BioMetals 1998, **11**, 277–295

Classification and evolution of EF-hand proteins

H. Kawasaki[1]*, S. Nakayama[2] & R. H. Kretsinger[3]

[1]*Department of Molecular Biology, Tokyo Metropolitan Institute of Medical Science, 3–18–22, Honkomagome, Bunkyo-ku, Tokyo 113, Japan.* [2]*Department of Biochemistry, Nagasaki University School of Medicine, 1–12–4 Sakamoto, Nagasaki 852–8523, Japan.* [3]*Department of Biology, University of Virginia, Charlottesville, VA 22903–2477, U.S.A.*

Received 15 June 1998; accepted for publication 19 June 1998

Forty-five distinct subfamilies of EF-hand proteins have been identified. They contain from two to eight EF-hands that are recognizable by amino acid sequence as being statistically similar to other EF-hand domains. All proteins within one subfamily are congruent to one another, i.e. the dendrogram computed from one of the EF-hand domains is similar, within statistical error, to the dendrogram computed from another(s) domain. Thirteen subfamilies – including Calmodulin, Troponin C, Essential light chain, Regulatory light chain – referred to collectively as CTER, are congruent with one another. They appear to have evolved from a single ur-domain by two cycles of gene duplication and fusion. The subfamilies of CTER subsequently evolved by gene duplications and speciations. The remaining 32 subfamilies do not show such general patterns of congruence; however, some – such as S100, intestinal calcium binding protein (calbindin 9kd), and trichohylin – do not form congruent clusters of subfamilies. Nearly all of the domains 1, 3, 5, and 7 are most similar to other ODD domains. Correspondingly the EVEN numbered domains of all 45 subfamilies most closely resemble EVEN domains of other subfamilies. Many sequence and chemical characteristics do not show systemic trends by subfamily or species of host organisms; such homoplasy is widespread. Eighteen of the subfamilies are heterochimeric; in addition to multiple EF-hands they contain domains of other evolutionary origins.

Keywords: evolution, classification, EF-hand, domain, homology, chimera, congruence, gene duplication, gene fusion, eukaryote, dendrogram, calmodulin, troponin C, light chain of myosin, S100, parvalbumin, calcineurin, recoverin, calpain, sorcin, diacylglycerol, calbindin, aequorin, phospholipase C, BM-40

Introduction

At least 300 distinct and different protein domains have been observed in the tertiary structures of proteins and/or inferred from amino acid sequences. Our world of protein structures is estimated to consist of one to ten thousand domains (Chothia 1992; Orengo *et al.* 1994). Each has its own story. We tell one of those stories. We emphasize that we take an evolutionary perspective. Chemical and functional implications are discussed as related to evolution.

Each domain has its own structural and functional characteristics. We will focus on the unique characteristics and evolution of the EF-hand domain. However, many of the concepts and problems seen here are encountered in the analyses of other families. Hence, we first summarize the conceptual framework within which we, and most people, work.

01. Each domain has its own characteristic architecture; all members of that domain are recognizably similar. This similarity is difficult to

*Present address: H. Kawasaki; Division of Plant Genetic Engineering, Kihara Institute for Biological Research, Yokohama City University, Maioka-cho 641–12, Totsuka-ku, Yokohama 244-0813, Japan

define; usually, with proper alignment the mean displacement of α-carbons of corresponding residues, is <2.0 Å. A good sculptor would know it if he saw it.

02. Domains form thermodynamically stable structures when shorn of precursors, terminii, or other domains. They have a hydrophobic core, excepting those that are integral membrane proteins, and are reasonably compact with an axial ratio less than 3:1. There is a single connection to the preceding and following domains; that is, the polypeptide chain does not weave back and forth between two domains.
03. Most domains are at least 20 residues long and seldom more than 200 residues. Sometimes two members of one domain cannot be recognized as being similar, or homologous, based on comparison of their optimally aligned, amino acid sequences. Homologs that lie in this 'twilight' zone are easily recognized as being, or not being, similar when their structures are seen.
04. All members of one domain family have evolved by descent from a common precursor domain; they are homologous. Sometimes (members of) two different domains resemble one another. Usually the burden of proof of convergent evolution rests with the advocate of analogy.
05. Many polypeptide chains consist of several domains – homologous or not – fused together. The history of such a chimeric protein involves the evolution of the individual domains prior to fusion, the fusion event(s), and the subsequent evolution of the chimera.
06. Exons are regions of DNA that encode protein domains. Introns mark sites of splicing in (or out) exons that encode domains of chimeric proteins.
07. Gene duplication and splicing gave rise to different isoforms within the domain family. With the passage of time homologous domains in different proteins and/or in different organisms diverged in sequence, and in the resultant structure, to form different subfamilies.
08. Two proteins (domains) are orthologs if they are found in two different species (organisms) and if their most recent common precursor protein occurred in the species (organism) that was ancestral to the two species (organisms) under consideration. If two homologous proteins are not orthologs, then they are paralogs. For instance, if two homologs, **a** and **b**, existed in the ancestral organism prior to speciation, then **a** in one descendent species is a paralog of protein **b** in the sister species.
09. Many domains are found in archae, eubacteria, as well as eukaryotes. Some, such as the EF-hand, are found only in eukaryotes.
10. Each precursor, or *ur-domain*, sprang forth, fully developed, in the organism that was ancestral to all species in which that domain now exits. This statement is obviously a straw man meant to emphasize our ignorance of the origin of domains.

These ten commandments are frequently valid and provide a reasonable starting point for a discussion of the evolution of proteins in general and of the EF-hand in particular. There are, however, important qualifications or exceptions to each. Those exceptions lend flavor to each domain family and challenges to those who would formulate more comprehensive generalizations:

01. In truth many nominally different domains do resemble one another. This is especially vexing for small domains such as the EF-hand, which consists of only ten residues in an α-helix, ten in a loop, and ten in a second helix (see Chazin, 1998, this issue). The angle between helices varies among EF-hands and within a single EF-hand depending on its functional state. The loop varies in length and path, depending on whether (among other things) calcium can be bound and if so, is bound. A search of all protein structures yields many helix-loop-helix conformations that resemble the canonical EF-hand. One resorts to other criteria to identify and distinguish EF-hands. Usually, but not always (note parvalbumin), an EF-hand occurs paired with an adjacent EF-hand (except in calpain in which two EF-hands from two monomers pair to form a dimer).
02. Pairs of EF-hands can be stable as seen in vitro or inferred from in vivo experiments. Single EF-hands, with or without calcium, are less stable (Donaldson *et al*. 1995). A single EF-hand has a small, partial hydrophobic core. The exposure of the core, or opening, of a pair of EF-hands upon binding calcium and/or binding to a target protein is fundamental to transduction of information in intracellular signaling, the function of many EF-hand proteins. A single EF-hand or a pair of hands are both compact with the exception that the two EF-hands of BM40 grasp an α-helix fifty residues upstream. We note in passing that many fibrous proteins, or extended tails of otherwise globular proteins, show tandem, repetitive sequences. Such (partially) repetitive proteins do not conform to the domain paradigm.

03. Two sequences, each 200 residues long and aligned with few gaps, that are 18% identical probably have a similar structure and in turn are homologs. For sequences as short as thirty residues, one might demand 25% identity. As discussed in **Methods** the Z score takes into account the probability of chance alignment given that particular search sequence and that data base. We assume that no single sequence (or structure) will be significantly similar to two different domains. If this occurs (often), the fundamental concept that the world of (globular) proteins consists of distinct domains is invalid. Several putative EF-hand sequences fall into this twilight zone and one resorts to other correlates. For example, the sequence of domain 2 of aequorin is beyond the twilight zone of homology; yet, its presumed mate is definitely an EF-hand – guilt by association.
04. Certainly the most parsimonious postulate for large domains is descent from a common precursor domain. However, one might argue that a simple helix-loop-helix motif 'arose' independently several times. We will discuss EF-hands from the perspective of a single origin; however, the thousand different sequences in 170 domains of 45 different subfamilies cannot be represented in a simple, tidy dendrogram. We assume, but certainly cannot prove, that this diversity reflects a complex evolution, not multiple, independent origins and convergence to a helix-loop-helix energy minimum.
05. As noted, most EF-hands occur in pairs. At least eighteen of the EF-hand subfamilies are in heterochimeric proteins, i.e., fused with different sorts of domains. In many of these heterochimeras the EF-hands bind Ca^{2+} ions functioning as second messengers. They transduce that information; and the change in structure of the EF-hand(s) is assumed to be transmitted to the catalytic domains of the heterochimeric enzymes.
06. Introns are scattered throughout and between the genes encoding EF-hands. The EF-hand family is one of the strongest exceptions to Gilbert's (1987) interesting, and still sometimes valid, generalization equating exons with domains.
07. Eighteen gene loci encoding EF-hand proteins of six subfamilies have been identified on thirteen different chromosomes in *Homo*. Some are clustered; for example ten of the thirteen known loci for S100 proteins are at 1q21. Others are scattered throughout the genome (Berchtold, 1993). Obviously both gene duplication with adjacent splicing and translocation have been important in this dispersal. The basic assumption underlying the generation of dendrograms based on comparison of aligned amino acid sequences is the accumulation of random point mutations. Especially in short domains this chromosomal turbulence of duplication and translocation, sometimes splitting domains, blurs the history of individual domains.
08. In order to deduce correct phylogenies from dendrograms of proteins one must use orthologs. Paralogs give false phylogenies. The skeletal troponins C of *Homo* and *Rattus* resemble one another more closely than do the skeletal TNC of *Homo* and the cardiac TNC of *Gorilla*. Many EF-hand subfamilies are ill suited for reconstructing the phylogenies of species because they have several gene loci, hence isoforms. Further, as will be elaborated, the relationships among subfamilies are convoluted. We cannot sort out all of the gene duplications, translocations, and splicings that gave rise to these 45 subfamilies, each with multiple alleles.
09. EF-hands have been found only in eukaryotes with the single exception of the four EF-hand domain protein in *Saccharopolysporin*. If no or few other EF-hand proteins are found in eubacteria or in archae, one might suspect that the encoding gene was transformed or transduced into the ancestor of *Saccharopolysporin* from a eukaryote. Such transformations are assumed to occur infrequently but even the rare event would compromise the deduction of evolutions of species from dendrograms of their constituent proteins. This is one reason why phylogenies based on multiple protein families are much more reliable than those based on a single character.
10. It seems unlikely that the many different domains that (so far) have been found only in eukaryotes sprang from the head of Medusa. Yet, we have little idea how new domains arose in eukaryotes and not a clue as to how the first dividing cells got their proteins.

We summarize our methods and strategies for sorting out some patterns and insights into the evolution of the EF-hand proteins – first, characterization of subfamilies, then generation of dendrograms within subfamilies and among subfamilies. For clarity these processes are described as sequential. In fact they are iterative and involve some subjective judgments. Finally we discuss the inferences that can be made from these classifications and dendrograms, about the evolution and about the functions of EF-hand proteins.

Methods

Subfamilies

Most subfamilies of EF-hand proteins, were originally, and correctly, identified by functional and/or by chemical characteristics of the first known representative(s). With hindsight we now ask that all members of one subfamily be congruent and that they have similar functions. That is, the optimal dendrogram of the various EF-hand domains of the putative subfamily will show all domains 1 clustering together, all domains 2 clustering together, etc. as illustrated for troponin C (Figure 1). The second defining characteristic is that, within statistical error, the relationship among the species hosting those proteins will be similar in the domain 1 dendrogram and in the domain 2 dendrogram, etc. The subdendrograms show congruence with one another. The concept of congruence has been described in terms of a chimeric family; however, it is equally valid for a single domain. The subdendrogram of the N-halves should be congruent with that of the C-halves – a nice control.

If several subfamilies are congruent with one another, e.g. the S100's, the intestinal calcium binding proteins, and the trichoylin/profilaggins (Figure 2), the decision whether to classify them as one or several subfamilies depends on information regarding structure, function, or gene loci. For example, the ICBP's, like the S100's, consist of only two EF-hands. They are put into two distinct subfamilies for four reasons. The ICBP's cluster apart from the S100's; this condition must by definition be satisfied for two distinct subfamilies. The ICBP's are monomers and S100s are dimers. The ICBP's have been suggested to function in transcelluar transport of calcium; several S100's are inferred to interact with target proteins in the cytosol. The genes encoding ICBP are on the X chromosome; those encoding S100's are clustered on chromosome 1 (Schäfer *et al.*, 1995) and one each on 21 and 4 in humans. The HYFL's cluster together and have ~2000 residues in other types of domains in their C-terminii.

The computation of the optimal dendrogram of >1000 domain sequences, or even of ~170 subfamily domains is practically impossible. Even if such a massive dendrogram were computed or even approximated, it would contain so much detail as to be incomprehensible; hence, several simplifications. As discussed, proteins that contain identified EF-hand domains are assigned to subfamilies. Then (a representative sequence of) each of the subfamily domains is compared with all of the other domains to determine which are more similar. The other domain having the most similar sequence is indicated for each domain in table 1.

This simple procedure for determining most closely related domains provides a valuable starting point for determining congruence but it is imperfect for at least three reasons. More slowly evolving (domains of) proteins like calmodulin will appear to be closely related to many proteins because the branch length from the common precursor to CAM is short, reflecting its slow rate of change. The mathematical procedure to generate dendrograms optimizes the connectivity and branch lengths among all EF-hand domains; nearest neighbors are closely related but not necessarily the most similar. Standard alignment procedures do not weight the functional similarities of the domains, the functional importance of the amino acids, nor the nature of insertions (and deletions).

Calmodulin, troponin C, essential light chain, and regulatory light chain of myosin show strong congruence with one another. The dendrogram of these four subfamilies, collectively referred to as *CTER*, is very robust because it contains 142 proteins or 568 domain sequences. CAM is inferred to be encoded in the genomes of all eukaryotes; further it is expressed in (nearly) all cells. CAM interacts with over twenty target proteins within the cytosol (Rhoads & Friedberg, 1997); hence, its sequence has been highly conserved. Since CAM is present in all eukaryotes, its four domain precursor was present in the organism ancestral to all eukaryotes. It evolved very slowly, it is almost a fixed point in time; many EF-hands are closely related to it.

Scores of (regions of) proteins have been suggested to contain EF-hands without strong justification. The issue is usually resolved when the structure becomes available. To test a putative sequence we first compare it with the SWISSPROT data base and ask that its Z score (Altschul *et al.*, 1994) be greater than 100 as compared with known EF-hand in the data base and that the top scoring matches be EF-hand proteins. Since it is sometimes difficult to get a high Z score with only a thirty residue domain imbedded in a longer sequence we may cut out just the putative thirty residues to use as a probe sequence. If its better Z-scores, against known EF-hand domains, fall in the twilight zone, then we exploit the observation that most EF-hands occur in pairs and include a stretch of residues before and/or after the putative EF-hand to include in the probe. We also compare the test sequence against our own data base of known EF-hands to determine which known EF-hands are most similar (table 1). Although in principle the two procedures should give the same result, our data base is easier to use and interpret. Either data base may have recent entries not present in the other.

Dendrograms

The goal of many analyses of the evolution of proteins, or of their encoding DNA's, is to generate a dendrogram. The assumption implicit in such a representation is that all of the terminal nodes, classically referred to as operational taxonomic units (OTU's), represent (domains of) proteins that are related by descent from a common origin. The internal nodes, or branch points, could result from either speciation or gene duplication.

Within a subfamily or among congruent subfamilies, such as *CTER*, the optimal alignment of the sequences among domains is unambiguous. We then used MOLPHY (Adachi & Hasegawa, 1996) or other standard programs to generate dendrograms that show the relationship among the domains. If the evolutionary relationships

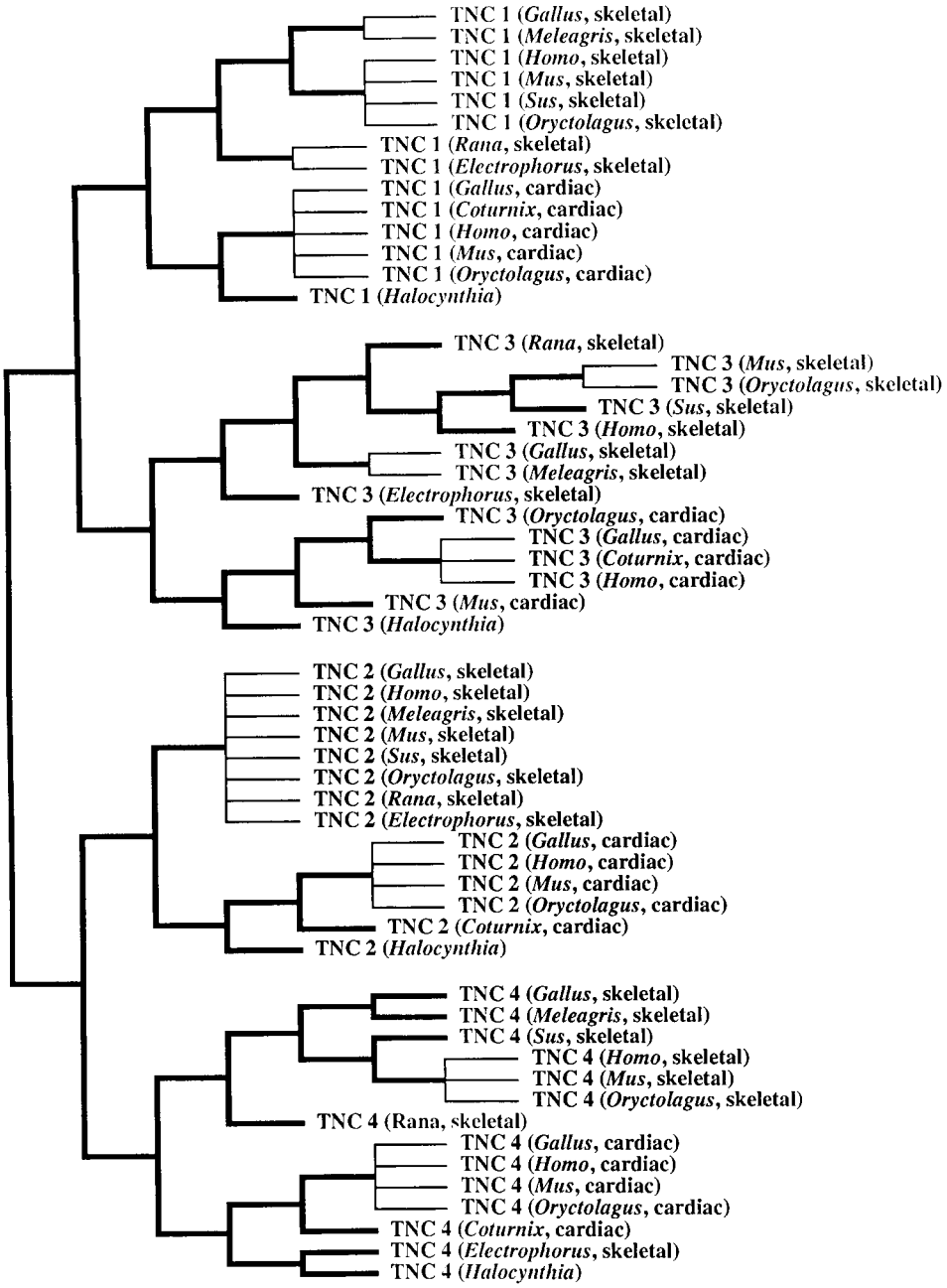

Figure 1. Amino acid sequences are available for 14 troponins C from ten different species. To date TNC has been found only in animals; most animals have several isoforms encoded at separate loci. The dendrogram illustrates the two essential characteristics of congruence. All domains 1 are more closely related to one another than to other domains; correspondingly all domains 2 cluster together as do domains 3 and 4. Second, the distribution of domains within each of the four subdomains is (nearly) identical. An additional interesting characteristic, not inherent to the concept of congruence, is that the cluster of domains 1 is most closely related to domains 3 and the domains 2 are more closely related to domains 4

among species are known, the interpretation of such dendrograms of domains is usually straight forward.

However, we often encounter a situation in which one domain of one subfamily is most closely related to a domain of subfamily **a**; while another domain of the subfamily in question is most closely related to a domain of subfamily **b**. One is then faced with two challenges – first to estimate the statistical reliability of that relationship, then to interpret the probable evolutions of these domains, prior to and following the gene splicing event(s).

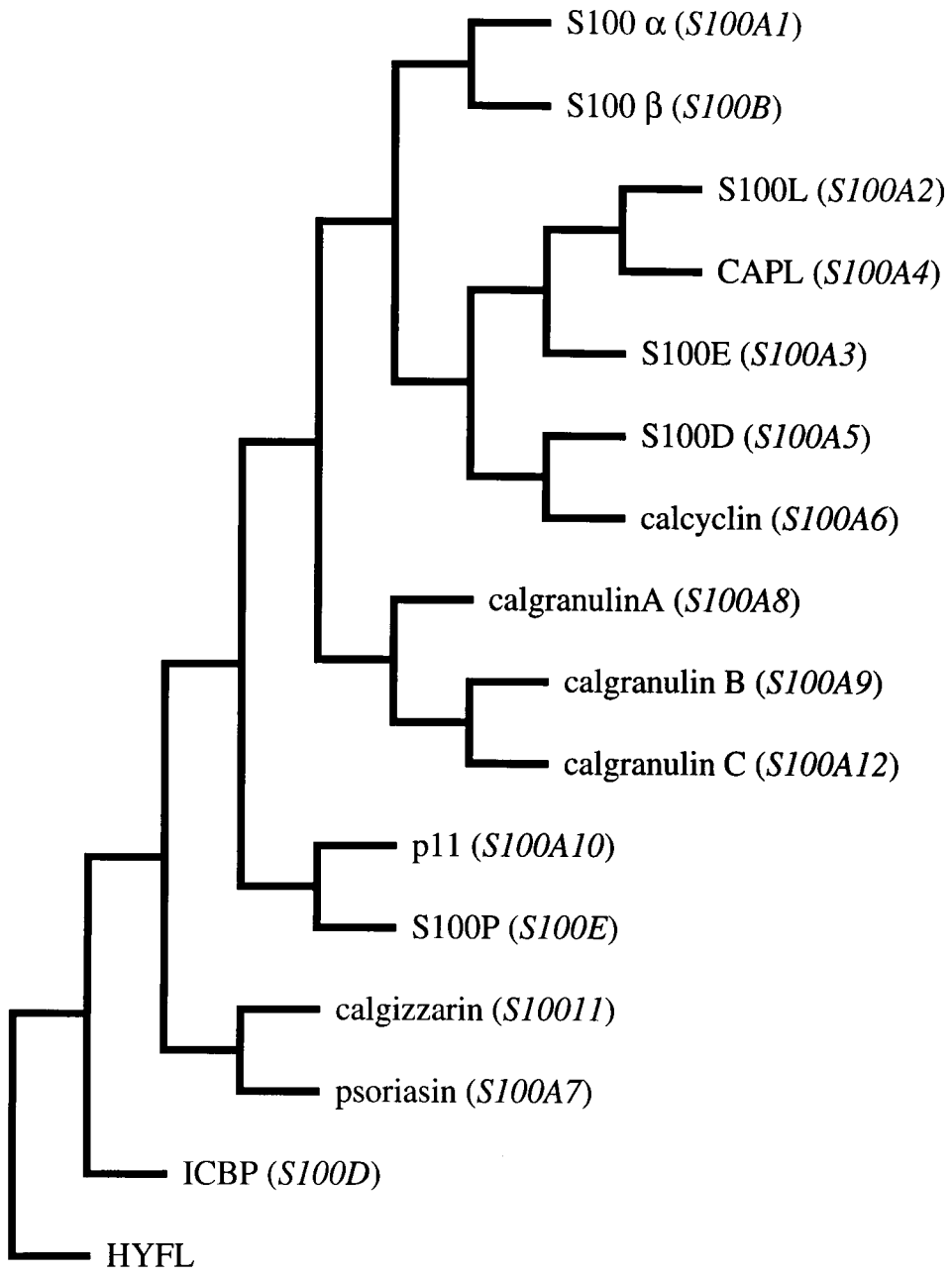

Figure 2. Dendrogram of the congruent thirteen S100, two ICBP, and one HYFL proteins of *Homo*. Fifty-seven sequences are available for vertebrates. Within each of the sixteen isoforms the species cluster by their known phylogenies; orthologous isoforms are correctly identified. For clarity all of the non-human sequences are omitted from the figure; although, all 57 were used to compute the dendrogram.

These classifications and dendrograms are valuable in their own right; they represent the basic anatomy of the subject. Hopefully, they provide the foundations for interpretations of function. Here a note of caution is appropriate. One can often infer binding sites, etc. in proteins by looking at conserved residues in either mutation studies or in alignments of similar sequences. These valuable deductions do not rely explicitly on the concept of evolution. We will distinguish those conclusions based on evolution and those based solely on alignments with indication of conserved or consensus residues.

Table 1. (See p. 284) The EF-hand subfamilies are described in several groups – ***CTER*** those that are congruent with calmodulin, **Pairs** those closely related among or between themselves but not closely related to other subfamilies, **Self** those whose domains are most closely related to other domains within the same subfamily, and **Miscellaneous** those whose domains do not show a strong and consistent pattern of similarity with other EF-hand subfamilies. For 16 subfamilies only one sequence is available; this is indicated by inclusion of genus name in parentheses or as part of the name of that subfamily. **Kingdom** refers to whether the protein is found in **A**nimals, **P**lants, **F**ungi, and/or **P**rotists. The **Function/Struct** column indicate whether a function is known '+' or not '?' and whether a crystal structure (or model for CVP) is available 'X' or not '?'. The symbol '****' in the **Het/ch**. column or in the last domain column indicates a hetero-chimeric protein with the non-EF-hand domain(s) to either the N- or C-side of the EF-hands. For AEQ and PPTS the '**-**' indicates that domain is not recognizable as an EF-hand by analysis of it sequence; however, its proximity to an otherwise unpaired EF-hand suggests that it may be an EF-hand and inferred not to bind calcium. For each EF-hand domain, 1 – 4 (or if up to eight domains, wrap around 5 – 8) is indicated whether calcium binding is observed (or inferred from sequence) '+' or not '-'. For some subfamily domains there are instances of both binding and not binding calcium '+/-'. There are three examples – 'a, b, & d' – of noncanonical EF-hand loops that bind calcium, figure 4. Some loops inferred not to bind calcium may provide additional examples of noncanonical calcium binding loops. The second entry in each domain column indicates the domain most closely related, as described in **Methods**, to that subfamily domain.

Discussion

Characteristics of Subfamilies

In order to make the EF-hand proteins accessible to those outside the field we provide a very brief description of each subfamily.

Calmodulin (CAM) is inferred to be encoded in all eukaryotes and expressed in most cells. In its four calcium form it activates at least twenty target enzymes or structural proteins (Rhoads & Friedberg, 1997; Crivici & Ikura, 1995). Pairs of domains 1&2 are connected to lobe 3–4 by an eight residues linker. Helix F2, the linker, and helix E3 form a seven turn α-helix. The linker serves as a flexible tether. In the apo-form both lobes are in the closed conformation and the tether is extended; CAM is then dumbbell shaped. In the Ca_4-form the tether is bent, both lobes are in an open conformation, and the two lobes enfold a helix of the target protein (Finn & Forsén 1995).

Troponin C (TNC) together with TNI and TNT forms the heterotrimer, troponin, which imparts calcium sensitivity to skeletal and cardiac muscle. TNC has eleven residues in the linker; hence, the F2 α-helix, linker, E3 helix is eight turns long. TNC is dumbbell shaped in the apo form. The two lobes of TNC in the Ca-form enfold the target peptide, TNI 1–47, (Vassylyev *et al.* 1998) in a conformation similar to the CAM*target complex. In the dendrograms based on both protein and DNA the vertebrate skeletal and cardiac forms cluster separately indicating that the common precursor of all vertebrates already has these two isoforms. *Halocynthia*, a protochordate, has the most divergent TNC; its TNC clusters near the base of the cardiac branch (figure 1).

Essential Light Chain (ELC) has two alternate mRNA splice forms. L1 has ~55 residues before the first EF-hand; L2 has about ten. The ancestor to vertebrates had four isoforms of ELC. The skeletal form is most divergent then the smooth muscle form. The cardiac atrial and cardiac ventricular forms are most similar. Molluscan ELC is inferred to bind calcium in its first EF-hand; vertebrate ELC's do not bind calcium. ELC and RLC both enfold the α-helical portion of the myosin heavy chain (Rayment *et al.* 1993)

Regulatory Light Chain (RLC) of vertebrate myosins do not bind calcium but are an alternate source of calcium activation (in smooth muscle) or modulation (striated and cardiac) via a CAM activated kinase. Comparison of the crystal structure of molluscan RLC with a Ca^{2+} ion bound in domain 1 (Houdusse & Cohen, 1996) with the complex of chicken ELC and RLC complexed with the S1 fragment of myosin (Rayment *et al.* 1993) has permitted them to present a model of ELC and RLC complexed with myosin in the active and inactive states.

Troponin, *nonvertebrate* (TPNV) is found in various nonvertebrates; for example lobster has three isoforms (Garone *et al.* 1991). It is a close homolog of TNC; however, its mode of function remains unknown and it clusters in a subfamily distinct from TNC. Its first and third domains closely resemble RLC3; its second and fourth domains resemble TNC2. It appears to have arisen by a relatively recent duplication of a two domain precursor. The interdomain linkers of CAM, TNC, ELC, and RLC have been shown to be bent when those proteins are bound to a target. In contrast, 2,3 linker of TPNV is strongly predicted to be helical and hence not to bend.

EF-hand Proteins

Acronym	Name	King-dom	Function Struct	Het ch.	Domain 1(5) Ca-binding Homolog	Domain 2(6) Ca-binding Homolog	Domain 3(7) Ca-binding Homolog	Domain 4(8) Ca-binding Homolog
CAM - TNC - ELC - RLC								
CAM	calmodulin	A P F P	+ X		+ TNC1	+ TNC2	+ ELC3	+/- RLC4
TNC	troponin C	A	+ X		+/- CAM1	+ TPNV4	+/- CAM3	+ CDPK4
ELC	essential light chain	A F	+ X		a/- CAM1	+/- RLC2	+/- CAM3	+/- RLC4
RLC	regulatory light chain	A F P	+ X		+ CAM1	- ELC2	- CAM3	- CAM4
TPNV	troponin, nonvertebrate	A	+ ?		- RLC3	+ TNC2	+/- RLC3	+ TNC2
CAL	cal1 (*Caenorhabditis*)	A	? ?		+ TNC1	+ CAM2	+ CAM3	+ CAM4
SQUD	squidulin (*Loligo*)	A	? ?		+ CAM1	+ CAM2	+ CAM3	+ CAM4
CDC	CDC31 & caltractin	F P	? ?		+ CAM1	+/- CAM2	+/- CAM3	+ CAM4
CVP	Ca vector protein (*Branch.*)	A	? ?		- RLC1	- S100.2	+ ELC2	+ CAM4
EFH5	EFH5	P	? ?		- CAM1	+/- ELC2	+/- RLC3	- CDC4
CLAT	CAM like leaf (*Arabidopsis*)	P	? ?		+ CAM1	+ CDC4	+ CAM3	+ CAM4 ****
CAST	CAST (*Solanum*)	P	? ?	****	+ RLC1	- ELC2	+ RLC1	+ CDPK4
PARV	parvalbumin, oncomodulin	A	? X		deleted	- ELC2	+ CLBN3	+ TPNV4
Pairs								
S100	S100	A	? X		c/- ICBP1	+ ICBP2		
ICBP	intestinal Ca binding protein	A	? X		c/- S100.1	+ S100.2		
HYFL	trichohylin, profilaggrin	A	? ?		c/? S100.1	+ S100.2****		
CLNB	Calcineurin B	A F	+ X		+ P22.1	+ P22.2	+ P22.3	+ P22.4
P22	p22	A	? ?		+ CLNB1	- CLNB2	+ CLNB3	+ CLNB4
VIS	visinin & recoverin	A	+ X		- CLNB1	+ CLNB2	+ CLNB3	- CLNB4
CALP	calpain	A	+ X	****	+ SORC1	+ SORC2	- SORC3	- SORC4
					- SORC5			
SORC	sorcin	A	? ?	****	+ CALP1	+ CALP2	- CALP3	- CALP4
					- CALP5			
SPEC	*Strongylocentrotus* CaBP	A	? ?		+/- LPS1	+ LPS2	+ LPS3	+/- LPS4
LPS	*Lytechinus pictus* SPEC	A	? ?		+ LPS5	+ LPS6	+ LPS7	+ LPS8
					+ LPS1	+ LPS2	+ LPS3	- LPS4
DGK	diacylglycerol kinase	A	+ ?	****	+ GPD1	+ GPD2 ****		
GPD	glycerol-P dehydrogenase	A	+ ?	****	- DGK1	+ DGK2		
ACTN	α-actinin	A	+ ?	****	+/- FDRN1	+/- FDRN2		
FDRN	α-spectrin, α-fodrin	A	+ ?	****	- ACTN1	- ACTN2		
Self								
CLBN	calbindin 28 kDa, calret.	A	? ?		+ CLBN3	+/- CLBN4	+ CLBN1	+ CLBN2
					+ CLBN3	- DGK1		
TCBP	*Tetrahymena* CaBP	P	? ?		- TCBP3	+ TCBP4	+ TCB1	+ TCB2
TPP	p24 thyroid protein	A	? ?		+ 1FB3	+ VIS2	+ TPP4	? TPP3
Miscellaneous								
RTC	reticulocalbin (*Mus*)	A	? ?	****	+ AEQ3	+ VIS4	+ SARC3	+ CLBN4
					+ CLBN2	+ PLC4		
PFS	surface prot (*Plasmodium*)	P	? ?	****	+ PLC2	+ DGK2	+ PLC1	+ CLBN2
					+ SARC4****			
1FB	1F8 & TB17	P	? ?		+ VIS1	+ SARC2	+ TPP1	+ VIS4
SARC	sarcoplasm Ca binding prot	A	? X		+ RLC1	+/- 1F8.2	+ TPP3	+/- CALP4
CMSE	CaBP (*Saccharopolyspora*)	eubact.	? ?		+ SARC1	+ SARC2	+ CLNB4	+ SARC4
AEQ	aeuqorin & luciferin bind pro	A	+ ?		+ LPS7	**-** FDRN2	+ CDPK1	+ EFH5.4
PPTS	protein phos (*Drosophila*)	A	+ ?	****	- EFH5.1	**-** S100.2	+ AEQ3	+ SARC4
LAV	LAV1 (*Physarum*)	F	? ?	****	+ VIS1	+ TPNV2	+ TCBP3	+ EFH5.4
CDPK	Ca dep. protein kinase	P	+ ?	****	+ AEQ3	+ RLC2	+ CLBN3	+ TNC4
PFPK	prot kinase (*Plasmodium*)	P	+ ?	****	+ VIS1	+ RLC2	+ EFH5.3	+ SARC2
PMAT	memb assoc (*Arabidopsis*)	P	? ?	****	+ LPS1	+ DGK2	+ CAM3	+ CALP4
PLC	phospholipase C	A F	+ +	****	- TNC4	- PARV3	- ICBP1	- P22.2 ****
FIMB	fimbrin	A F	+ ?		+/- CALP1	+/- CAM2 ****		
BM40	osteonectin, SPARC	A	? X	****	d/+ P22.1	+ VIS2		
CRPG	CAM rel gene prod (*Homo*)	A	? ?		+ CLBN6	+ CDPK4		

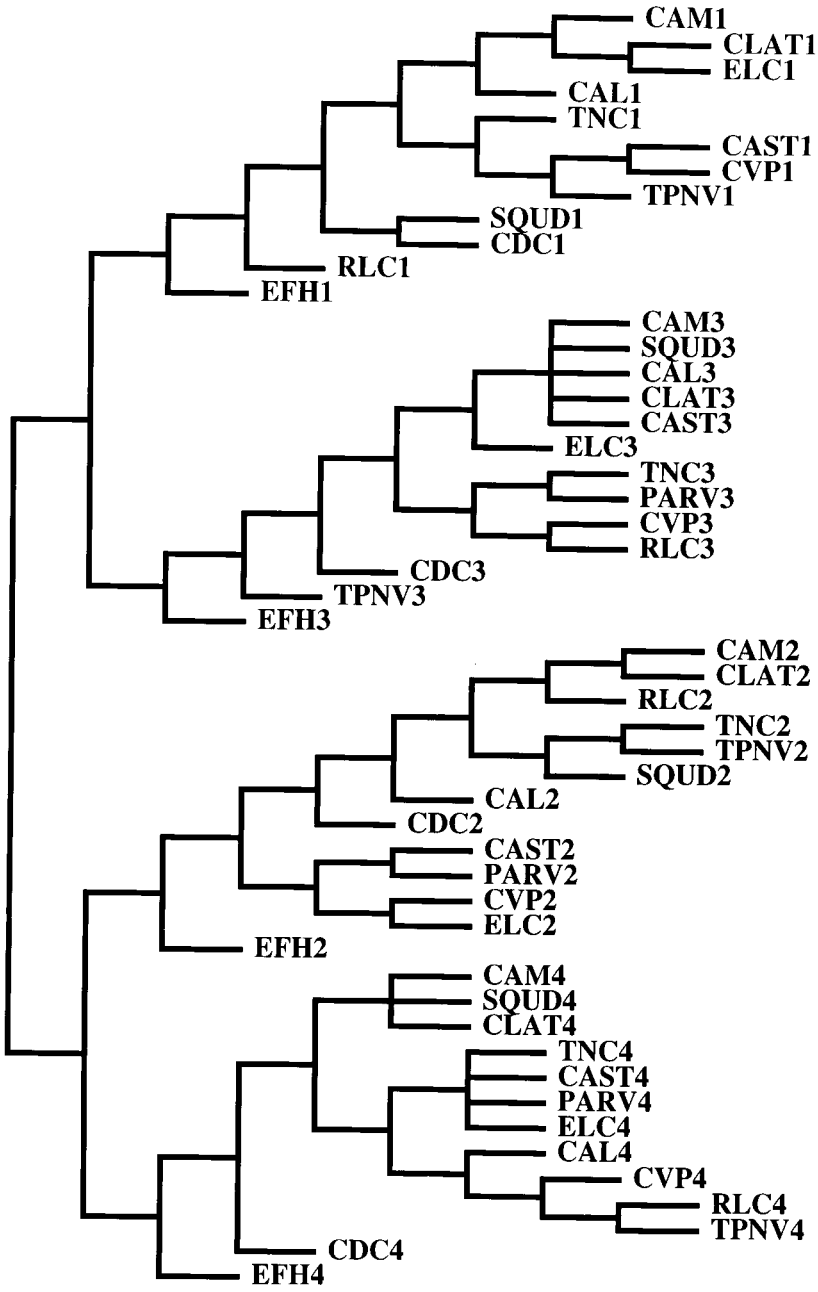

Figure 3. Dendrogram relating the 51 domains of the 13 *CTER* subfamilies – CAM, TNC, ELC, RLC, TPNV, CAL, SQUD, CDC, CVP, PARV, CLAT, CAST, and EFH5. All of the ~25 domains 1 of CAM are represented by a single terminal node; however, the information of amino domain sequences was used in constructing the dendrogram. All domains 1 cluster together, as do domains 3, 2, and 4. Within each domain cluster the relationships among subfamilies is similar but not identical.

Cal1 (CAL): The *cal-1* gene was isolated from the nematode, *Caenorhabditis elegans*, by hybridization with a human CAM cDNA probe (Salvato *et al.*, 1986). CAL has four EF-hands in its 161 residues. It is distinct from CAM but strongly congruent with *CTER*.

Squidulin (SQUD) is identical to vertebrate CAM in 101 of 149 positions; however its Lys115 is not methylated (Head, 1989); Lys115 of CAM is trimethylated. Further, it has a Pro at the second residue of the 2–3 linker, which is helical in CAM (when not bound to a target). Bearer *et al.* (1996) found that SQUD functions as a light chain of myosin in the axoplasm of squid.

Caltractin (CDC) was initially cloned from the unicellular green alga, *Chlamydomonas*. It is a struc-

tural component of the basal body complex, the major microtubule organizing center in *Chlamydomonas* and the functional homolog of the centrosome in the animal cell. The *CDC31* gene product, a yeast homolog of caltractin, interacts with the *KAR1* product. *KAR1* encodes an essential component of the yeast spindle pole body that is required for karyogamy and duplication of the spindle pole body (Biggins & Rose, 1994).

Calcium vector protein (CVP) of *Amphioxus* is inferred to bind to a target protein. In the model (Cox *et al.* 1990) of CVP, based on the crystal structures of CAM and of TNC, a target α-helix is bound between the two lobes of CVP and the disulfide bond in domain is accommodated with no steric interference.

EFH5 (EFH5) of trypanosomes is encoded by the *TbEFH5* gene, which is transcribed as a polycistronic messenger. It follows the CAM gene *C* by 111 bp and precedes ubiquitin-EP52/1 by 116 bp (Ajioka & Swindle, 1993). *TbEFH5* is inferred to be under the control of a distant upstream promoter in *T. brucei, T. cruzi, and L. tarentolae* (Wong *et al.*, 1993). EFH5 is congruent with *CTER*; however it is the most distant subfamily of *CTER*.

Calmodulin-like leaf protein from Arabidopsis thaliana (CLAT) is congruent with *CTER* but differs from CAM in that it has a 47 residue domain of unknown homology at its C-terminus. Its linker between lobe 1,2 and lobe 3,4 is TYSEK in contrast to the MKDTDSEE of CAM (Ling & Zielinski, 1993).

pCAST (CAST) clone from *Solanum tuberosum* (potato) encodes four EF-hands preceded by 38 residues (Gellatly & Lefebvre 1993). Its 2,3 linker is 23 residues long. It is not in the same subfamily as PMAT, which is not congruent with CTER; however, it is interesting to note that both contain $(Ser)_5$ in their N-terminal domains.

Parvalbumin (PARV) consists of three EF-hands, the first of which does not bind calcium and lies over the (potential) hydrophobic patch of the lobe formed by the C-terminal two domains. The domains are numbered 2, 3, & 4 and are, with that numbering, congruent with *CTER*. The first domain is inferred to have been deleted. EF-hands 3 & 4 of PARV have high affinity for calcium and for magnesium. In the resting muscle cell PARV is in the two magnesium form. Following a pulse of messenger calcium the bound Mg^{2+} ions dissociate slowly and the excess Ca^{2+} ions are bound by parvalbumin thereby permitting relaxation of skeletal muscle. Whether PARV also performs this relaxing function in other tissues is not known.

S100 (S100) family consists of at least thirteen distinct isoforms. S100's A1 – A10 in humans are all encoded on chromosome 1; whether they are subject to coordinate expression has not been determinded. Most have been demonstrated to be dimers. S100a $(S100A1)_2$ regulates giant protein kinase (Heierhosrt *et al.*, 1996) and S100A9 forms a heterotetramer with annexin II. The NMR solution structure of dimeric S100B free and complexed with a peptide from protein p53 (Rustandi *et al.* 1998) indicate binding the cleft between domains 1 and 2. Certain neurons are characterized by their contents of S100, as well as contents of parvalbumins. Several S100's function in culture as neurite extension factors (Schäfer & Heizmann 1996).

Intestinal calcium binding protein (ICBP) is strongly congruent with S100. It is placed in a different subfamily because it is a monomer; its encoding gene is on the X chromosome; and it is inferred to function in transcellular transport of calcium (Wooding *et al.* 1996).

Trichohyalin and *profilaggrin* (HYFL) have two EF-hands at their N-termini, both of which are congruent with S100 and with ICBP. Trichohylin contains seven tandemly repeated domains homologous to involucrin at its C-terminus (Lee *et al.* 1993). Profilaggrin has ten to twelve filaggrin domains at its C-terminus (Markova *et al.* 1993). Both associate with keratin intermediate filaments.

Calcineurin B (CLNB) is the regulatory subunit and calcineurin A is the catalytic subunit of protein phosphate. Calmodulin also binds to calcineurin A; both CNLB and CAM in their calci-forms are required for catalytic activity. The crystal structure of Ca_4-CNLB, complexed with an α-helical fragment (350–370) of calcineurin A and with the drug complex, FKBP12-FK506, (Kissinger *et al.* 1995; Griffith *et al.* 1995) shows both lobe 1,2 and lobe 3,4 embracing the fragment (350–370) in their grooves homologous to those of CAM and TNC. In CLNB the two lobes bind to the same side of the target and are displaced about 22 Å along it. In contrast the two lobes of CAM enfold the target peptide from myosin light chain kinase on opposite sides and are related by an approximate two fold rotation axis.

P22 (p22) is required for constitutive exocytotic membrane traffic in rat liver. p22 is N-myristoylated as is the congruent recoverin. The N-myristoylate group and the calcium induced change in conformation are essential to the function of P22 (Barroso *et al.* 1996). Lin and Barber (1996) reported that P22 inhibits the GTPase stimulated, Na^+/H^+ exchanger.

Recoverin or *visinin* (VIS) regulates guanylate cyclase activity in retinal cone cells of vertebrates. In the dark, cationic channels are kept open by bound cGMP. Ca^{2+} ions enter through these channels; this

influx is matched by efflux through a Na^+-K^+/Ca^{2+} exchanger. Light activates an enzyme cascade that stimulates cGMP hydrolysis, leading to channel closure, which blocks calcium influx but not its efflux. Free [Ca^{2+}] in the cell drops within 0.5 s of light stimulation. Recoverin prolongs the photo-response by blocking the phosphorylation of photexcited rhodposin by rhodopsin kinase. Isoforms, visinin and neurocalcin, are found in nerve cells; frequenin was characterized in *Drosophila*. Recoverin is N-myristylated. In the crystal structure (Flaherty *et al.* 1993) of nonmyristoylated recoverin a single Sm^{3+} ion is bound in place of Ca^{2+} in domain 2. The hydrophobic patches of lobe 1,2 and of lobe 3,4 are exposed and orientated so that their two clefts could not enfold a target helix as seen in CAM, TNC, ELC, RLC, CLNB and BM40. NMR studies (Ames *et al.* 1995) showed that the myristoyl side chain is buried between the two lobes in the apo-form but is exposed following calcium binding to EF-hands 2 and 3.

Calpain (CALP) (EC 3.4.22.17) is an intracellular cysteine protease whose two isoforms require either either μM or mM free [Ca^{2+}] for activity. The large subunit of this heterodimer contains a domain homologous to sulfhydryl protease and five EF-hand domains at the C-terminus. The small subunit contains a Gly-rich domain at its N-terminus and five EF-hands at its C-terminus. In both subunits domains 1 & 2 and 3 & 4 form two lobes. The fifth domains of large and small subunits pair, as usually found in adjacent EF-hands, to join the two in the heterodimer (Hohenester *et al.* 1996).

Sorcin (SORC) or grancyclin is a homodimer, congruent with CALP, and inferred to dimerize via its fifth EF-hand. Its concentration is greatly increased in multidrug resistant cells and has been shown to bind to the ryanodine receptor (Lokuta *et al.*, 1997) and to the N-terminal domain of annexin VII (synexin) (Brownawell *et al.* 1997).

Strongylocentrotus purpuratus ectodermal calcium binding protein (SPEC) mRNA's begin to accumulate twenty hours after fertilization in aboral cells (Hardin *et al.* 1988).

Lytechinus pictus SPEC-resembling protein (LPS) evolved by a recent gene duplication and fusion from a four domain precursor, apparently common to the precursor of SPEC. The distribution of introns in the two halves of LPS is similar to one another and to that in SPEC. Further, interdomains 2,3 and 6,7 of LPS are similar to interdomain 2,3 of SPEC (Xiang *et al.* 1991).

Diacylglycerol kinase (DGK) (EC 2.7.1.107) has two EF-hands that span residues 114–202. The domain at the N-terminus is not similar to other known protein families. The domain (203–735) at the C-terminus contains an ATP binding site and cysteine-rich zinc-finger sequences similar to those found in protein kinase C. Isoforms α, β and γ show varying levels of calcium sensitivity (Yamada *et al.*, 1997).

FAD-dependent, mitochondrial glycerol-3-phosphate dehydrogenase (GPD) (EC 1.1.99.5) is located in the outer surface of the inner mitochondrial membrane. The mammalian enzyme is homologous to the yeast and eubacterial FAD-glycerol-3-phosphate dehydrogenases and has two EF-hands at its C-terminus, unlike its yeast and bacterial homologs (Brown *et al.*, 1994).

α-Actinin (ACTN) is an actin filament cross-linking protein. It is a homodimer whose subunits assemble in an antiparallel fashion to form a rod like structure (Taylor & Taylor 1993). The difference between nonmuscle and muscle forms is that the cross linking activity of the former is completely inhibited by μM calcium, while the activity of the muscle form is not calcium sensitive. Both forms have two EF-hand domains at their C-terminii.

α-*Spectrin* (FDRN) was first described in the erythrocyte in which it reinforces the cell membrane by cross linking ankyrin and protein 4.1. Nonerythroid αII-spectrin and its close homolog, α-fodrin, form elongated double stranded filaments. Each strand consists of segments of triple stranded α-helices that are formed by the strand going forward, back, forward. The two EF-hands at the C-terminus impart a calcium sensitive (Lundberg *et al.* 1997) interaction with actin and other cytosolic proteins.

*Calbindin D*28k (CLBN) is found in intestinal epithelium and may be involved in transcellular transport of calcium; however, it is not congruent with ICBP, sometimes referred to as calbindin 9kDa. Domains 1, 3 & 5 and domains 2, 4 & 6 are more closely related. Interdomains 1,2 and 3,4 and 5,6 are similar and are inferred to be homologous. CLBN appears to have evolved by two duplications and fusions from a two domain precursor. The six domains, individually synthesized, aggregate in pairwise combination to approximate the spectral properties of the intact CLBN (Linse *et al.* 1997).

Tetrahymena pyriformis calcium binding protein (TCBP) may be involved in the calcium sensitive changes in ciliary beating. The isoforms, TCBP-23 and TCBP-25 are 35% identical in amino acid sequence; however, their encoding DNA's are 49% identical. Lobe 1,2 most closely resembles lobe 3,4; the origin of the inferred two domain precursor is obscure (Takemasa *et al.* 1990).

p24 thyroid protein (TPP), or calcyphosine, consists of four EF-hands. Domains 3 and 4 resemble one another most closely; however, this deviation from the *ODD*EVEN* pattern is barely significant. The function of TPP is unknown. Another member of the subfamily has been found in lobster and crayfish muscle (Sauter *et al.* 1995).

Reticulocalbin (RTC) contains six EF-hand domains and is inferred to be in the lumen of the endoplasmic reticulum because it has an N-terminal leader sequence and a glycosylation site. The C-terminal sequence is HDEL; this is similar to the KDEL that serves as a signal to retain proteins within the ER. RTC is not closely related to the other known six EF-hand protein, CLBN (Ozawa & Muramatsu, 1993). Four of six EF-hands (1, 4, 5 & 6) bind calcium but 2 & 3 do not (Tachikui *et al.* 1997)

Plasmodium falciparum surface protein (PFS) has five EF-hand domains and has other sorts of domains at both N- and C-terminii. Its fifth domain, like that of SORC, might function as a linker in dimerization (Rawlings & Kaslow, 1992). Templeton *et al.* (1997) found that PFS is localized not to the surface but to an intracellular membrane.

1F8 and *TB17* (1F8) cDNAs were prepared from highly redundant mRNA's from *Trypanosoma cruzi* (Engman *et al.*, 1989) and *T. brucei* (Lee *et al.* 1990). 1F8, or calcimedin, (Wu *et al.* 1992) is located in the flagellum.

Sarcoplasm calcium binding protein (SARC) binds three Ca^{2+} or Mg^{2+} ions competitively. It is found both as a monomer and as a dimer; it may function as a calcium buffer. Neither the sequences of its four EF-hands nor of the three interdomain linkers closely resemble other EF-hands or linkers. Crystal structures of SARC are available from the sandworm, *Nereis diversicolor*, (Cook *et al.* 1991) and from amphioxus, *Branchiostoma lanceolatum* (Cook *et al.* 1993). Although the sequences share only 12% identity overall, the main chain traces of the two are nearly superimposable. All four domains have calcium bound. The hydrophobic faces of lobes 1,2 and 3,4 contact one another.

Calcium binding protein from Saccharopolyspora erythraea (CMSE) is the only known EF-hand protein from eubacteria (Swan *et al.* 1989); none are known from archea. Its indicated congruence with SARC is tenuous. The function of neither is known.

Aequorin (AEQ) (EC 1.13.12.5 renilla-luciferin 2-monooxygenase) is a calcium-dependent photoprotein that oxidizes luciferin (coelenterazine) and thereby produces the luminescence of the marine coelenterate, *Aequoria victoria* (Prasher *et al.* 1987). In the anthozoan, *Renilla reniformis*, calcium induced bioluminescence involves two proteins – luciferase and luciferin binding protein. LBP has four EF-hands but lacks the luciferase activity of aequorin (Kumar *et al.* 1990). AEQ has only a few residues N- and C-terminal to the four EF-hand domains; perhaps its luciferase activity is associated with its aberrant second domain.

Protein phosphatase (PPTS) (EC 3.1.3.16) is encoded by the *rdgC* gene of *Drosophila*; it has a phosphoprotein phosphatase domain at its N-terminus. Three EF-hands are easily recognized and indicated in table 1 as 1, 3 & 4 (Steele *et al.* 1992). Nominal domain 2, occupying 56 residues between domains 1 and 3, may be a divergent EF-hand paired with domain 1.

LAV1 (LAV) cDNA encodes a 355 residue protein. Different isoforms are expressed in amoebae and in plasmodia of *Physarum polycephalum* (Laroche *et al.* 1989). Although LAV appears not to be congruent with any other EF-hand protein, its interdomains 1,2 and 3,4 resemble 1,2 and 3,4 of *CTER*. Linker 2,3 consists of a sole Leu. This places sever constraints on the spatial relationship between the lobe 1,2 and the C-terminal lobe. The function of LAV is unknown and no homolog has been identified for the ~200 residues at its N-terminal.

Calcium dependent protein kinase (CDPK) (EC 2.7.1.37) consists of an N-terminal kinase domain (homologous to the b subunit of Ca/CAM-dependent protein kinase II) followed by four EF-hands. A pseudosubstrate sequence and a CAM-binding motif are found at the junction region between the kinase domain and the four EF-hands. The range of substrates of CDPK include membrane ATPase (Saijo *et al.* 1997).

Plasmodium falciparum protein kinase (PFPK) (EC 2.7.1.37) is encoded by the *PfCPK* gene. Residues 56–325 encode a region homologous to CDPK, a protein kinase from plant, with which it shares ~40% sequence identity (Zhao *et al.* 1993). However, the four EF-hand domains of PFPK are not congruent with *CTER*, as are those of CDPK. The two chimeric kinases probably acquired their respective four EF-hands by different duplication and splicing events.

PM129 clone from Arabidopsis thaliana (PMAT) cDNA, enriched for plasma membrane associated proteins, contains a single continuous open reading frame (Bartling *et al.* 1993). The 37 residues before the first of four EF-hand domains may comprise a distinct domain. Each of the linkers 1,2 2,3 and 3,4

is eight residues long. There are three Cys's in the fourth EF-hand and one in the 3,4 linker; a disulfide bond could be formed without distorting the canonical EF-hand.

Phospholipase C (PLC) (EC 3.1.4.10) hydrolyzes phosphatidylinositol 4,5-bisphosphate to inositol 1,4,5-trisphosphate and diacylglycerol. It has an absolute requirement for calcium and is activated by guanine nucleotide binding proteins (PLC-β), or by receptor tyosine kinases (PLC-γ), or by transglutaminase II (PLC-δ). All three vertebrate forms of PLC have a plekstrin homolog domain, followed by four EF-hands, a triosephosphate isomerase-like β-barrel catalytic domain, and a C2 domain (Grobler & Hurley 1998). Although calcium is absolutely required for activity, it is bound to the catalytic domain. None of the four EF-hands have canonical calcium binding loops nor is calcium bound in the crystal structure (Essen *et al.* 1996).

Fimbrin (FIMB) is an actin filament bundling protein from chick. It has two EF-hands at its N-terminus and four, 125 residue actin binding domains, homologous to those of α-actinnin, at its C-terminus (de Arruda *et al.* 1990). The human homologs are I, L, and T-plastin whose genes are on chromosomes 3, 13, and X (Lin *et al.* 1994).

BM-40 (BM40), also known as osteonectin or SPARC (secreted protein acidic and rich in cysteines), consists of a short stretch of acidic amino acids, a region homologous to follistatin and an extracelluar calcium (EC) binding module. At least four other extracellular glycoproteins – testican, tsc, Qr1 and SC1 – all of unknown function, have homologous follistatin and EC modules. The latter half of the EC module, residues 209–286, consists of a pair of EF-hands (α-helices D, E, F & G of the EC module) as confirmed in the crystal structure (Hohenester *et al.* 1996). The second EF-hand binds calcium in canonical coordination. The first EF-hand binds calcium in the third (after S100 and regulatory light chain of molluscs) known variation of the canonical pattern of calcium binding (figure 4). Since BM40 is extracellular, calcium binding supposedly imparts stability, not information, to it. The cleft between the two EF-hands is occupied by the first α-helix of the EC module, 140–160, in an orientation very similar to that of a target helix bound by calmodulin; this appears to be another example of stabilization, not information transduction.

Calmodulin related gene product (CRGP) is encoded by the gene *T+*, which contains a transposon-like human repeat element, *THE1*, in the 3' untranslated region of its message (Deka *et al.* 1988).

Conclusions

Several of our conclusions were anticipated in the introduction and are elaborated. Some are general and could apply to many families; other are specific to the EF-hand family.

When the structure of parvalbumin was determined nearly three decades ago (Kretsinger 1972) it seemed quite daring to think that it evolved by gene triplication and that troponin C might share the EF-hand domain. One now marvels that there are 45 subfamilies. Yet many protein (domain) homolog families contain scores to hundreds of subfamilies. If only ~3000 different domains comprise over 60,000 different (subfamilies of) proteins, it is hardly surprising to find each domain put to multiple uses. Whether this ~3000 reflects some thermodynamic and/or steric limit on possible domains or whether evolution found it easier to adapt an existing domain than to create a new domain is one of the fundamental questions in evolution.

The concept of the unique origin and subsequent divergent evolution is fundamental to understanding the evolution of any protein. One of its fundamental implications is that each domain in a chimeric protein must be evaluated independently of the other domains in that protein; hence, the importance of congruence. Within each of the 29 subfamilies with more than one representative, the dendrograms of the domains are congruent (to statistical error) as demanded by the definition of a subfamily of chimeric proteins, for example TNC, Figure 1.

The members of two, or several, subfamilies may be congruent with one another. This is illustrated for S100, ICBP, and HYFL in figure 2. Several other triplets – CLNB, P22, VIS – or pairs –CALP & SORC, SPEC & LPS, DGK & GPD, and ACTN & FDRN – are also locally congruent.

The question whether to 'lump' or to 'split' surely preceded Linnaeus. For instance, the S100 proteins and intestinal calcium binding protein had been considered to be in the same subfamily. They are now in two distinct subfamilies based on separation in their joint dendrogram, differences in structures, inferred functions, and chromosomal loci of encoding genes. ICBP is a monomer and has been suggested to function in transcellular transport of calcium. Several S100's have been demonstrated to dimerize and at least one, A10 (p11), interacts with two molecules of annexin II (calpactin). The gene encoding ICBP is on X; those encoding S100 are on chomosomes 1, 4, and 21. HYFL is also a distinct subfamily; profilaggrin and trichohyalin both have ~2000 residues C-terminal to their two EF-hands.

Canonical

```
                                        ---------- Loop ----------
αααααααα E αααααααα                                  αααααααα F αααααααααα
                                  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2
1  2  3  4  5  6  7  8  9  0  1  2  3  4  5  6  7  8  9  0  1  2  3  4  5  6  7  8  9
E  n  *  *  n  n  *  *  n  X  *  Y  *  Z  G  #  I -X  *  *-Z  n  *  *  n  n  *  *(n)
```

S-100, domain **1**

```
                                 -8    -5   -3    0          +5
                                                            2  2  2  2  2  2  2  2  2
1  2  3  4  5  6  7  8  9                                   1  2  3  4  5  6  7  8  9
*  n  *  *  n  n  *  *  n  *  *  *E G  D  * * *  L  *  K  * E  n  *  *  n  n  *  *  *
                           X      Y  Z   -Y                -Z
                           =O     =O =O  =O
```

ELC, domain **1**

```
                                                  1  1  1  2  2  2  2  2  2  2  2  2  2
1  2  3  4  5  6  7  8  9                        7  8  9  0  1  2  3  4  5  6  7  8  9
$  n  *  *  V  F  $  L  F  D F W D G R D  G * V  D  A  n  K  n  G  D  n  n  R  C  L
                           X     -X    -Y
                           Y     Z          -Z
                           =O    =O         =O
```

BM40, domain **1**

```
                                          1  1  1  1  1  1  2  2  2  2  2  2  2  2  2
1  2  3  4  5  6  7  8  9  0  1  2  3  4  5  6  7  8  9  0  1  2  3  4  5  6  7  8  9
P  n  *  W  *  n  *  *  L  D  *H~P *  D  *  #  L -X  *  *-Z  n  *  P  n  *  *  P  n
                          =O
```

Figure 4. The canonical EF-hand consists of (-helix E, (residues 1 – 10), loop around the Ca^{2+} ion (10 – 21) and α-helix F (19 – 29). Residue 1 is often Glu ('E'); the insides of the helices usually have hydrophobic residues 'n'. The side chains of five residues, approximated by the vertices of an octahedron (X, Y, Z, -X, & -Z) provide oxygen atoms to coordinate calcium; residue 16, '#' at -Y, bonds to the Ca^{2+} ion with its carbonyl oxygen. The calcium is actually seven coordinate in a pentagonal bipyramid with X and –X vertices. There are five oxygen atoms in the Y, Z plane; since the –Z ligand, usually Glu, coordinates calcium with both oxygen atoms of its carboxlyate group. A Gly, 'G', at 15 permits a tight bend; 'I' at 17 has a hydrophobic side chain that attaches the loop to the hydrophobic core of the pair of EF-hand domains.

Three variations to this canonical calcium coordination scheme have been confirmed in crystal structures. In S100 there are two residues inserted into the first loop of those domains 1 that bind calcium. The Glu at –Z coordinates with both oxygens of its carboxylate group. Four carbonyl oxygens (X, Y, Z, & -Y) coordinate calcium. The seventh ligand is the oxygen of water. Loop 1 of molluscan ELC also has two residues inserted but not at homologous positions to those inserted in loop 1 of S100. Asp10 coordinates calcium or magnesium with both carboxylate oxygen (X) and carbonyl oxygen (Y); Asp12 (Trp is inserted between 11 and 12) coordinates with both carboxylate (-X) and carbonyl oxygen (Z); Asp14 (Arg is inserted between 13 and 14) uses its carboxylate (-Y); and a carbonyl oxygen comes from 16 (-Z). The '$…$' in the molluscan ELC's is ether Glu…Asp or Asp…Glu. Loop 1 of BM40 is similar to the canonical loop except that a Pro is inserted between 12 and 13. The His~Pro peptide bond is cis; the result is that the carbonyl oxygen of Pro bonds to calcium at Y instead of the usual side chain of residue 12. Pro's are at 24 and 28. For all three variants conserved or consensus residues that differ from canonical are indicated.

These local congruences probably reflect a more recent gene duplication and similarity of function. However, we cannot with confidence infer the relationships of these six locally congruent groups with the *CTER* group, described next.

Thirteen subfamilies – *C*AM, *T*NC, *E*LC, *R*LC, TPNV, CAL, SQUD, CDC, CVP, EFH5, CLAT, CAST, and PARV – comprising 51 domains are congruent, Figure 3. They are referred to collectively as *CTER*. The most parsimonious interpretation is that a gene duplication, fusion and reduplication generated the four domain precursor of all of these *CTER* subfamilies. If one of these subfamilies is found in only one kingdom, e.g. the animals, it is difficult to discern whether the gene duplication event occurred in the organism that was ancestral to the four kingdoms or whether it arose within the animal lineage. If one subfamily, e.g. CAM or RLC, is found in several kingdoms, it was probably present in the organism ancestral to all eukaryotes.

This pairing of *ODD* and *EVEN* domains seen in *CTER* has several important implications. First as noted above, it is most readily interpreted in terms of the gene encoding an *ur-domain* duplicating and fusing to form an *ODD*EVEN* pair. After some divergence a subsequent duplication generated the four domain precursor of *CTER*. Only ten of the 119 (170 – 51) domains of the remaining 32 (45 – 13) subfamilies depart from this *ODD*EVEN* pattern (Table 1).

Three subfamilies show significantly higher congruence between their own domains than between their domains and those of other subfamilies. TCBP1 and TCBP3 are closely related as are TCBP2 and TCBP4; certainly there was a recent two domain precursor; however, we cannot relate this TCBP*OD**TCBP*EV* precursor to other EF-hand proteins found in protists. Correspondingly, a CLBN*OD**CLBN*EV* precursor gave rise to a CLNB*OD1**CLBN*EV2* *CLBN*OD3**CLBN*EV4* precursor. Probably the first two domains were duplicated and spliced to form the six domain CLBN observed today. The situation in TPP is more complex; TPP3 and TPP4 are most similar but not closely related to TPP1 or to TPP2. Perhaps a three domain precursor duplicated its C-terminal domain.

Finally there are fifteen subfamilies whose domains show no discernable patterns of domain relationships with themselves or with other EF-hand domains. They do, however, honor the *ODD*EVEN* pattern next discussed .

In *CTER* the clusters of domains 1 and domains 3 are closely related, as are domains 2 and 4 (Figure 3). This implies that the gene encoding an *ur-EF-hand* domain duplicated and fused to form a gene encoding an *ODD*EVEN* pair of EF-hands. After some divergence a second gene duplication, fusion event produced the four domain precursor – *OD1*EV2*OD3*EV4* – of all *CTER* proteins.

There are four interesting variations on this *ODD*EVEN* generalization. We noted above the close similarity of TPP3 and TPP4. Parvalbumin has three domains; in linear sequence, N to C, they are most closely related to *EVEN, ODD, EVEN* domains of other EF-hand subfamilies. We conclude that the first domain was deleted and that the remaining three are properly labeled 2, 3 and 4. The unpaired domain 2 covers the potential hydrophobic patch of the 3–4 pair. In calpain and sorcin the unpaired fifth domain of the monomer pairs with another fifth domain to form a dimer. The other subfamily with an odd number of domains is the surface protein of *Plasmodium*; it might be expected to form a similar dimer. Aequorin and protein phosphatase both have second domains whose sequences are unrecognizable as EF-hands. We assign both as EF-hands on the precedents of *ODD*EVEN* pairs.

An approximate two-fold rotation axis relates the two domains of an *ODD*EVEN* pair. If the two EF-hands in a pair are equivalent; it is strange that an inversion of *ODD*EVEN* order so seldom occurs. Perhaps this assumed two fold equivalence is broken by residues at the C-terminus of the *ODD* domains, or by the *ODD*EVEN* linkers, or by residues at the N-terminus of the *EVEN* domains.

Why can we not trace with confidence the evolution of the 32 EF-hand subfamilies that are not in the *CTER* group back to **the** postulated *ur-EF-hand* or to **the** *ODD*EVEN* pair? It is conceivable that EF-hands originated several times and their progeny have structures that honor similar functional or thermodynamic constraints. Or there could have been a unique origin of the EF-hand with initial divergent evolution but with some pathways subsequently converging. Such homoplasy is observed in many contexts.

The conventional interpretation, the one that we favor, is that each of the (6 groups of paired, 3 self duplicated, and 15 miscellaneous) 24 groups has a complex evolution, hidden in the stochastic processes of mutation, gene duplication, and splicing. With 85 postulates we could 'explain' the evolutions of the 86 EF-hand domains in the 24 groups. We turn to the structures and functions of these proteins for help.

A model for the function of EF-hand proteins, such as calmodulin, states that: They are found in the cytosol or associated with a membrane facing

the cytosol. In the unstimulated cell they are in the apo state. Following stimulation, messenger calcium binds to the EF-hands causing a change in conformation. The calcium bound form activates a target enzyme or structural protein. This paradigm of information transduction provides a valuable reference for listing all of the exceptions.

For most of the EF-hand proteins the affinity for the Ca^{2+} ion is about $10^{4.2}$ greater than for the Mg^{2+} ion. This difference does not reflect an inherent affinity of the oxygen ligands for divalent cations; many small molecules have a higher affinity for magnesium than for calcium. It must reflect some selection for function. One implication is that in a cytosol with $[Mg^{2+}] \sim 10^{-2.8}$ M and $[Ca^{2+}] \sim 10^{-7.2}$ M an EF-hand protein, such as parvalbumin, with a high affinity for divalent cations binds magnesium whereas a protein, such as the (regulatory) lobe 1,2 of skeletal troponin C will have no cations bound. This leads to the counterintuitive situation in which the entering messenger calcium binds first to the lower affinity troponin C because the higher affinity parvalbumin has to wait until its bound magnesium slowly dissociates. The change in free energy of the system associated with binding calcium is dependent on the conformation of the entire molecule (or complex). One cannot predict calcium, or magnesium, affinity from the amino acid sequence, except by using close homologs.

One variation on this theme of transduction is to splice the EF-hand domains to their (inferred) targets as seen for the enzymes DGK, GPD, CALP, PPTS, CDPK, PFPK, and PLC. The second EF-hand domains of DGK and of GPD are similar; however, one does not know whether this implies similar spatial relationships to their respective catalytic domains. BM40 provided an interesting precedent. The cleft in its lobe is filled with an α-helix fifty residues N-ward. Some of the pairs of EF-hands in these hetero-chimeras may exert their regulatory effect by binding and releasing similar target helices within their own polypeptide chains.

Other heterochimeras appear not to be enzymes but structural or regulatory proteins – CLAT, CAST, HYFL, SORC, ACTN, FDRN, RTC, PFS, LAV, PMAT, and BM40. The four EF-hands of CLAT and CAST are congruent with *CTER*. The two EF-hands of HYFL are congruent with S100 and ICBP. The two EF-hands of ACTN and of FDRN are similar. Yet no general pattern emerges; these splicings that generated heterochimeras appear to have occurred independently of one another.

Twelve EF-hand proteins have been demonstrated or strongly inferred to transduce information to a target. CAM, TNC, ELC, RLC, TPNV, SQUD, CDC, and CVP are in the *CTER* group. CLNB, P22 and VIS are congruent with one another. To this extent one sees an evolutionary pattern. S100 may be involved in transduction but its congruent subfamilies ICBP functions in transcellular transport of calcium and HYFL is a heterochimera. BM40 is extracellular, there is no evidence of an information transducing function. Some S100s function in culture as neurite extension growth factors (Schäfer & Heizmann, 1996); whether this is their function in vivo has not been established.

Several proteins, notably parvalbumin and ICBP, have been proposed to function in calcium buffering and transport of calcium. To some extent these are default assignments made after no targets for these EF-hand proteins have been found.

There are in the known world of protein structures many examples of helix-loop-helix motifs that resemble EF-hands. We have been able to distinguish questionable domains by considering the structural context. For instance the two EF-hands of BM40 are in the twilight zone of sequence similarity; however, the α-carbons of the pair are superimposable on lobe 1,2 or 3,4 of calmodulin to within 1.5 Å r.m.s.

Usually protein domains consist of contiguous residues; the polypeptide strand does not weave back and forth between domains. Yet, the first α-helix of the extracellular, calcium binding module of BM40 lies in the cleft of the EF-hand pair fifty residues down stream. BM40 is not an enzyme and is not known to undergo conformational change; however, this phenomenon of looping back (or forward) to another domain is characteristic of self-inhibitory regions of enzymes.

The canonical EF-hand with calcium binding loop consists of 29 residues. There are three distinct variations on this calcium binding loop, figure 4, seen in domains 1 of S100, of ELC, and of BM40. All coordinate calcium with seven oxygen atoms in a pentagonal bipyramid. We assume that the *ur-ODD*EVEN* pair of EF-hands was competent to bind calcium. The mutations giving rise to these three variant loops appear to have occurred independently. At least a third of all EF-hand domains appear not to be competent to bind calcium but given the precedents of S100, ELC, and BM40, one is well advised to wait for the crystal structures. The important point is that there appear to be no system trends or generalizations as to which subfamilies bind calcium and, if so, with what affinity.

We cannot infer whether the assumed original, *ur-EF-hand*, bound calcium nor whether the Adam EF-hand could function without an Eve mate. The

details of the calcium binding loops of various domains are best explained, not by steady divergence but by extensive exploration of the possibilities of calcium coordination (or lack thereof). Again, we face the same statistical barrier, with only thirty residues and with only six calcium coordinating residues, it is difficult to reach strong conclusions regarding parallel or convergent evolution within the EF-hand family.

As fascinating as these evolutionary patterns may be, one is left frustrated by our inability to tease out the full sequence of duplications, translocations, and splicings that gave rise to these 45 subfamilies of EF-hand proteins.

Note added in proof

Vito *et al.* (1996) reported new EF-hand protein, ALG-2 that mediates calcium-regulated signal along the death pathway. This protein is classified tentatively as a member of SOCR subfamily.

Vito P. Lacana, E, D'Adamio L. 1996 Interfering with Apoptosis: Ca^{2+}-Binding Protein ALG-2 and Alzheimer's Disease Gene ALG-3. *Science* **271,** 521-525.

References

Adachi J, Hasegawa M. 1996 MOLPHY Version 2.3: Programs for molecular phylogenetics based on maximum likelihood. *Computer Science Monographs*, **28**, 1–150. Institute of Statistical Mathematics, Tokyo.

Ajioka J, Swindle J. 1993 The calmodulin-ubiquitin associated genes of *Trypanosoma cruzi*: their identification and transcription. *Mol. Biochem. Parasitol*. **57**, 127–136.

Altschul SF, Boguski MS, Gish W, Wootton JC. 1994 Issues in searching molecular sequence databases *Nature Genetics* **6**, 119 – 129.

Ames JB, Porumb T, Tanaka T, Ikura M, Stryer L. 1995 Amino-terminal myristoylation induces cooperative calcium binding to recoverin. *J. Biol. Chem*. **270**, 4526–4533.

Barroso MB, Bernd KK, DeWitt ND, Chang A, Mills K, Stzul ES. 1996 A novel Ca^{2+}-binding protein, p22, is required for constitutive membrane traffic. *J. Biol. Chem*. **271**, 10183–10187.

Bartling D, Bülter H. Weiler EW. 1993 *Arabidopsis thaliana* cDNA encoding a novel member of the EF-Hand superfamily of calcium-binding proteins. *Plant Physiol*. **102**, 1059–1060.

Bearer EL, DeGiorgis JA, Jaffe H, Medeiros NA, Reese TS. 1996 An axoplasmic myosin with a calmodulin-like light chain. *Proc. Natl. Acad. Sci. USA* **93,** 6064–6068

Berchtold MW. 1993 Evolution of EF-hand calcium-modulated proteins. V. The genes encoding EF-hand proteins are not clustered in mammalian genomes. *J. Mol. Evol*. **36**,489–496.

Biggins S, Ros, MD. 1994 Direct interaction between yeast spindle pole body components: Kar1p is required for Cdc31p localization to the spindle pole body. *J. Cell Biol*. **125**, 843–852.

Blanchard H, Grochulski P, Li Y, Arthr SC, Davies PL, Elce JS, Cygle M. 1997 Structure of a calpain Ca^{2+}-binding domain reveals a novel EF-hand and Ca^{2+}-induced conformational changes. *Nature Struct. Biol*. **4**, 532–538.

Brown LJ, MacDonald MJ, Lehn DA, Moran SM. 1994 Sequence of rat mitochondrial glycerol-3-phosphate dehydrogenase cDNA. Evidence of EF-hand calcium-binding domains. *J. Biol. Chem*. **269**, 14363–14366.

Brownawell AM, Creutz CE. 1997 Calcium-dependent binding of sorcin to the N-terminal domain of synexin (annexin VII). *J. Biol. Chem*. **272**, 22182–22190.

Chazin WJ. 1998 Structural aspects of calcium-binding proteins of the EF-hand type. *BioMetals* this issue.

Chothia C. 1992 One thousand families for the molecular biologist. *Nature* **357**, 543–544.

Cook WJ, Ealik SE, Babu YS, Cox JA, Vijay Kumar S. 1991 Three-dimensional structure of a sarcoplasmic calcium-binding protein from *Nereis divesicolor. J. Biol. Chem*. **266**, 652–656.

Cook WJ, Jeffrey LC, Cox JA Vijay Kumar S. 1993 Structure of a sarcoplasmic calcium-binding protein from amphioxus refined at 2.4 Å resolution. *J. Mol. Biol*. **229**, 461–471.

Cox JA, Alard P, Schaad O. 1990 Comparative molecular modeling of *Amphioxus* calcium vector protein with calmodulin and troponin C. *Prot. Engineering* **4**, 23–32.

Crivici A, Ikura M. 1995 Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct*. **24**, 85–116.

de Arruda MV, Watson S, Lin C-S, Leavitt J, Matsudaira P. 1990 Fimbrin is a homologue of the cytoplasmic phosphoprotein plastin and has domains homologous with calmodulin and actin gelatin proteins. *J. Cell Biol*. **111**, 1069–1079.

Deka N, Wong E, Matera AG, Kraft R, Leinwand LA, Schmid CW. 1988 Repetitive nucleotide sequence insertions into a novel calmodulin-related gene and its processed pseudogene. *Gene* **71**, 123–134.

Donaldson C, Barber KR, Kay CM, Shaw GS. 1995 Human S100b protein: Formation of a tetramer from synthetic calcium-binding site peptides. *Protein Sci*. **4**, 765–772.

Engman DM, Krause K-H, Blumin JH, Kim KS, Kirchhoff LV, Donelson, JE. 1989 A novel flagellar Ca^{2+}-binding protein of trypanosomes. *J. Biol. Chem*. **264**, 18627–18631.

Essen L-O, Perisic O, Cheung R, Katan M, Williams RL. 1996 Crystal structure of a mammalian phosphoinositide specific phospholipase Cδ. *Nature* **380**, 595–602.

Finn BE, Forsén S. 1995 The evolving model of calmodulin structure, function and activation. *Structure* **3**, 7–11.

Flaherty KM, Zozulys S, Stryer L, McKay DB. 1993 Three

dimensional structure of recoverin, a calcium sensor in vision. *Cell* **75**, 709–716.

Gellatly KS, Lefebvre DD 1993 Identification of a cDNA clone coding for a novel calcium-binding protein from potato tuber. *Plant Physiol*. **101**, 1405–1406.

Gilbert W. 1987 The exon theory of genes. *Cold Spring Harbor Symp. Quant. Biol*. **52**, 901–905.

Griffith JP, Kim JL, Kim EE, Sintchak MD, Thomson JA, Fitzgibbon MJ, Fleming MA, Caron PR, Hsiao K, Na via MA. X-ray structure of calcineurin inhibited by the immunophilin-immunosuppressant FKBP12-FK506 complex. 1995 *Cell* **82**, 507–522.

Grobler JA, Hurley JH. 1998 Catalysis by phospholipase Cδ requires that Ca^{2+} bind to the catalytic domain, but not the C2 domain. *Biochemistry* **37**, 5020–5028.

Hardin PE, Angerer LM, Hardin SH, Angerer RC, Klein WH. 1988 Spec2 genes of Strongylocentrotus purpuratus. Structure and differential expression in embryonic aboral ectoderm cells. *J. Mol. Biol*. **202**, 417–431.

Head JF. 1989 Amino acid sequence of a low molecular weight, high affinity calcium-binding protein from the optic lobe of the squid, *Loligo pealei. J. Biol. Chem*. **264**, 7202–7209.

Heierhorst J, Kobe B, Feil SC, Parker MW, Benian GM, Weiss KR, Kemp BE. 1996 Ca2+/S100 regulation of giant protein kinases. *Nature* **380**, 636–639.

Hohenester E, Mauer P, Hohenadl D, Timpl R, Jansonius JN, Engel J. 1996 Structure of a novel extracellular Ca^{2+}-binding module in BM-40. *Nature Struct. Biol*. **3,** 67–73.

Houdusse A, Cohen C. 1996 Structure of the regulatory domain of scallop myosin at 2 Å resolution: Implications for regulation. *Structure* **4**, 21–32.

Kawasaki H, Kretsinger RH. 1995 Calcium-binding proteins 1: EF-hands. *Protein Profile* **2**, 305–490.

Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, Tempczyk A, Kalish VJ, Tucker KD, Showalter RE, Moomaw EW, Gastinel LN, Hakuba N, Chen X, Maldonado F, Barker JE, Bacquet R, Villafranca JE. 1995 Crystal structure of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature* **378**, 641–644.

Kretsinger RH. 1972 Gene triplication deduced from the tertiary structure of a muscle calcium binding protein. *Nature New Biology* **240**, 85–88.

Kumar S, Harrylock M, Walsh KA, Cormier MJ, Charbonneau H. 1990 Amino acid sequence of the Ca^{2+}-triggered luciferin binding protein of *Renilla reniformis*. *FEBS Lett*. **268**, 287–290.

Laroche A, Lemieux G, Pallotta D. 1989 The nucleotide sequence of a developmentally regulated cDNA from *Physarum polysephalum. Nucleic Acids Res*. **17**, 10502–10502.

Lee MG-S, Chen J, Ho AWM, D'Alesandro PA, Van der Ploeg LHT. 1990 A putative flagellar Ca^{2+}-binding protein of trypanosomatid protozoan parasites. *Nucleic Acids Res*. **18**, 4252–4252.

Lee S-C, Kim I-G, Marekov LN, O'Keefe EJ, Parry DAD, Steinert PM. 1993 The structure of human trichohyalin. Potential multiple roles as a functional EF-hand-like calcium-binding protein, a cornified cell envelope precursor, and an intermediate filament-associated (cross-linking) protein. *J. Biol. Chem*. **268**, 12164–12176.

Lin CS, Shen W, Chen ZP, Tu YH, Matsudaira P. 1994 Identification of I-plastin, a human fimbrin isoform expressed in intestine and kidney. *Mol. Cell. Biol*. **14**, 2457–2467.

Lin X, Barber DL. 1996 A calcineurin homologous protein inhibits GTPase-stimulated Na-H exchange. *Proc. Natl. Acad. Sci*. **93**, 12631–12636.

Ling V, Zielinski RE. 1993 Isolation of an *Arabidopsis* cDNA sequence encoding a 22 kDa calcium-binding protein (CaBP-22) related to calmodulin. *Plant Mol. Biol*. **22**, 207–214.

Linse S, Thulin E, Gifford LK, Radzewsky D, Hagan J, Wilk RR, (kerfeldt KS. 1997 Domain organization of calbindin D_{28K} as determined from the association of six synthetic EF-hand fragments. *Prot. Sci*. **6**, 2385–2396.

Lokuta AJ, Meyers MB, Sander PR, Fishman GI, Valdivia HH. 1997 Modulation of cardiac ryanodine receptors by sorcin. *J. Biol. Chem*. **272**, 25333–25338.

Lundberg S, Buevich AV, Sethson I, Edlund U, Backman L. 1997 Calcium-binding mechanism of human nonerythroid α-spectrin EF-structures. *Biochemistry* **23**, 7199–7208.

Markova NG, Marekov LN, Chipev CC, Gan S-Q, Idler WW, Steinert PM. 1993 Profilaggrin is a major epidermal calcium-binding protein. *Mol. Cell. Biol*. **13**, 613–625.

Nakayama S, Kretsinger RH. 1994 Evolution of the EF-hand family of proteins. *Annu. Rev. Biophys. Biomol. Structure* **23**, 473–507.

Orengo CA, Jones, DT, Thornton JM. 1994 Protein superfamilies and domain superfolds. *Nature* **372**, 631–634.

Ozawa M, Muramatsu T. 1993 Reticulocalbin, a novel endoplasmic reticulum resident Ca^{2+}-binding protein with multiple EF-hand motifs and a carboxyl-terminal HDEL sequence. *J. Biol. Chem.* **268**, 699–705.

Prasher DC, McCann RO, Longiaru M, Cormier MJ. 1987 Sequence comparisons of complementary DNAs encoding aequorin isotypes. *Biochemistry* **26**, 1326–1332

Rawlings DJ, Kaslow DC. 1992 A novel 40-kDa membrane-associated EF-Hand calcium-binding protein in *Plasmodium falciparum. J. Biol. Chem*. **267** 3976–3982.

Rayment I, Rypniewski WR, Schmidt-Bäse K, Smith R, Tomchick DR, Benning MM, Winkelmann DA, Wesenberg G, Holden HM. 1993 Three-dimensional structure of myosin subfragment-1: A molecular motor. *Science* **261**, 50–58.

Rhoads A, Friedberg F. 1997 Sequence motifs for calmodulin recognition. *FASEB J*. **11**, 331–340.

Rustandi RR, Drohat AC, Weber DJ. 1998 The Ca^{2+}-dependent interaction of S100B(ββ) with a peptide derived from p53. *Biochemistry* **37**, 1951–1960.

Saijo Y, Hata S, Sheen J, Izui K. 1997 cDNA cloning and prokaryotic expression of maize calcium-dependent protein kinases. *Biochim. Biophys. Acta* **1350**, 109–114.

Salvato M, Sulston J, Albertson D, Brenner S. 1986 A novel calmodulin-like gene from the nematode *Caenorhabditis elegans. J. Mol. Biol*. **190**, 281–289.

Sauter A, Staudenmann W, Hughes GJ, Heizmann, CW. 1995 A novel EF-hand Ca^{2+}-binding protein from abdominal muscle of crustaceans with similarity to calcyphosine from dog thyrodea. *Eur. J. Biochem* **227**, 97–101.

Schäfer BW, Wicki R, Engelkamp D, Mattei M-G, Heizmann CW. 1995 Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21: Rationale for a new nomenclature of the S100 calcium-binding protein family. *Genomics* **25**, 638–643.

Schäfer BW, Heizmann CW. 1996 The S100 family of EF-hand calcium-binding proteins: functions and pathology. TIBS 21, 134–140.

Smith SP, Shaw GS. 1998 A novel calcium-sensitive switch revealed by the structure of human S100B in the calcium-bound form. *Structure* **6**, 211 222.

Steele FR, Washburn T, Rieger R, O'Tousa JE. 1992 Drosophila *retinal degeneration C* (*rdgC*) encodes a novel serine/threonine protein phosphatase. *Cell* **69**, 669–676.

Swan DG, Cortes J, Hale RS, Leadlay PF. 1989 Cloning, characterization, and heterologous expression of the *Saccharopolyspora eythraea* (*Streptomyces erythraeus*) gene encoding an EF-Hand calcium-binding protein. *J Bacteriol*. **171**, 5614–5619.

Tachikui H, Navet AF, Ozawa M. 1997 Identification of the Ca^{2+}-binding domains in reticulocalbin, an endoplasmic reticulum resident Car^{2+}-binding protein with multiple EF-hand motifs. *J. Biochem. (Tokyo)* **121**, 145–149.

Takemasa T, Takagi T. Kobayashi T, Konishi K, Watanabe Y. 1990 The third calmodulin family protein in *Tetrahymena*. Cloning of the cDNA for *Tetrahymena* calcium-binding protein of 23 kDa (TCBP-23). *J. Biol. Chem*. **265**, 2514–2517.

Taylor KA, Taylor DW. 1993 Projection image of smooth muscle α-actinin from 2-D crystals formed on positively charged lipid layers *J. Mol. Biol*. **230**, 196–205.

Templeton TJ, Fujioka H, Aikawa M, Parker KC. 1997 *Plasmodium falciparum* Pfs40, renamed Pf39, is localized to an intracellular membrane-bound compartment and is not sexual stage-specific. *Mol. Biochem. Parasitol*. **90**, 359–365.

Vassylev DG, Takeda S, Wakatsuki S, Maeda K, Maeda Y. 1998 Crystal structure of troponin C in complex with troponin I fragment at 2.3 Å resolution. *Proc. Natl. Acad. Sci*. **95** 4847–4852.

Wong S, Morales TH, Neigel JE, Campbell DA. 1993 Genomic and transcriptional control linkage of the genes for calmodulin, EF-hand 5 protein, and ubiquitin extension protein 52 in *Trypanosoma brucei. Mol. Cell. Biol*. **13**, 207–216.

Wooding FBP, Morgan G, Jones GV, Care AD. 1996 Calcium transport and the localisation of calbindin D9k in the ruminant placenta during the second half of pregnancy. *Cell Tissue Res*. **285**, 477–489.

Wu Y, Haghighat NG, Ruben L. 1992 The predominant calcimedins from *Trypanosoma brucei* comprise a family of flagellar EF-hand calcium-binding proteins. *Biochem. J*. **287**, 187–193.

Xiang M, Ge T, Tomlinson CR, Klein WH. 1991 Structure and promoter activity of the *LpS1* genes of *Lytechinus pictus. J. Biol. Chem*. **266,** 10524–10533.

Yamada K, Sakane F, Matsushima N, Kanoh H. 1997 EF-hand motifs of α, β and γ isoforms of diacylglycerol kinase bind calcium with different affinities and conformational changes. *Biochem J*. **321**, 59–64.

Zhao Y, Kappes B, Franklin RM. 1993 Gene structure and expression of an unusual protein kinase from *Plasmodium falciparum* homologous at its carboxyl terminus with the EF Hand calcium-binding proteins. *J. Biol. Chem*. **268**, 4347–4354.