions. In contrast, only less than half of the variations are found when comparing the nucleotide coding sequences of the three CaM genes in human versus rat and mouse.

Even though mammalian CaM genes encode identical proteins, as described above, the nucleotide sequence varies quite markedly [53]. As an example, Fig. 3a shows extensive differences in the nucleotide coding sequence of the three CaM genes in human, presenting more than one hundred nucleotide differences, almost exclusively from differences in the third codon position. Similar numbers of variations are also found in rat (Fig. 3a) and mouse (not shown) and mostly in the same positions. In contrast, only less than half of the variations are found when comparing the nucleotide coding sequences of the three CaM genes in human versus rat and mouse.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Human calmodulin genes and miRNAs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Properties</td>
<td>CALM1</td>
</tr>
<tr>
<td>Chromosome locus</td>
<td>14q24-q31 (14q32.11)</td>
</tr>
<tr>
<td>Gene length (bp)</td>
<td>11,240</td>
</tr>
<tr>
<td>Exons length (bp)</td>
<td>11,240</td>
</tr>
<tr>
<td>Exons length (bp)</td>
<td>11,240</td>
</tr>
<tr>
<td>Introns length (bp)</td>
<td>11,240</td>
</tr>
<tr>
<td>Introns length (bp)</td>
<td>11,240</td>
</tr>
<tr>
<td>Promoter</td>
<td>4</td>
</tr>
<tr>
<td>TATA box</td>
<td>4</td>
</tr>
<tr>
<td>CAAT box</td>
<td>4</td>
</tr>
<tr>
<td>GC box</td>
<td>1</td>
</tr>
<tr>
<td>AGGGA motif</td>
<td>1</td>
</tr>
<tr>
<td>CRE bind. site</td>
<td>1</td>
</tr>
<tr>
<td>AP1 bind. site</td>
<td>1</td>
</tr>
<tr>
<td>AP2 bind. site</td>
<td>1</td>
</tr>
<tr>
<td>Poly-A signals</td>
<td>(5)</td>
</tr>
<tr>
<td>(number, sequences)</td>
<td>5′-AATAA-3′</td>
</tr>
<tr>
<td>5′-TATATA-3′</td>
<td>5′-TATATA-3′</td>
</tr>
<tr>
<td>5′-TATATA-3′</td>
<td>5′-TATATA-3′</td>
</tr>
<tr>
<td>CPEB-binding sites</td>
<td>1</td>
</tr>
<tr>
<td>(number, sequences)</td>
<td>5′-UUUUUAU-3′</td>
</tr>
<tr>
<td>Potential CTCF-binding sites</td>
<td>1</td>
</tr>
<tr>
<td>(sequence)</td>
<td>5′-GAGACTGCT-3′</td>
</tr>
<tr>
<td>miRNA let7a binding to 3′-UTR</td>
<td>5′-UGGCCUC-3′</td>
</tr>
<tr>
<td>Stau2 binding</td>
<td>Located in a retained intron within the 3′-UTR</td>
</tr>
</tbody>
</table>

1. The potential CTCF-binding sites shown were determined using the CTCFBSDB 2.0 prediction tool. For each gene, the sequence shown corresponds to that with the highest score in the coding orientation from the output reported. bp, base pair; CPEB, cytoplasmic polyadenylation element binding protein; CTCF, CCCTC-binding protein.
rat, mouse or cow (see Fig. 3b). This shows that there are more differences in the three CaM genes among them in a given species than in a given CaM gene among different vertebrates.

3.1.2. CaM pseudogenes

The lack of introns, truncation and mutational disruption of the CDS are the main characteristics of eukaryote pseudogenes (e.g. [52,54,55]), although these characteristics do not always apply. Pseudogenes are widely distributed in the genome of many living organisms, and despite early assumptions of lack of function, a growing number of evidences support the functionality of some of them (reviewed in [56–58]). Some of the mechanisms by which pseudogenes could have a functional role include: (i) they could encode a sequence-altered or truncated functional protein reminiscent of the parental one; (ii) they could be transcribed into siRNAs or antisense RNAs, preventing translation of the parental gene, or yielding lncRNAs, which directly interact and regulate the expression of certain proteins; and (iii) they could favor DNA folding allowing interaction of different chromatin regions, and this could positively or negatively affect the expression of nearby genes, and to allow the conversion of their parental genes into dysfunctional ones by transferring segments via non-allelic recombination (reviewed in [57,58]).

Pseudogenes derived from the CaM genes have been identified in multiple vertebrates, including human [59–61]. Among them, a couple of presumably non-functional intronless pseudogenes, derived from the genuine human CALM1, denoted CALM1P1 and CALM1P2, presenting disruptive mutations in their CDS were first described by Rhyner et al. [52], as well as three retro-pseudogenes derived from the human CALM2 gene [62]. Alu elements, repetitive sequences that are cut by the restriction endonuclease Alu, were identified upstream of the 5′-UTR and downstream of the 3′-UTR in one of these pseudogenes [62]. Alu elements are dispersed by retro-transposition. Transcription of Alu elements, reverse transcription of their RNA and insertion of the resulting cDNA into the targeted genes are the three steps required in the dispersion process, usually leading to multiple effects by modifying functions of other genes [63]. The retro-pseudogenes derived from...
The transcripts in the databases Ensemble (https://www.ensembl.org/index.html) or NCBI (https://www.ncbi.nlm.nih.gov/) used in the analysis were as follows:

From human: CALM1, ENSMUST00000356978.9; CALM2, ENSMUST00000272298.12; and CALM3; ENSMUST00000259129.14; from rat: Calm1, ENSRNOT00000006479.2; Calm2, ENSRNOT00000022603; and Calm3, GenBank X14265.1; from mouse: Calm1, ENSMUST00000110082.11; Calm2, ENSMUST00000040440.7; and Calm3, ENSMUST00000019514.10; and from chicken: Calm1, ENSGALT00000045488.3; Calm2, ENSGALT00000016279.5. Exon sequence segments are highlighted in capital letters. The site of splicing in the codon and the amino acid involved are shown in parenthesis.

### Table 2

<table>
<thead>
<tr>
<th>Species</th>
<th>Intron/exon boundary</th>
<th>CALM1</th>
<th>CALM2</th>
<th>CALM3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>ATGgttagg... (+3 M1)</td>
<td>ATGgttagg... (+3 M1)</td>
<td>ATGgttagg... (+3 M1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tagagCGTGA</td>
<td>...tagagCGTGA</td>
<td>...tagagCGTGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGCGTgtaag... (+1 E12)</td>
<td>TGCGTgtaag... (+1 E12)</td>
<td>TGCGTgtaag... (+1 E12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGGATgtaag... (+1 G60)</td>
<td>TGGATgtaag... (+1 G60)</td>
<td>TGGATgtaag... (+1 G60)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tagagGTAAT</td>
<td>...tagagGTAAT</td>
<td>...tagagGTAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACAAGTaatg... (+3 K95)</td>
<td>ACAAGTaatg... (+3 K95)</td>
<td>ACAAGTaatg... (+3 K95)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...cagaGATGG</td>
<td>...cagaGATGG</td>
<td>...cagaGATGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGAAGTaatg... (+1 E141)</td>
<td>TGAAGTaatg... (+1 E141)</td>
<td>TGAAGTaatg... (+1 E141)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>ATGgttagg... (+3 M1)</td>
<td>ATGgttagg... (+3 M1)</td>
<td>ATGgttagg... (+3 M1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tagagCGTGA</td>
<td>...tagagCGTGA</td>
<td>...tagagCGTGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGCGTgtaag... (+1 E12)</td>
<td>TGCGTgtaag... (+1 E12)</td>
<td>TGCGTgtaag... (+1 E12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGGATgtaag... (+1 G60)</td>
<td>TGGATgtaag... (+1 G60)</td>
<td>TGGATgtaag... (+1 G60)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tagagGTAAT</td>
<td>...tagagGTAAT</td>
<td>...tagagGTAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACAAGTaatg... (+3 K95)</td>
<td>ACAAGTaatg... (+3 K95)</td>
<td>ACAAGTaatg... (+3 K95)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...cagaGATGG</td>
<td>...cagaGATGG</td>
<td>...cagaGATGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGAAGTaatg... (+1 E141)</td>
<td>TGAAGTaatg... (+1 E141)</td>
<td>TGAAGTaatg... (+1 E141)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>ATGgttagg... (+3 M1)</td>
<td>ATGgttagg... (+3 M1)</td>
<td>ATGgttagg... (+3 M1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tagagCGTGA</td>
<td>...tagagCGTGA</td>
<td>...tagagCGTGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGCGTgtaag... (+1 E12)</td>
<td>TGCGTgtaag... (+1 E12)</td>
<td>TGCGTgtaag... (+1 E12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGGATgtaag... (+1 G60)</td>
<td>TGGATgtaag... (+1 G60)</td>
<td>TGGATgtaag... (+1 G60)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tagagGTAAT</td>
<td>...tagagGTAAT</td>
<td>...tagagGTAAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACAAGTaatg... (+3 K95)</td>
<td>ACAAGTaatg... (+3 K95)</td>
<td>ACAAGTaatg... (+3 K95)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...cagaGATGG</td>
<td>...cagaGATGG</td>
<td>...cagaGATGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGAAGTaatg... (+1 E141)</td>
<td>TGAAGTaatg... (+1 E141)</td>
<td>TGAAGTaatg... (+1 E141)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td>...tcagATTTC</td>
<td></td>
</tr>
</tbody>
</table>

### 3.2. Regulation of CaM expression

Since CaM is involved in a huge variety of physiological functions, even within a single cell, it is of no surprise that a certain level of CaM expression is not only required in a specific tissue and cell with a given set of CaM targets of different abundance, locations and affinities for CaM, but also in different subcellular compartments. It has been shown in living smooth muscle cells that only 5% of the total CaM concentration is freely diffusible [69] and that, upon increase of cytosolic Ca²⁺, CaM could be translocated to the nucleus. This has been confirmed in cardiomyocytes where soluble free CaM was found in very low concentration (ca. 1% of total) and increasing intracellular Ca²⁺ led to translocation of CaM to the nucleus as shown by Bers and collaborators [70]. It is believed that the presence of CaM binding sites exceeds the amount of available CaM, making diffusion of free CaM difficult and indicating that there must be a selection of targets based on affinities and kinetics [71]. Therefore, subcellular CaM pools could play an important role in target regulation [72].

Another factor regulating the available CaM concentration is the presence of CaM targets that bind in the absence of calcium, e.g. GAF-3/neuromodulin, which has been proposed to act as CaM buffer [73] to prevent Ca²⁺ dependent signaling through CaM. In addition to this CaM storage hypothesis, it has been proposed that CaM could be temporally “masked” and thereby not anymore able to regulate targets by posttranslational modification, e.g. phosphorylation is shown to reduce its activity in regulating targets (reviewed in [13,14]). Moreover,
as CaM contains two Ca\textsuperscript{2+} binding lobes, each with two EF-hands with different Ca\textsuperscript{2+} affinities and binding kinetics that could act in a somewhat independent way, this may contribute to its extraordinary flexibility in Ca\textsuperscript{2+}/CaM-dependent signaling.

The available information on the presence and especially the significance of cis-regulatory elements controlling the differential expression of the three human CaM genes and their alternatively polyadenylated mRNAs is rather scanty. Obviously, a major obstacle is the fact that it is not possible to discriminate on the level of protein accounted for each gene namely, which CaM gene is responsible for the final protein product as CaM from all three CALM genes are identical.

Whether and how these potential regulatory elements are used by the relevant transcription factors and other regulatory proteins as well as RNAs and whether there is cross-talk among these elements to facilitate or repress the relative expression of a particular CaM gene and its mRNA product over the others, depending on different physiological states of the cell, is only very fragmentarily known so far. As work on CaM gene regulation carried out before and around 2000 has been reviewed by the group of Strehler [12], we will only discuss some of the major conclusions and focus on a few more recent examples of cis-regulatory elements and their functions, as well as cases of epigenetic control of the CaM genes in human and other organisms and list and compare these potential regulatory elements in table form.

### 3.2.1. Promoters, insulators

The promoter regions of the three human CaM genes contain a variable number of regulatory elements, such as TATA, CAAT, GC boxes; AGGGA and CRE motifs, transcription factor-binding sites (e.g. for SP1, AP1, AP2, CREB), as shown in Table 1. This may impinge on the relative efficiency of their transcriptional activity [74]. The function of insulators in vertebrates is mediated upon binding to CTCF, an eleven zinc finger DNA-binding domain protein, which could interfere with the communication between promoters and enhancers or by connecting two CTCF binding sites thereby mediating chromatin looping (reviewed in [75]). Of interest, potential CTCF-binding sites have been identified in all three human CaM genes (see Table 1). Possible functional

---

**Fig. 3.** a.b. Alignment of nucleotide sequences of the three human and other vertebrate CaM genes. Alignment of the three human CALM genes (a), and alignment of each gene compared to the orthologs from four species (b). Nucleotide sequences were retrieved from NCBI database under the following accession numbers: NM_006888.6, NM_001743.6, and NM_005184.4 for human CALM1, CALM2 and CALM3, respectively; NM_031969, NM_017326.3, and BC063187.1 for rat Calm1, Calm2 and Calm3; NM_001242572.1, NM_001242587.1, and NM_001046249.2 for Cow Calm1, Calm2 and Calm3, and NM_001319341.1, NM_007589.5, and NM_007590.3 for mouse Calm1, Calm2 and Calm3. Only coding sequences were aligned with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) using default settings. The start and stop codons are included in the alignments as well as the nucleotide position numbers. Non-identical nucleotides are highlighted in red. The third position in each codon is marked in red=AT3 or green=GC3.
significance of the cAMP response element found in human, rat, and the Xenopus CaM II promoter has been supported by findings showing that cAMP upregulates CaM I and CaM II in rat PC12 cells [76].

It was demonstrated that a 1218 bp segment of the promoter and the 5’-leading sequence of the chicken Calm2 gene containing CpG islands, typical for housekeeping genes and in many cases not containing TATA boxes, was sufficient for the initiation and regulation of transcription [77]. Gene promoters may contain single or multiple dispersed transcriptional start sites (reviewed in [78]), and in mammals CpG island-containing promoters usually are associated with dispersed transcriptional start sites and housekeeping genes (reviewed in [79]). The above mentioned chicken Calm2 promoter was shown to have lower activity in differentiated cells, as compared to proliferating ones, in agreement with the lower endogenously-produced CaM level observed during differentiation [77]. As the chicken Calm2 promoter shares similarities with mammalian CALM2 promoters, it could be predicted that it works in a similar way as the mammalian promoter. Similar studies were conducted using promoter segments of the rat Calm2 and Calm3 genes in subsequent studies. In these cases, transgenic mice expressing constructs of promoter segments with the β-galactosidase gene as a reporter were used to monitor the expression in different regions and cell types of the central nervous system. In most but not all cases, β-galactosidase expression matched the expression of the endogenous CaM genes in adult mice [80,81]. Additional studies were performed in testis, specifically demonstrating Calm2 expression in spermatocytes, and not in other cellular stages during spermatogenesis, when Calm1 and Calm3 are expressed [82].

3.2.2. Epigenetic control

Epigenetic labeling of DNA and histones plays a central role in the pattern of gene transcription during development, differentiation and the physiological status of the cell [83–86]. Of interest, the cytosine methylation level of human CALM1 in spermatocytes progressively decreases from 25 to 60 years of age, as observed with other genes (e.g. CDH13 and STMN2 encoding T-cadherin and stathmin-2, respectively) [87], suggesting a differential expression level of CaM during aging. Also, a prospective study in a rather small cohort of type 2 diabetes patients demonstrated hypomethylation of CpG islands in the promoter of CALM2 in peripheral leukocytes, contradicting the hypothesis that methylation of this gene contributes to the onset of type 2 diabetes as previously suspected [88].

3.2.3. Tissue and cell type specific expression of the three mammalian CaM genes

CaM is expressed in all cells of eukaryotes in variable amounts. To determine which CaM gene is expressed in each case in a specific cell type and physiological condition may shed light on how the three CaM genes are contributing to the function of the cell or tissue. Expression of the individual CaM genes can only be measured on the transcript level, because the proteins coded by the three genes are identical. Therefore, transcription data cannot be directly linked to the levels of CaM protein. Human CALM1 expression was detected in all organs tested including heart, placenta, lung, and kidney, though it was found to be particularly abundant in brain and skeletal muscles, and very scarce in liver and pancreas [52]. CALM2 also presents very high expression in human brain, followed by similar expression levels in heart, placenta, lung, liver, skeletal muscle and kidney, and in much lesser extent in pancreas [51]. In the heart, the expression of different CaM genes is not homogeneous through its different tissue structures. Thus, in the Purkinje fibers, specialized to conduct the electrical stimuli to the left and right heart ventricles, the expression of CALM1 is higher than in the working atrial and ventricular non-specialized cardiomyocytes, while the expression of CALM3 is higher in both the left and right epicardium than in the endocardium [89]. This suggests that each cardiac structure may require a different quantitative and/or temporal output of CaM for their correct functioning. The specific downregulation of only one of the three CaM genes could give some hints on the functionality exerted on the cell. Several reports have been published in this context [90,91]. However, these studies do not specifically address whether the knocked-down CaM gene was the only one regulating the cell function under study, but rather that the supply of CaM derived from the untouched genes is unable to fully compensate for the observed deficiencies. This could be due to lack of sufficient tempo-spatial supply of CaM to the targeted CaM-binding proteins.

3.2.4. Differential expression of CaM induced by signaling processes, during cell differentiation and development

The expression of the CaM genes has been shown to be dependent on the physiological conditions of the cell. For instance, the half-life of CALM1 and CALM3 mRNAs is longer than the one of CALM2 mRNA in mouse myoblasts during both proliferation and differentiation, while the half-life of all three CALM mRNAs increases during differentiation, but the actual level of these transcripts decreases, as well as the CaM level, although with distinct kinetics [92]. Calm expression is also subjected to endocrine control. Thus, it has been shown that resection of the adrenal glands decreased the level of CALM3 mRNA in the rat cerebral cortex and the hippocampus, without affecting CALM1 and CALM2 mRNAs, while corticosterone administration prevented CALM3 down-regulation as expected [93]. Evidence that differential transcription of the CALM genes in rats provide a means to control the spatial and temporal availability of CaM pools was reviewed in [12]. The CALM mRNAs exhibit cell type and cell state-specific mRNA abundance. In rats CALM1 and CALM2 appear in neurites and dendrites, while CALM3 predominates in the cell body. Humans express a long CALM1 transcript (4.1 kb) as well as a shorter 1.8 kb version. The longer CALM1 (4.1 kb) human mRNA has been demonstrated to have increased stability. It was suggested that this protects the transcripts from degradation during trafficking to distal locations (reviewed in [12]). We searched for evidence of differential expression of the two CALM1 RNA isoforms, in the RNA sequence database GTExPortal (gtexportal.org). It appears that the 4.1 kb version is more specific to the brain (Fig. 4). The shorter version appears to be the most abundant in all tissues sequenced, except in brain, where the long form dominates. This may reflect the need for

![Fig. 4. RNA isoforms from alternative polyadenylation in CALM1. RNA-sequencing reads from representative tissues. The figure shows RNA sequencing results of CALM1. The 3' UTR starts at position 0.6 Kb of the transcript. Evidence is seen of reverse transcriptase activity starting from d-TTT primers at position 1.8 Kb (red arrow a) and at position 4.1 Kb (red arrow b).](https://www.gtexportal.org/j).
stable CALM mRNA when trafficking to distal dendrites in brain.

3.2.5. 5′-UTR and 3′-UTR

The 5′-UTR of mRNAs may be involved in transcriptional or translational regulation whereas the 3′-UTR of mRNAs may act, among other functions, in transport and location of the mRNA to specific intracellular locations where translation is required, as well as controlling mRNA stability by promoting or inhibiting its degradation. In mammals, both UTRs are unique in the three CaM mRNAs, presenting highly conserved sequence homology (76–95%) among human, rat and mouse, and 85–96% between human and several cetaceans.

The importance of the 5′-UTRs was demonstrated in the three human CaM genes in experiments testing transcriptional activity with the use of the luciferase reporter gene in teratoma cells, as its absence strongly decreased transcription [51] and discussed in Toutenhoofd and Strehler [12]. The least conserved 3′-UTR is found in the CALM3 transcript (79.7% and 78.2% comparing human with rat and mouse, respectively), while the most conserved corresponded to the CALM2 transcript (90.2% and 92.7% comparing human with rat and mouse, respectively), closely followed by the CALM1 transcript (89.3% and 86.2% comparing human with rat and mouse, respectively) [94,95]. This is close to the somewhat higher conservation (91.1–92.7%) of the CDSs of the three CaM genes comparing human with rat and mouse transcripts indicating an important function of the 3′-UTR in CALM1 and CALM2 mRNAs (see Table 1). The CPEB protein, first identified in Xenopus oocyte, binds to the 3′-UTR at specific sites and regulates mRNA polyadenylation and translation (reviewed in [96]). CPEB activity is regulated by phosphorylation with diverse protein kinases, including CaMK-II [97,98]. Nevertheless, CPEB phosphorylation does not affect RNA binding, a process requiring Zn2+, and which is mediated by two RNA recognition motifs and the Cys-His-rich C-terminus of CPEB needed for interaction with the consensus sequence 5′-UUUUUAUU-3′ [98] (see Table 1).

The 3′-UTR of the human CALM3 transcript contains three polyadenylation sites, accounting for the identified 0.9, 1.9 and 2.3 kb UTRs are unique in the three CaM mRNAs, presenting highly conserved stability by promoting or inhibiting its degradation. In mammals, both locations where translation is required, as well as controlling mRNA translational regulation whereas the 3′-UTR of mRNAs may act, among other functions, in transport and location of the mRNA to specific intracellular locations where translation is required, as well as controlling mRNA stability by promoting or inhibiting its degradation. In mammals, both UTRs are unique in the three CaM mRNAs, presenting highly conserved sequence homology (76–95%) among human, rat and mouse, and 85–96% between human and several cetaceans.

The importance of the 5′-UTRs was demonstrated in the three human CaM genes in experiments testing transcriptional activity with the use of the luciferase reporter gene in teratoma cells, as its absence strongly decreased transcription [51] and discussed in Toutenhoofd and Strehler [12]. The least conserved 3′-UTR is found in the CALM3 transcript (79.7% and 78.2% comparing human with rat and mouse, respectively), while the most conserved corresponded to the CALM2 transcript (90.2% and 92.7% comparing human with rat and mouse, respectively), closely followed by the CALM1 transcript (89.3% and 86.2% comparing human with rat and mouse, respectively) [94,95]. This is close to the somewhat higher conservation (91.1–92.7%) of the CDSs of the three CaM genes comparing human with rat and mouse transcripts indicating an important function of the 3′-UTR in CALM1 and CALM2 mRNAs (see Table 1). The CPEB protein, first identified in Xenopus oocyte, binds to the 3′-UTR at specific sites and regulates mRNA polyadenylation and translation (reviewed in [96]). CPEB activity is regulated by phosphorylation with diverse protein kinases, including CaMK-II [97,98]. Nevertheless, CPEB phosphorylation does not affect RNA binding, a process requiring Zn2+, and which is mediated by two RNA recognition motifs and the Cys-His-rich C-terminus of CPEB needed for interaction with the consensus sequence 5′-UUUUUAUU-3′ [98] (see Table 1).

The 3′-UTR of the human CALM3 transcript contains three polyadenylation sites, accounting for the identified 0.9, 1.9 and 2.3 kb UTRs [94]. The close homology of these UTRs denotes close phylogenetic relationship as well. The 3′-UTRs are unique in the three CaM mRNAs, presenting highly conserved AU-rich elements, where miR-4661-3p is expected to bind. Suppression of miR-4661-3p by peptide nucleic acid inhibitors lowered CaM levels and suppressed neuroinflammation mediated by microglial cells [106].

3.2.7. Differential subcellular localization of CaM mRNAs

Establishment of local mRNA pools, which could be far away from the cell body in neurons, is believed to contribute to the post-transcriptional regulation leading to temporal and spatial fine-tuning of CaM gene expression. The initial discovery of specific subcellular localization for several CaM mRNAs in differentiated PC12 cells in 1993 by Zhang et al. [107] showed that CALM1 and CALM2 mRNAs mostly localized to the neurites and that CALM3 mRNA was positioned in the cell body or close to it.

The group of Gulya investigated the differential subcellular localization of the different mRNAs of the 3 CaM genes in neuronal rat tissue, which contains the highest CaM concentration in the mammalian body (reviewed in [108]). The presence of different patterns of mRNAs derived from the same CaM gene, detected on Northern blots from different tissues or cell populations as well as degradation sequence elements in certain transcripts of the same CaM gene but not in others, pointed to posttranscriptional regulation on the level of mRNA stability due to differential polyadenylation (see also chapter 3.2.8 on codon usage).

Targeting specific mRNA to dendritic sublocalization, where CaM mRNA is colocalized together with other transcripts, including alpha CaMKII, and assumed to be translated there, has been shown in a number of articles indicating an increased need of CaM signaling for postsynaptic activities (reviewed in [108]) and [109]. Differential transport of mRNAs from the 3 CaM genes to functional different compartments in neurons and glia cells of the rat hippocampus [110] as well as in the midbrain-brain stem region has been shown by the Gulya group [111].

The intracellular distribution of the mRNAs derived from the 3 CaM genes also varies in isolated mouse ventricular cardiomyocytes. It was shown that the mRNA from the CALM2 gene is more concentrated in the central part of the cell, close to the nucleus, and the CALM3 mRNA is concentrated in the cell periphery, while the Calm1 mRNA is mostly located between the other two [112]. It is possible that the differential intracellular distribution of the three mRNAs is correlated with their subcellular specific localization of the mRNA in the cytoplasm. Moreover, treatment of animals with isoproterenol, an agonist of adrenergic-β receptors, enhanced the level of Calm3 mRNA, but not Calm1 and Calm2 mRNAs, coinciding with enhanced overlap of this mRNA with Ryr2 mRNA at the intercalated discs [112]. This is in agreement with the expected increase of Ca2+ mobilization required for heart contraction during sympathetic stimulation, and therefore suggestive of a specialized role of a given CaM gene for particular cell functions.
Saucerman and Bers [113] added further evidence of local regulation of Ca\(^{2+}\) signaling depending on the presence and properties of Ca\(^{2+}\)-binding proteins. They hypothesized that based on the known properties of three important cardiac myocyte channels, the L-type calcium channel, the ryanodine, and the IP\(_3\) receptors, there exists two functional different pools of CaM, one of them containing CaM permanently attached to the channels, called “dedicated” CaM, and another one consisting of “promiscuous” CaM sensing local Ca\(^{2+}\). It appears that local Ca\(^{2+}\) concentrations may be much higher than the overall cytosolic levels, as the affinity of CaM for Ca\(^{2+}\)-binding is quite low (\(K_a\) ca. 5 mM) further indicating that relatively high CaM concentrations possibly in the apo form sequenced by “storage” CaM-binding proteins, such as GAP43/neuromodulin, must be readily available at distinct places in the cell and may be differentially regulated by the 3 CaM genes.

A general concept, as a result of these investigations, has been put forward by Palifi et al. [109] that in the rat brain CaMII and CaMI are the genes that are mostly regulated in response to a variety of signals and stimuli in contrast to CaMIII that is rather considered a housekeeping gene, even though there is no clear distinction and most of the data about differential CaM gene expression derives from neuronal studies. This partial specificity of the expression of the 3 CaM genes enabling the control of local CaM concentration in microdomains, which is important for its availability to interact with targets with high accuracy, may explain at least in part the evolutionary maintenance of the 3 genes.

3.2.8. Codon usage affecting RNA stability and translational efficiency

mRNA stability has long been considered as a factor influencing the final protein production from a gene. cis-regulatory elements in the 5′-UTR, which can bind proteins and miRNAs, and destabilize the mRNA or repress translation, are covered (Section 3.2.6.), as features, which separate the different transcripts from human CALM1,2,3. Here, we will focus on the codon usage bias (CUB) of the three CALM coding regions, and relate these to (predicted) differences in stability. We postulate that the CUBs may have evolved as a means for the cell and organism to regulate the CaM production required depending on time, location, and physiological conditions.

It has been known that the availability of tRNAs has an effect on mRNA stability, but no well-defined way to predict stability from RNA sequences has been described before 2015. This changed when Presnyak et al. [114], studying yeast codon usage, introduced the codon stabilization coefficient (CSC) as a way to give each codon a score, which links the codon usage of mRNA to its half-life. If a codon is overrepresented in stable mRNAs the CSC score is positive, and vice versa the CSC score is negative for codons overrepresented in unstable mRNAs. Several studies have since then expanded on this, so that we now also have CSC scores for zebrafish, Xenopus [115], and several human cell types [116,117]. The CSC scores are based on experimental data, in which RNA-seq is done at different time points after transcription was blocked by actinomycin D. Experimental approaches vary, and can relate to endogenous mRNA or to mRNA expressed from lentiviral infection [117].

In animals, CSC appear to be somewhat shared between different species, but not with yeast [118]. In addition, the CSC score is not an entirely stable factor in an organism, one reason being that tRNA values can vary in different cell types of the same organism [117]. Interestingly, genes with similar function appear to coordinate their mRNA stabilities in part by similar codon usage [119].

It has been proposed that synonymous codons are translated at different rates, and that slower translation can stimulate mRNA decay [116]. This appears to be linked to longer dwell times at the amino acid entry (binding) site of the ribosome during translation. This slowdown is sensed by the decay machinery leading to deadenylation by the CCR4-NOT complex and eventual decapping of the 5′-end. Hia et al. [120] proposed that interleukin enhancer binding factor 2 (ILF2) mediates mRNA stability by binding to transcripts with low GC3 content (that is, a G or C at the wobbler position 3 in a codon), and noted that GC3 codons tend to have higher CSC scores. This observation is our starting point for comparing the three CALM transcripts. Fig. 3ab shows that human CALM3 has a remarkably higher GC3 content (70%), compared to the 37% in CALM1 and 30% in CALM2. When only considering synonymous codons the values are 33%, 25%, 68% for CALM1,2,3, respectively. As discussed above, CALM1 and CALM2 appear to be closer related to each other than to CALM3, based on similar promoter, enhancer and UTR regions and, as we show here, they are also both predicted to have relatively unstable mRNAs based on their low GC3 content. Next, we applied published CSC scores from eight independent datasets for four human cell lines, from two publications using three different experimental approaches [116,117] (Fig. S4a,b). As shown in Fig. 5a, when averaging the CSC scores from four experiments done in 293T HEK cells, the GC3/AT3 discrimination does indeed appear to be very highly linked to CSC scores, so that there is only a single AT3 codon among the possible CSC codons. Indeed, when calculating the CSC scores for CALM1,2,3 for all cell types, the same pattern as in Fig. 3ab for GC3 content is seen, so that CALM3 has a much higher average CSC score, with CALM2 having the lowest.

In Fig. 5c, we have applied the 293T HEK cell CSC scores along the three transcripts, using the color gradients from Fig. 5a. It seems likely that what is important is not only the average CSC score of a transcript, but also the presence of one or more rows of low scoring codons as seen in CALM1 and CALM2. Hypothetically, a longer row of CSC low scoring codons could slow down translation and trigger the decay machinery, compared to if the low scoring codons are more dispersed. Not only does CALM3 have a higher average CSC score, but the slow codons are also more dispersed and not clustered as compared to CALM1 and CALM2.

Since all three CaMs are identical in their amino acid sequence, the codon usage is limited. For instance, CaM contains nine methionines, which are coded for by only one codon, thereby fixing the CSC scores at those positions. On the other hand, some amino acids are coded for by six different codons, leaving a wide window of CSC scores to be selected for by evolution. Considering the commonly preferred codon usage in human, the coding region of the 148 amino acids CaM protein would be expected to have a GC3 content of 60% (Table 3). This again points to the most salient observation being the low GC3 content of CALM1 and CALM2.

It can be calculated that any CALM mRNA selected for the highest possible CSC score at all sites would have a CSC score of 0.077; on the other hand, any CALM mRNA with a minimal stability would have a CSC score of –0.085. The average CSC score for a theoretical CALM coding region is 0.007. Considering this, it appears that human CALM3 is selected for stability, but more evident, that CALM1 and especially CALM2 are selected for instability.

In Fig. 5d, we show graphically, by cumulative CSC scores, how the codon stability is distributed along the mRNAs of the three CALMs, and for a theoretical CALM mRNA, which has CSC scores calculated by the average score of the codons at each AA position, when considering the common human codon usage (see Table 3). CALM3 has a higher stability than the average, while CALM1,2 are below, as illustrated by arrows. The CALM1 mRNA is only slightly unstable in the first half of the protein, after which it is highly unstable. The CALM2 mRNA is highly unstable throughout the coding region, except a neutral region at the 5′end. Conversely, CALM3 is stable throughout the coding region, except for the very last part of the 3′end.

Fig. 6 shows the GC3 content of the coding sequence of the three CALM genes for eight mammals. They all share the pattern of CALM3 having the highest GC3 content and CALM2 the lowest, except for rat, which has a slightly higher score for CALM2 than CALM1. Mice/rats have the least CALM differences in GC3 content of the seven species, given their high CALM2,2 scores. Cow has the highest scoring CALM3 gene, and also a relatively high CALM1 GC3 content, which is even higher than the GC3 content of mice/rat CALM3, while human/gorilla has the lowest scoring CALM2 gene.

Interestingly, it has been reported that the GC3-rich and GC3-poor gene coding regions predict distinct sub-cellular spatial distributions.
Fig. 5. Comparative stability prediction of the three human CALM transcripts. a. Codons are separated from high to low CSC score, by using the average of four independent measurements of the 293T HEK cells transcriptome \([116,117]\). Color gradient from dark green to dark red is used to discriminate the CSC scores, and GC3/AT3 wobbler position is indicated by green or red for each codon. b. Data from four different human cell lines and three experimental approaches (Endog, ORF and SLAM). In the top section CSC score data from four cell lines are shown. The middle section shows how the data from transcriptomes correlate between different cell lines, and experimental setups. Data from [116] in dark gray, and [117] in light gray. The bottom section shows statistics when applying the CSC scores on the three CALM transcripts. c. The color gradients and data from panel a is used to show how predicted high and low stability codons are distributed along the transcripts. d. Accumulative CSC scores of CALM1,2,3 transcripts. The X-axis shows the codon number of the CALM coding region and the Y-axis the codon stabilization coefficients for each codon (CSC scores) are accumulatively added. The gray CALM is theoretical to how the CSC would be distributed based on the average CSC score for each codon position, based on average human codon usage.
so that GC3 rich gene products are located more in the cell periphery/membrane/projections, while GC3 poor products are closer to the nucleus. As covered in Section 3.2.7, the CALM2 mRNA was reported to be located close to the nucleus in human cells, and CALM3 mRNA in the periphery [112] thus reflecting what would be predicted based on the GC3 content here, provided the mRNA locations correspond to the local translation locations. However, as also mentioned in 3.2.7, in rats CALM1,2 mRNAs are reported to be localised more in the periphery/membrane/projections while CALM3 mRNA was reported close to the nucleus in human cells.
neurites (peripheral) compared to CALM3 mRNA. This apparent difference compared to humans, may reflect the relatively small differences in GC3 content between the 3 rat genes (Fig. 6), giving less locational predictive power.

If these theoretical estimations reflect real properties of the three CALM mRNAs, what could be the functional significance for an organism in having two low and one high stable CALM? Variation of mRNA stability could give the cell an additional mechanism in fine-tuning gene expression, besides the role of promoters, enhancers, and other UTR features. It could be that having low stability CALM coding regions would allow for rapid regulation of CaM production at the transcriptional level, since the overall mRNA level would be mainly regulated by newly transcribed mRNA. Thus, CALM3 having a more stable transcript, could be necessary when transcription is halted, as occurs during spermatogenesis, when a more stable CALM3 may be necessary to prevent the sperm cell from running low of CALM mRNA. Another factor could be the need for stable CALM mRNA in certain cell types, even though transcription is not halted. For instance, translational activity in brain dendrites may need a more stable mRNA, given the relatively long physical distance from the nucleus, making the mRNA more prone to degradation. Epithelial cells might also need more stable mRNAs as a guard against UV light exposure or other mutagens as it is known that chemical damage to mRNA is likely to result in accumulation of aberrant protein products [122].

3.2.9. Transcriptome data

Given the apparent dominant negative effects from single allele CALM mutations causing heart arrhythmia, it would be desirable to know the relative protein levels produced from each CALM gene. This is, however, not a simple task, as the amino acid sequence of CaM produced by 3 independent genes is identical. The relative CALM1,2,3 expression in adult human heart (left ventricle) was reported, by qPCR analysis, to be approximately 14%, 25% and 61%, respectively [123]. If these numbers reflect the actual protein levels, and every allele is expressed at an equal level, it would mean that 7% mutated CALM1 (one of six alleles) is sufficient to cause heart arrhythmia (see also part 4.1). In this section we will review human public RNA-seq data from high-throughput studies where the expression levels of large numbers of genes were measured simultaneously. We will first cover data from bulk tissue gene expression (GTexportal.org) and then data from single cells (scRNA, Proteinatlas.org).

Average CALM1,2,3 expression from 53 tissues is 302, 661 and 329 TPM (transcripts per million), respectively (Fig. 7a). All three genes show the highest expression across thirteen brain tissues with averages of 686, 1012 and 692 TPM (29%, 42%, 29%). The overall expression in testis is, on average, 490 TPM, but here the relative contribution from CALM3 is higher than normal with 54% (506 TPM), and only 8% (124 TPM) from CALM1. Average data from two heart tissues (atrium and left ventricle) are 113, 220 and 136 TPM (24%, 47%, 29%), and in all three genes heart tissue is among the tissues with the lowest expression having...
an average of 156 TPM.

Bulk RNAseq measures the average expression of genes from multiple cells, but is limiting when one wishes to understand gene expression patterns within the cell, and among subpopulations of cells, like for instance cardiomyocytes in heart tissue and the spermatocytes in testis. Looking across 76 single cell types (Fig. 7b), the overall rank between the three CALM5 is not changed compared to bulk tissue with averages of 596, 943 and 218 (34%, 54%, 12%), respectively, but the contribution from CALM3 is lower compared to tissue samples. In neuronal cells the CALM1 contribution is unusually high with 58% (567 nTPM, N normalized transcripts per million), while CALM3 is relatively low with 9% (86 nTPM). Interestingly, when looking at cardiomyocytes, the mRNA load from CALM1 is also the highest (868 nTPM, 51%), and here the CALM3 contribution is only 11% (178 nTPM).

However, the most intriguing observation from these scRNA data is the CALM3 contribution of germ cells. As mentioned above, CALM3 is upregulated in testis tissue (34%), but when looking at certain single germ cell types, CALM3 is dramatically overexpressed. Across 4 different germ cells, the CALM3 contribution is 58% (1393 nTPM), and in early and especially late spermatids the contribution is even higher with 80% (1864 nTPM) and 87% (2697 nTPM), respectively. Since transcription can be expected to be low, or absent at late stages of spermatogenesis, it can be speculated that CALM3 given its theoretically higher stability, becomes more important for the organism than CALM1,2 which would degrade faster due to a lower mRNA half-life.

### 3.3. CaM gene transcription in pathological conditions

The three human CaM genes have been shown to be expressed at different levels in cancer cells. For example, CALM3 presents higher expression than CALM1 and CALM2 in lymphoblastic cells [124]. Also, the expression of the three CaM genes tested in human teratoma cells follows the sequence CALM3 >> CALM2 = CALM1 [51].

CALM1 and the EGFR gene are overexpressed in esophageal squamous cell carcinoma. Knocking-down CALM1 by CRISPR-Cas9 and inhibiting EGFR activation with Aftatinib in these tumor cells synergistically inhibited cell proliferation, migration and invasion in vitro and in vivo and also induced apoptosis [90]. The observed synergy may be due, at least in part, to the stimulatory effect that CaM exerts on EGFR activation [125]. Likewise, in hepatocellular carcinoma the downregulation of CALM2 using siRNA induced the inhibition of cell migration and proliferation, inhibiting as well the formation of colonies and the invasive capacity of these tumor cells via the induction of apoptosis [91].

Increased expression of the CALM1 gene has been shown to occur as well in gastric adenocarcinomas, as compared to normal peritumoral tissue, and this gene was considered a potential inflammation-related target to treat this type of tumors [126]. Also, increased expression of the CALM2 gene has been detected in a series of cell lines derived from anaplastic large cell lymphomas, but to a highly variable extent [127].

CALM1 is upregulated about 5-fold in ovary tissues of patients with polycystic ovary syndrome [128]. In a transgenic mouse model of ALS, it follows the sequence 3.3. CaM gene transcription in pathological conditions

### 4. Human CaM mutations and gene polymorphisms

#### 4.1. CaM mutations leading to heart arrhythmias and polymorphisms associated with cardiovascular disease

A surprising discovery was made and published in 2012, as for the first time mutations in human CaM were found to be linked to life-threatening arrhythmias [131]. As CaM is an essential and extremely conserved protein taking a central role as a calcium sensor in many basic cell-signaling processes [5], it was earlier assumed that mutations in CALM genes would not be tolerated and propagated. Two mutations in CALM1, one at position N54I (this numbering includes the start Met), inherited in a large Swedish family and characterized by repeated events of syncope and heart stop and another de novo mutation at position N98S in a person from Iraq were identified. Both mutations were found to cause catecholaminergic polymorphic ventricular tachycardia (CPVT), which is a severe arrhythmia known to cause sudden cardiac arrest and death during exercise or emotional stress without previously known heart defect. This was followed by the finding of further CaM mutations affecting the CALM1 and CALM2 genes causing long-QT syndrome (LQTS), which affected the patients in a similar way, including strongly enhanced QT episodes of T wave alternans and heart stop in the first year of life [123]. A number of additional CaM mutations were then reported to cause the mentioned syndromes and also idiopathic ventricular fibrillation (IVF) (reviewed in [35,132–135]). Based on these findings, pathologies caused by CALM mutations were termed calmodulinopathy in the International Calmodulinopathy Registry by Schwartz and collaborators [132]. As reported by the registry in 2019, 74 patients from 51 families with CALM mutations had been identified with a distribution among the three CALM genes of 36/23/15 (rel. percentages of 49%/31%/20%) in CALM1,2,3, respectively [132]. Most of these rare mutations were life-threatening and found to be de novo mutations causing early onset of adrenergically induced arrhythmia. The majority of these mutations caused LQTS, and interestingly the same mutations had different outcomes in different patients.

CaM mutations causing heart arrhythmia were to a major degree localized to the C-terminal half of CaM at residues involved in calcium coordination as part of the calcium binding domains EF-hand III and IV (see Fig. 8). Some of these mutations were found in more than one patient with the most prevalent ones being (N98, D130, D132, and F142). These residues were altered to different amino acids in different patients. It was noted that even considering the relatively small number of patients with arrhythmogenic CaM mutations, most CALM-LQTS mutations were found in EF-hand IV replacing evolutionary highly conserved residues important for Ca\(^{2+}\) binding. On the other hand, CALM-CPT mutations were located in EF-hand III or in the EF-hand I-II linker. Interestingly, some mutations including N98S, D132E and Q136P caused both CVT and LQTS (reviewed in [35,132]).

CaM has many binding partners, interacting with them in a huge number of different ways, working as linker and/or adaptor protein due to its structural flexibility [8]. In the heart, several CaM targets have been shown to be affected by CaM mutations. Most prominent are the rydogeminal receptor 2 (RyR2) governing Ca\(^{2+}\) release from the sarco-plasmatic reticulum (reviewed in [136]), and the L-type Ca\(^{2+}\) channel, responsible for voltage-gated Ca\(^{2+}\) entry into the sarcoplasm of the cardiomyocyte (reviewed in [137]). As calcium signaling is the main small signaling molecule controlling and triggering cardiomyocyte contraction and relaxation, unbalanced calcium signaling is an important reason for heart failure. Since CaM is an essential player in calcium signaling of all eukaryotic cells and has many important targets outside the heart, one can only speculate why CaM mutations primarily affect the heart and no other organ functions. It can be speculated, given the low overall expression of CaM and the high abundance of important CaM binding proteins in the heart, the pool of free CaM might be very low, and therefore the function of mutated CaM becomes more pronounced, since it cannot be substituted by free WT CaM. However, there...
are some reports suggesting that \textit{CALM} gene mutations may be causing other pathologies (see Sections 4.2–4.4).

Interestingly, it was found that two newly discovered \textit{CALM} mutations, causing heart failure, were present only in some tissues (genetic mosaicism) of parents from affected individuals [138]. This was documented in one case with a mutation in \textit{CALM3} \textit{E41K} with 25% of sequencing reads and in another case with \textit{CALM3} \textit{D130G} with 6% mutated alleles, in the peripheral blood of the mother in the first case and the father in the second case. This indicates that germline parental mosaicism may explain unexpected fetal arrhythmia in some children of the same parents and multiple intrauterine fetal demise observed in one case. This finding will be of potential relevance for consulting families with \textit{CaM} mutation causing heart disorders.

Limited information is available about genetic polymorphisms in the three \textit{CALM} genes that are correlated with cardiovascular disease [139]. The polymorphisms \textit{rs3174843} and \textit{rs3179089} in the \textit{CALM1} 3-UTR of the gene (CC and GG genotypes, respectively) have been associated with enhanced risk of ischemic stroke, particularly the \textit{rs3179089} polymorphism in females [140,141], and have also been found to be a potential marker for sudden cardiac death [142].

In addition, a polymorphism of unassigned number, in the \textit{CALM3} promoter region at position –347A→T, directly adjacent to the 5-ends of a putative SP1 transcription factor binding site, was found to affect the level of transcription of this gene, with a lower transcriptional activity of the T-\textit{CALM3} version in a promoter analysis in HEK293 cells and cardiomyocytes compared to the wild-type and to exhibit a clearly higher frequency in patients with familial hypertrophic cardiomyopathy (FHC) compared to controls [143]. According to this, a small decrease in wild-type \textit{CaM} would increase the probability of causing FHC.

How can a mutation in one out of 6 alleles of the \textit{CALM} genes cause severe and life-threatening heart diseases? It is known that mutant \textit{CaM} can pre-associate at low calcium with the two most prominent \textit{CaM} targets in cardiomyocytes, the \textit{Ca}_{1.2} and \textit{RyR2} channels which regulate \textit{Ca}^{2+} influx into the sarcoplasm and therefore may compete with wild-type \textit{CaM} in closing the channels after calcium is increased, as shown for the \textit{Ca}_{1.2} channel which has an IQ-motif \textit{CaM} binding site (reviewed in [133–135]). Therefore, it was proposed that mutations affecting calcium binding (the majority of all known arrhythmia causing mutations) would mostly affect the \textit{L}-type calcium channels leading to LQTS and, on the other hand, mutations with minor effect on calcium binding (N98S and N54I) would rather interfere with the more complex situation in the regulation of the \textit{RyR2} channel, which could lead to CPVT. In the case of the \textit{RyR2}, one explanation of the reported dominant negative effect of \textit{CALM} mutations (e.g. of \textit{CaM} \textit{F90L} causing VF and sudden heart arrest [144]) could be based on the fact that \textit{RyR2} consists of 4 subunits each of which have a \textit{CaM} binding domain. Binding of only one mutated \textit{CaM} to the complex could affect the other subunits and interfere with the know clustering of \textit{RyR2} into 2D oligomers. A similar situation of dominant negative effect of \textit{CALM} mutations has been demonstrated for another complex \textit{CaM} target, the \textit{CaMK-II}, which plays as well an important role in heart physiology, and consists of 12 or 14 subunits, which interact with each other depending on the \textit{Ca}^{2+}/\textit{CaM} binding [38,145]. Interestingly, \textit{CaMK-II} regulates both the \textit{L}-type calcium channel \textit{Ca}_{1.2} and the \textit{RyR2} by phosphorylation, thus facilitating channel opening (reviewed in [125,126]). The effect of individual \textit{CaM} mutations occurring in one out of 6 alleles is of course dependent on the relative expression of the three genes. Measuring mRNA levels of the three genes in cardiac tissue has been done by several groups and is so far inconclusive as largely divergent results were obtained (reviewed in [35]). It cannot be ruled out that these divergent data are the result of the theoretically large differences in mRNA stability of three \textit{CALM}s, resulting in higher lab to lab variations than normal. As covered in Section 3.2.9, \textit{CALM2} appears to be the generally highest expressed \textit{CALM}, also in heart tissue, with 47% contribution. The codon stability data covered in Section 3.2.8 suggest that \textit{CALM1,2} produce unusually unstable mRNA, while \textit{CALM3} is predicted to have high stability. It can be imagined how this may result in higher degradation of \textit{CALM1,2} compared to \textit{CALM3} during sample preparation, and so give a general trend towards lab results overestimating \textit{CALM3} contribution. Unpublished data from our lab point into this direction. When we analyzed relative expression from the three \textit{CALM} genes, we have observed striking differences comparing fresh and old tissue samples. qPCR results from cells and fresh tissue showed \textit{CALM3} being the least expressed, but in old rat brain tissue, \textit{CALM1,2} were virtually absent, whereas \textit{CALM3} seemed in the normal range, resulting in a relative \textit{CALM3} expression of 98% and 100%, respectively from two independent measurements. Also, other sources report that the \textit{CALM1,2} transcripts are especially unstable at room temperature [146]. \textit{CALM1,2} were found to correlate with RNA integrity, so that when the RNA quality is low, the relative expression of \textit{CALM1,2} decreased, indicating that \textit{CALM1,2} are more prone to RNA damage compared to genes producing more stable RNA.

The long 4.1 kb mRNA isoform of \textit{CALM1} does appear to be relatively overexpressed compared to the short form in heart left ventricular tissue (no scRNA data for the two forms in cardiomyocytes are available, see Fig. 4). But another, perhaps, more important point when considering mRNA expression in relation to calmodulinopathy, is the nature of the information given when sampling a bulk heart tissue, since cardiomyocytes are estimated to account for only 30% of the 4 cell types in the human heart [147]. As covered in Section 3.2.9, cardiomyocytes appear to express only 11% of the mRNA pool from \textit{CALM3} and 51% from \textit{CALM1}. How these numbers are translated into actual relative protein amounts is not known. As stated above, most patients with mutations in \textit{CALM1} and fewest in \textit{CALM3} are registered, possibly reflecting the cardiomyocyte expression levels of the three genes,
perhaps because a disease variant with higher expression is more likely to lead to severe symptoms. However, due to the small number of patients with CaM mutations this remains speculative.

Even if a clearer picture of mRNA levels will arrive, it is still an open question whether this would relate to the level of proteins, as mutant CaM may differ from the wild-type one in stability, subcellular location, target specificity, and might be post-translationally modified (e.g. phosphorylation, methylation, acetylation, oxidation) in different ways.

Even though a potential connection between the type of CALM mutations and the caused pathology has emerged, as outlined above, a clear correlation between phenotype and genetic alteration or properties of the resulting proteins has not been established yet. As CaM has many important targets in cardiomyocytes, it is likely that the pathophysiological effect of mutant CaM is the result of improper regulation of several critical players.

### 4.2. CaM mutations potentially leading to neural diseases

Ion channels are essential for neural function, and CaM has an important role in the regulation of many of these ion channels, including $Ca^{2+}$-dependent L-type $Ca^{2+}$ channels [148] (reviewed in [134,149]), voltage-gated Na$^+$ channels [150] (reviewed in [134,151]), voltage-gated K$^+$ channels (reviewed in [134]), Ca$^{2+}$-activated K$^+$ channels (reviewed in [152]), among others (reviewed in [35,134,149, 152]). Additionally, other CaM-dependent systems in neurons and/or glial cells, including the important signaling molecules CaMK-II and calcineurin, could also be affected likewise when CaM mutations occur and contribute to neural pathogenesis.

Patients with CALM1 and CALM2 (CALM3 have 29% average relative expression in brain tissue, but only 8% across 11 glial and neuronal single cell RNA measurements) mutations affecting residues at EF-hands III and IV, showing decreased $Ca^{2+}$ single cell RNA measurements) mutations affecting residues at EF-hands III and IV, showing decreased $Ca^{2+}$ affinity, exhibited in addition to cardiac arrhythmias as described above, neurodevelopmental delay, epilepsy, cognitive, and motor deficiencies. This was first ascribed to be secondary to brain injury after transient anoxic episodes, resulting from arrhythmias leading to syncope and cardiac arrest during early life, that was successfully resuscitated [123]. However, the authors of this report could not exclude primary neuronal affection associated to circulatory insufficiency. The same group hypothesized thereafter that perhaps only in cases where multiple cardiac arrests occurred, post-anoxic neural sequelae remain [153]. Nevertheless, this hypothesis was based considering only few cases with divergent neural clinical features, ranging from mental retardation, autism, developmental delay and seizures, and without history of cardiac arrest (as recorded in the International Calmodulinopathy Registry) [132,135,154]. This suggests, not unexpectedly, that CaM mutations may directly affect brain functions.

### 4.3. CaM mutations potentially leading to endocrine-associated illness

A female patient affected with sinus bradycardia presenting heterozygous mutation of CALM3 (E141K) suffered as well a serious hypo-glycemic condition, that was the ultimate cause of her death, and this was attributed to hyperinsulinemia [138]. The authors of this study suggested that this CaM mutation could be responsible for the dysregulated insulin secretion by affecting L-type $Ca^{2+}$ channels in pancreatic $\beta$-cells [138]. This report opens the possibility that CaM mutations could cause or contribute to some endocrine disorders.

### 4.4. CaM polymorphisms potentially associated with osteoarthropathies

Some single nucleotide polymorphisms (SNP) in CALM1 have been associated with osteoarthropathies. Mototani et al. [155] reported that a SNP in intron 3 (IVS3 -293C→T), as well as another SNP in the core promoter region of CALM1 (rs12885713 -16C>T) strongly correlated with hip osteoarthritis. Particularly, these authors claimed that the latter SNP decreased the rate of transcription of this gene, which lowers the CaM level and reduces the expression of other genes essential for cartilage matrix formation. These include COL2A1, encoding the pro-a1 chain of type II collagen [156], and AGC1, encoding chondroitin sulfate proteoglycan 1, also denoted aggrecan [157]. The study of Mototani et al. [155] was performed with Japanese patients. However, analogue studies with British and Greek Caucasian patients did not find a correlation between the occurrence of polymorphisms in the core promoter region of CALM1 and susceptibility to hip or knee osteoarthritis [158,159]. Hoaglund [160], commented that this apparent inconsistency could be explained by the different underlying pathology in the two patient groups, being mostly primary or idiopathic osteoarthritis in the Caucasian group, while being congenital hip dislocations or developmental acetabular dysplasia in the Japanese group, therefore giving validity to both studies. Nevertheless, two SNPs (i-CALM2-5 rs10153674 A>G; and i-CALM2-6 novel G>A) both located in intron 1 of CALM2 were significantly associated with hip osteoarthritis without acetabular dysplasia in the Japanese patients [161]. The authors of this study pointed out that this was observed only after stratification of the patients according to their acetabular dysplasia status, otherwise no significant correlation was found. In addition, SNP rs12885713 -16C>T occurrence in the core promoter region of CALM1, and SNP rs5871 located in the 3′-UTR, were associated with idiopathic scoliosis in adolescents Chinese patients [162]. Nevertheless, the genetic basis of these pathological conditions are multifactorial and many SNPs associated with idiopathic scoliosis have been found in other genes (reviewed in [163]). Furthermore, the association of SNP rs12885713 (−16C>T) in the core promoter region of CALM1 with osteoarthritis predisposition has been further questioned in a study carried out in Han Chinese patients and in a meta-analysis pooling Asian and Caucasian patients [164,165]. These findings suggest that ethnicity may play a role in the pathogenesis of these osteoarthropathies, although further studies are required to clarify this issue.

### 5. Concluding remarks

Based on a multiplicity of discoveries described earlier in [108], and extended in this article, it can be concluded that the three CaM genes in mammals cannot solely function as redundant copies replacing each other in case of need. The following issues would argue against this possibility:

1. Non-coding regions including promoter/enhancers 5′- and 3′-non-coding mRNA sequences, may lead to different stabilities of the RNA, e.g., due to target sequences of proteins and miRNA. The importance of these sequences is highlighted by the fact that they are almost as highly conserved among the corresponding genes in different vertebrates as the coding sequences.
2. Every individual CaM gene has highly conserved non-coding sequences, when different vertebrate species are compared. The conservation of these sequences among genes in the same species is lower, allowing the synthesis of mRNAs in different cell types with different stabilities and subcellular localizations. This can be responsible, at least in part, of multitude modes of fine-tuning regulation.
3. Another factor not previously focused on, but highlighted in this article, is the potential significance of differential codon usage by the 3 CaM genes, which may lead to different stabilities and translational efficiency. It is captivating to observe how the three CALMs have differentiated in their GC3 usage, considering that they originated from one single CALM gene by duplication. The pattern of two low GC3 CALM1,2 genes, and one high GC3 CALM3 gene in humans is also found in seven other mammals investigated. Especially, CALM2 appears to be highly selected for instability throughout its coding region. CaM is a very important protein, as highlighted by its many interacting partners, and uniquely 100% sequence conservation in all investigated
vertebrates. Higher organisms may have evolved to take advantage of three CALM genes with different mRNA stabilities. A low stability mRNA can be in some cell types where fast regulation of the available mRNA pool is especially advantageous, but can be a trade-off in other cell types. With three CALM genes evolution may have more freely tailored the three mRNAs. Following this hypothesis, it is interesting to observe how the high stability CALM3 is the main mRNA in late spermatids, where transcription is very low, and high stability is therefore needed. The organism ability to use an available stable CALM3 mRNA in spermatids may then have allowed the CALM1,2 mRNAs to evolve in a less stable direction, without evolutionary pressure selecting in opposite direction.

(4) Every individual CaM gene seems to be irreplaceable, as deduced from pathological conditions in human. First, mutations in a single CaM gene can lead to illness, despite co-expression of non-mutated paralogs. Second, the specific expression patterns of CaM genes change in cancer cells.

It would be interesting to know whether different subcellular pools of CaM created by the presence of specific mRNA pools would be prone to different post-translational modifications, such as phosphorylation, methylation, carboxymethylation or acetylation [13,166–169]. It can be imagined that the cellular location where the mRNA is translated, and in addition the speed of the translation machinery could result in different modifications of CALM1,2,3 products. Codon usage predict faster translation of the CALM3 compared to the other CaM genes. This could be relevant to determine the contribution of a specific CaM gene mutation or an isoform for the development of a pathology.

The advanced gene editing methods, using CRISPR technology, would allow to slightly modify CaM genes, for example adding different tags. This can be the basis for gene discrimination in mammals in order to study the differential expression of CaM from the 3 genes. However, this may of course be difficult, as the modification of CaM, even if minor, could lead to changes in function.

Data Availability

Data will be made available on request.

Acknowledgments

This work was funded by grants from the Danish Research Council (DFF-4004-00560) the Wedell Weedsborg Foundation and the Candys Foundation, Denmark, to MWB; grants from the program Consolidación de Grupos de la Junta de Andalucía (research group BIO-210) and from the Plan Estatal 2017–2020 (PID2019-103850GB-I00), Spain, to EV; and grant from the Consejería de Educación, Juventud y Deportes – Comunidad de Madrid (B2017/BMD-36), Spain, involving contributions from the European Funds for Regional Development, the Social European Fund, to AV.

References


