PROTEINS: Structure, Function, and Bioinformatics 65:643–655 (2006)

# Prediction of EF-Hand Calcium-Binding Proteins and Analysis of Bacterial EF-Hand Proteins

**Yubin Zhou, Wei Yang, Michael Kirberger, Hsiau-Wei Lee, Gayatri Ayalasomayajula, and Jenny J. Yang***
*Department of Chemistry, Georgia State University, Atlanta, Georgia 30303*

***ABSTRACT*** **The EF-hand protein with a helix–loop–helix $Ca^{2+}$ binding motif constitutes one of the largest protein families and is involved in numerous biological processes. To facilitate the understanding of the role of $Ca^{2+}$ in biological systems using genomic information, we report, herein, our improvement on the pattern search method for the identification of EF-hand and EF-like $Ca^{2+}$-binding proteins. The canonical EF-hand patterns are modified to cater to different flanking structural elements. In addition, on the basis of the conserved sequence of both the N- and C-terminal EF-hands within S100 and S100-like proteins, a new signature profile has been established to allow for the identification of pseudo EF-hand and S100 proteins from genomic information. The new patterns have a positive predictive value of 99% and a sensitivity of 96% for pseudo EF-hands. Furthermore, using the developed patterns, we have identified zero pseudo EF-hand motif and 467 canonical EF-hand $Ca^{2+}$ binding motifs with diverse cellular functions in the bacteria genome. The prediction results imply that pseudo EF-hand motifs are phylogenetically younger than canonical EF-hand motifs. Our prediction of $Ca^{2+}$ binding motifs provides not only an insight into the role of $Ca^{2+}$ and $Ca^{2+}$-binding proteins in bacterial systems, but also a way to explore and define the role of $Ca^{2+}$ in other biological systems (calciomics). Proteins 2006;65:643–655.** 

**Key words: EF-hand; S100; pattern search; bacterial genomes; prediction; evolution**

## INTRODUCTION

Calciomics describes the role of $Ca^{2+}$ in biological systems. $Ca^{2+}$, a messenger in cellular signal transduction, functions as a pivotal regulator of the cell life cycle including cell division, differentiation, and apoptosis.[1–5] The regulatory effects of $Ca^{2+}$ are influenced by the oscillation of intracellular $Ca^{2+}$ concentration, which ranges from submicromolar to millimolar levels.[6] $Ca^{2+}$ carries out its functions by binding to specific $Ca^{2+}$ receptors or $Ca^{2+}$-binding proteins (CaBPs). According to the role $Ca^{2+}$ ions or the proteins play in the biological context, most $Ca^{2+}$ binding proteins may fall into one of three categories: trigger or sensor proteins (e.g., calmodulin),[7] buffer proteins (e.g., S100G and parvalbumin),[8] or $Ca^{2+}$-stabilized proteins (e.g., thermolysin).[9] The EF-hand proteins can be found in each category and constitute more than 50% of all well-characterized $Ca^{2+}$-binding proteins. The EF-hand moiety is one of the most frequently used motifs in eukaryotic systems.[10]

Since the delineation of the EF-hand motif in 1973, the family of EF-hand proteins has expanded to include at least 66 subfamilies thus far.[11–13] EF-hand motifs are divided into two major groups: the canonical EF-hands as seen in calmodulin (CaM) and the prokaryotic CaM-like protein calerythrin [Fig. 1(A)], and the pseudo EF-hands exclusively found in the N-termini of S100 and S100-like proteins [Fig. 1(B)]. The major difference between these two groups lies in the $Ca^{2+}$-binding loop: the 12-residue canonical EF-hand loop binds $Ca^{2+}$ mainly via sidechain carboxylates or carbonyls (loop sequence positions 1, 3, 5, 12), whereas the 14-residue pseudo EF-hand loop chelates $Ca^{2+}$ primarily via backbone carbonyls (positions 1, 4, 6, 9) (Fig. 2). The residue at the -X axis coordinates the $Ca^{2+}$ ion through a bridged water molecule. The EF-hand loop has a bidentate ligand (Glu or Asp) at axis -Z. Among all the structures reported to date, the majority of EF-hand motifs are paired either between two canonical or one pseudo and one canonical motifs. For proteins with odd numbers of EF-hands, such as the penta-EF-hand calpain, EF-hand motifs were coupled through homo- or heterodimerization.[15–19]

Recently, EF-hand-like proteins with diversified flanking structural elements around the $Ca^{2+}$-binding loop have been reported in bacteria (Fig. 1).[20–22] Several lines of evidence indicate that these prokaryotic EF-hand-like proteins are widely implicated in $Ca^{2+}$ signaling and homeostasis in bacteria.[21,23–25] They contain flexible lengths of $Ca^{2+}$-binding loops that differ from the EF-hand motifs. However, their coordination properties resemble classical EF-hand motifs. For example, the semicontinuous $Ca^{2+}$-binding site in D-galactose-binding protein (GBP) contains a nine-residue loop (a.a.134-142). The $Ca^{2+}$ ion is coordinated by seven protein oxygen atoms, five of which are from the loop mimicking the canonical EF-loop whereas

Grant sponsor: NSF; Grant number: MCB-0092486; Grant sponsor: NIH; Grant numbers: GM 62999 and GM 070555.

*Correspondence to: Jenny J. Yang, Department of Chemistry, Georgia State University, University Plaza, Atlanta, GA 30302. E-mail: chejjy@langate.gsu.edu

Received 10 February 2006; Revised 12 May 2006; Accepted 24 June 2006

Published online 15 September 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21139

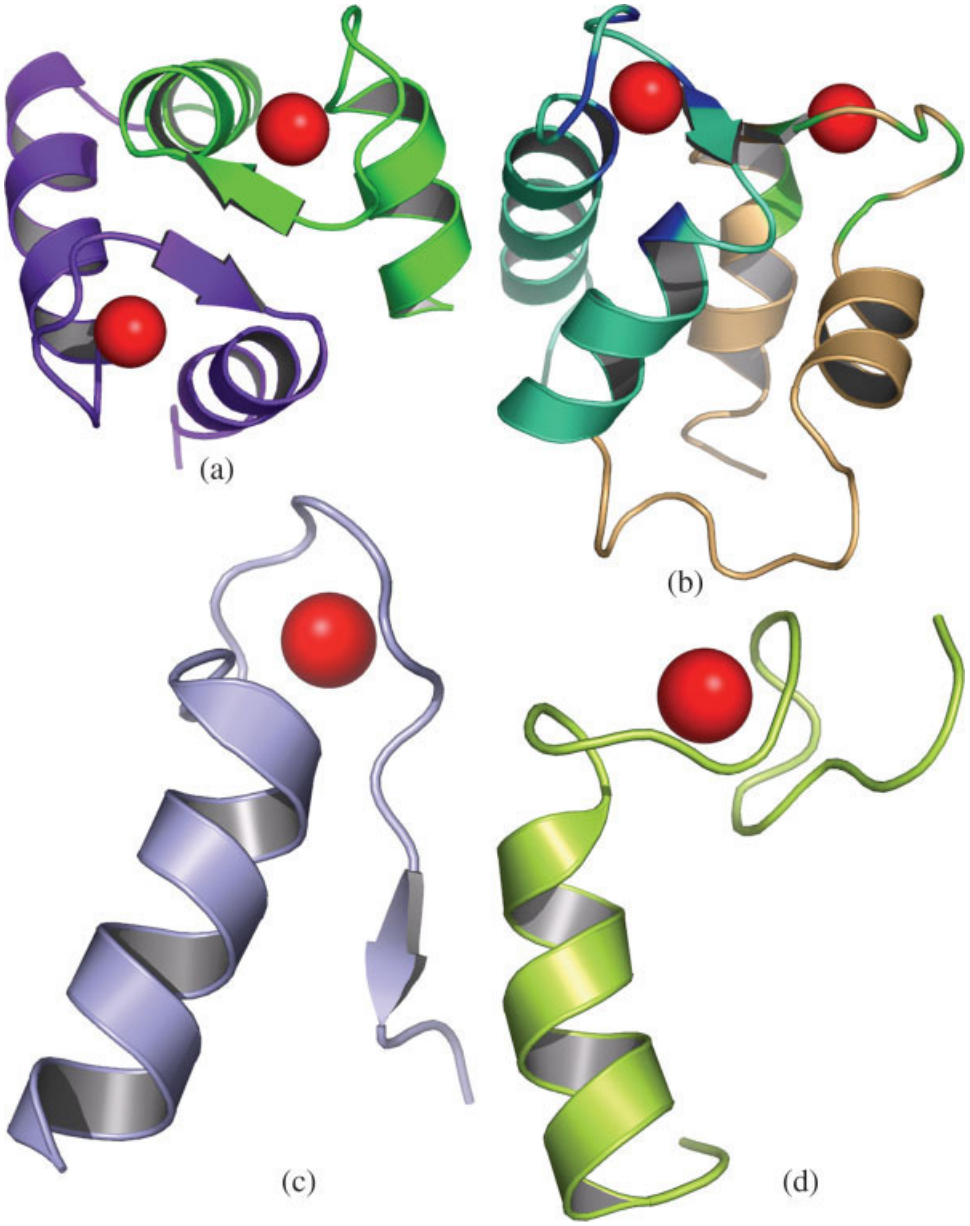



Fig. 1. Representative structures of canonical, pseudo EF-hand and EF-hand-like motifs. (**A**) Prokaryotic CaM-like protein calerythrin from *Saccharopolyspora erythrea* (PDB code: 1nya). (**B**) $Calbindin_{D9k}$ (PDB code: 1b1g). The N-terminal pseudo EF-hand (cyan) $Ca^{2+}$-binding ligands are highlighted by blue, while the C-terminal canonical EF-hand (gold) with green. (**C**) Periplasmic alginate-binding protein from *Sphingomonas sp.* (PDB code: 1kwh). The exiting helix was replaced by a β-strand. (**D**) The absence of entering helix at the EF-hand-like $Ca^{2+}$-binding site of dockerin from *Clostridium thermocellum* (PDB code: 1daq). The $Ca^{2+}$ ions are shown as red spheres.

the other two are from the carboxylate group of a distant Glu (a.a. 205). Another example is a novel domain named Excalibur (extracellular $Ca^{2+}$-binding region) isolated from *Bacillus subtilis*. This domain has a conserved 10-residue $Ca^{2+}$-binding loop strikingly similar to the canonical 12-residue EF-hand loop.[20] The diversity of the structure of the flanking regions is illustrated by the discovery of EF-hand-like domains in bacterial proteins. For example, a helix–loop–strand instead of the helix–loop–helix structure is seen in periplasmic galactose-binding protein (*Salmonella typhimurium*, 1gcg)[22] or alginate-binding protein (*Sphingomonas sp.*, 1kwh) [Fig. 1(C)][26]; and the entering helix is missing in protective antigen (*Bacillus anthracis*, 1acc)[27] or dockerin (*Clostridium thermocellum*, 1daq) [Fig. 1(D)].[28] Our studies have also shown that the single $Ca^{2+}$-binding loops from CaM are capable of binding $Ca^{2+}$ either alone or with the flanking helices when they are inserted into a non-$Ca^{2+}$-binding host protein CD2 domain 1 with β-

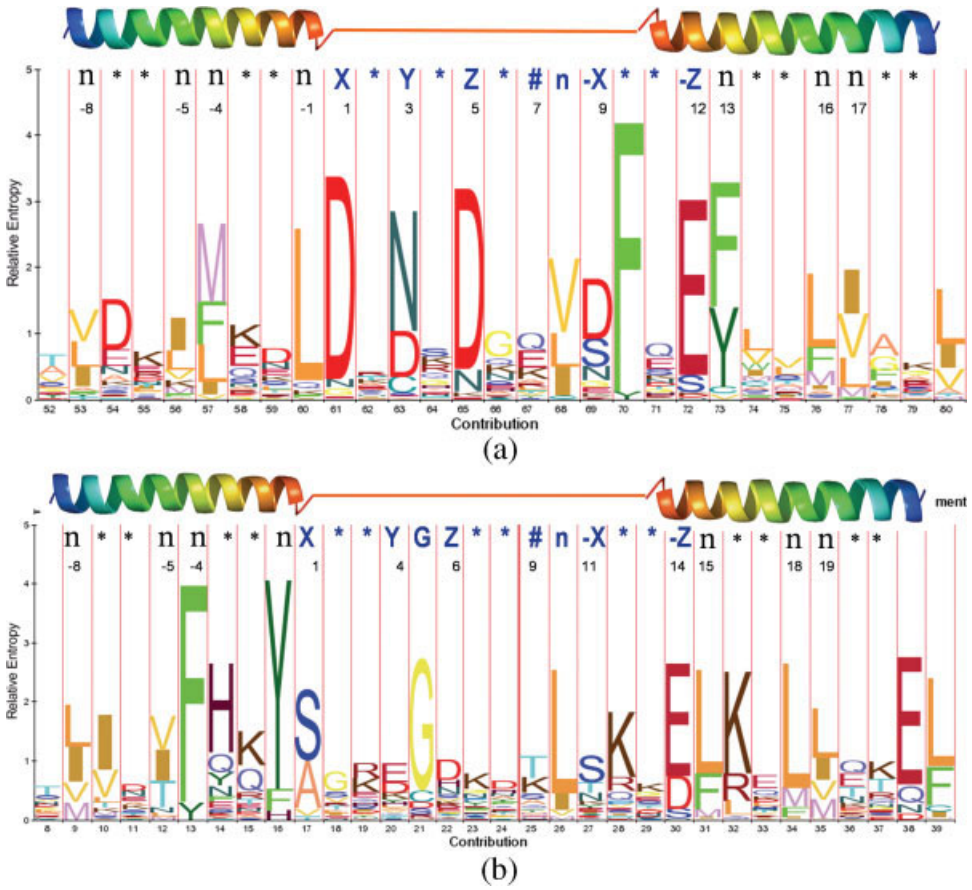



Fig. 2. Consensus sequence of canonical EF-hand (**A**) and pseudo EF-hand domains (**B**) drawn based on profiles HMM using LogoMat-M (http://logos.molgen.mpg.de/cgi-bin/logomat-m.cgi).[14] n: the hydrophobic residues within the flanking helices. #: the potential $Ca^{2+}$ binding ligands involving the mainchain carbonyl groups. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

strand structure.[29,30] The four EF-loops of CaM in the host protein have dissociation constants ($K_d$) ranging from 34 to 814 $\mu M$.[30] NMR studies revealed that the grafted EF-loop is directly involved in chelating $Ca^{2+}$.[31]

With the continuing expansion of genomic information, many efforts have been made to predict $Ca^{2+}$-binding proteins and to understand the role of $Ca^{2+}$ in biological systems. Pattern (motif signature) search is one of the most straightforward ways to predict continuous EF-hand $Ca^{2+}$-binding sites in proteins. Based on the sequence alignment results of canonical EF-hand motifs, especially the conserved side chains directly involved in $Ca^{2+}$ binding, a pattern, PS00018 (http://us.expasy.org/cgi-bin/nicesite.pl?PS00018), has been generated to predict canonical EF-hand sites. Alternative patterns have also been proposed with the addition of other conserved residues in the motif.[32,33] For pseudo EF-hand loop, however, each type of amino acid may serve as potential $Ca^{2+}$ binding ligands because of the use of the main chain, which makes prediction merely from the sequences relatively difficult. To circumvent this problem, the prediction of pseudo EF-hand sites was achieved by detecting the canonical EF-hands on the basis of the assumption that all the pseudo EF-hands are paired by a C-terminal canonical EF-hand. The currently available pattern PS00303 from EXPASY website (http://us.expasy.org/cgi-bin/nicedoc.pl?PDOC00275) predicts the S100 type $Ca^{2+}$ binding proteins by spanning the C-terminal canonical EF-hand motifs. It is worth pointing out that the prediction results obtained using this strategy do not directly provide the sequence of the pseudo EF-hand $Ca^{2+}$ binding loop.

Toward our goal of predicting and understanding the role of $Ca^{2+}$ in biological systems (denoted as calciomics), we report, herein, our progress in identifying EF-hand and EF-like motifs from the primary sequences. A series of patterns were generated by taking advantage of the metal binding properties of currently available EF-hand proteins and considering the helical structural context around the $Ca^{2+}$-binding loop. We modified the pattern PS00018 by allowing more choices (Glu, Gln, and Ser) at position 1 (axis X) and adding constraints at the flanking helical regions. By slightly loosening the constraints at the C-terminal canonical EF-hand, and simultaneously incorporating reserved residues in the N-terminal pseudo EF-hand, we also generated a modified pattern for the prediction of pseudo EF-hand sites. Compared with the

**TABLE I. The Pseudo EF-Hand Proteins**

| Protein | Synonyms | PDB codes | Accession number (species[a]) |
|---|---|---|---|
| S100A1 | S-100 protein alpha chain | 1k2h | P02639 (b), P23297 (h), P56565 (m), P35467 (r), Q7LZT1 (weatherloach) |
| S100A2 | S-100L, CAN19 | | P10462 (b), P29034 (h) |
| S100A3 | S-100E | 1kso | P62818 (m), P62819 (r), P33764 (h) |
| S100A4 | Metastasin, Calvasculin, Mts1 protein, 18A2, PEL98, Placental calcium-binding protein homolog, P9K, Nerve growth factor induced protein 42A | 1m31 | P35466 (b), P26447 (h), Q9TV56 (d), P07091 (m), P05942 (r) |
| S100A5 | S-100D | | P63084 (m), P63083 (r), P33763(h) |
| S100A6 | Calcyclin, Prolactin receptor associated protein, 5B10 | 1a03, 1cnp, 1jwd, 1k8u, 1k96, 1k9k, 1k9p, 2ncp | P14069 (m), P05964 (r), P06703 (h), P30801 (rb), Q98953 (c), O77691 (hs) |
| S100A7 | Psoriasin, Dermal allergen BDA11, Allergen Bos d 3 | 1psr, 2psr, 3psr | P31151 (h), Q28050 (b) |
| S100A8 | Calgranulin A, Neutrophil cytosolic 7 kDa protein P7, Migrion inhibitory factor-related protein 8, Chemotactic cytokine CP-10, MRP-8 | 1mr8 | P28782 (b), P05109 (h), P27005 (m), P50115 (r) |
| S100A9 | Calgranulin B, Neutrophil cytosolic 23 kDa protein, Migrion inhibitory factor-related protein 14 (MRP-14), P23, BEE22, P14, Leukocyte L1 complex heavy chain, Calprotectin L1H subunit | 1irj | P28783 (b), P06702 (h), P50117 (rb), P31725 (m), P50116 (r) |
| S100A10 | Calpactin I light chain, p10 protein, p11, Cellular ligand of annexin II, Nerve growth factor induced protein 42C | 1a4p, 1bt6 | P60902 (b), P60903 (h), P04163 (p), P620504 (rhesus macaque), P08207 (m), P05943 (r), P27003 (c), P27004 (African clawed frog) |
| S100A11 | Endothelial monocyte-activating polypeptide, Calgizzarin, S100C, MLN 70, EMAP | 1nsh, 1qls | P31949 (h), P24480 (rb), P50543 (m), Q6B345 (r), P31950 (p), P24479 (c) |
| S100A11P | Putative S100 calcium-binding protein A11 pseudogene | | O60417 (h) |
| S100A12 | Calgranulin C, CAGC, Calcium-binding protein in amniotic fluid 1, CAAF1, RAGE binding protein, Neutrophil S100 protein, p6 | 1e8a, 1gqm, 1odb | P79105 (b), P80310 (p), P80511 (h), O77791 (rb) |
| S100A13 | 8 kDa amlexanox-binding protein | | P79342 (b), Q99584 (h), P97352 (m) |
| S100A14 | S114 | | Q9HCY8 (h), Q9D2Q8 (m) |
| S100A15 | | | Q86SG5 (h) |
| S100A16 | S100F | | Q96FQ6 (h) |
| S100A17 | Clone: 5430400H23 product:hypothetical EF-hand/S-100/ICaBP type calcium binding protein | | Q9D3P1 (m) |
| S100B | S-100 protein beta chain | 1b4c, 1cfp, 1mho, 1dt7, 1mwn, 1psb, 1qlk, 1sym, 1uwo | P50114 (m), P04631 (r), P04271 (h), P02638 (b) |
| S100G | Vitamin D-dependent calcium-binding protein intestinal, Calbindin D9K, Cholecalcin | 1b1g, 1boc, 1bod, 1cb1, 1cdn, 1clb, 1d1o, 1ht9, 1ig5, 1igv, 1kcy, 1kqv, 1ksm, 1n65, 2bca, 2bcb, 3icb, 4icb | P29377 (h), P02632 (p), P02633 (b), P51964 (c), P02634 (r), P97816 (m) |
| S100H | Putative S100 calcium-binding protein H_NH0456N16.1 | | Q9UDP3 (h) |
| S100P | | 1ozo, 1j55 | P25815 (h) |
| S100Z | | | Q8WXG8 (h) |
| Hornerin | | | Q86YZ3 (h), Q8VHD8 (m) |
| Ictacalcin | | | Q91061 (channel catfish) |
| MRP-126 | | | P28318 (c) |
| Reptin | | | P97347 (m) |
| Trichohyalin | | | Q07283 (h), P37709 (rb), P22793 (s) |

[a]The abbreviation for species: b, bovine; c, chicken; d, dog; h, human; hs, horse; m, mouse; r, rat; rb, rabbit; s, sheep.

pattern PS00303, the new pattern, reflecting conserved genomic information in both the N- and C-terminal EF-hands, significantly improved the predictive accuracy and sensitivity. Finally we report our analysis of EF-hand proteins in bacterial genomes using the prediction method we developed. Our prediction results indicate that no pseudo EF-hand protein is found in the bacteria, which provides an additional piece of evidence suggesting that pseudo EF-hand motif may evolve later than canonical EF-hand motifs.

## MATERIALS AND METHODS

### Multiple Sequence Alignments and Phylogenetic Analysis

About 1904 proteins with potential canonical EF-hands and 84 proteins with pseudo-EF-hands (Table I) from SwissProt encompassing 66 distinct subfamilies of EF-hand proteins were included in our EF-hand databases. Typical members of each subfamily were collected to generate a subdatabase for multiple sequence alignments and phylogenetic analysis. Multiple sequence alignment (MSA) was performed using the ClustalW program with a gap open penalty of 10 and gap extension penalty set at 0.5.[34] The same program was applied to generate N-J tree for further display by the TreeView program.[35]

### Generation of Profile HMM and Patterns

Profile HMM (Hidden Markov Models) was generated from multiple sequence alignment results using HMMER by choosing both hmmbuild and hmmcalibrate algorithms. The statistical profile is subsequently visualized as HMM logo using LogoMat-M.[36,14] The EF-hand patterns were generated by taking into account highly conserved residues within both the pseudo and canonical EF-hand motifs.

### Evaluation of Canonic and Pseudo EF-Hand Pattern

The precision, sensitivity, and positive predictive values (PPV) of canonical EF-hand patterns (loop, eloop, loopf, and eloopf) were compared with that of the pattern PS00018. A total of 170 hits, including true positive, false negative, and false positive, were randomly selected from the results of PS00018 and set as the subdatabase for comparison. The newly generated pseudo EF-hand patterns, as well as the well-established S100 pattern PS00303, were used to search for possible pseudo-EF-hand $Ca^{2+}$ binding domains against major protein sequence databases such as SwissProt, iProClass, and NCBI reference sequences (RefSeq). The $Ca^{2+}$-binding properties of the proteins in the selected dataset have been experimentally verified and the prediction is compared with the verified information to determine the true positive, true negative, false negative, and false positive. Then the methods are applied to predict the proteins with unknown $Ca^{2+}$-binding properties in bacterial genomes. For statistical analysis, the precision, sensitivity, and PPV were determined as follows:

$$\text{Precision} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

where TP stands for true positive; TN, true negative; FN, false negative; FP, false positive.

## RESULTS AND DISCUSSION

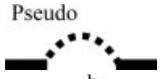
### Pattern Development

On the basis of the sequence alignment results and the statistical profile corresponding to each type of EF-hand $Ca^{2+}$-binding site, several patterns reflecting the most conserved information at particular positions have been developed and summarized in Table II. Among them, patterns 1–4, in addition to the commonly used pattern PS00018, can be used for the prediction of canonical EF-hand $Ca^{2+}$-binding sites with varying degrees of constraints on the sequences. Patterns PS00303 and PC (abbreviation of *p*seudo and *c*anonical EF-hands pattern) can be used for the prediction of $Ca^{2+}$-binding motifs within S100 and S100-like proteins. The patterns for EF-hand-like proteins are applied to the prediction of EF-hand-like $Ca^{2+}$-binding motifs with the loop length ranging from 10 to 15 residues.

### Canonical EF-Hand Motif

The widely applied pattern PS00018 has a stringent restraint at loop sequence position one (axis *X*) that only allows Asp, although Asn or Ser also occupy the position in a few EF-hands ($Ca^{2+}$ and integrin binding protein 2 (Q9Z309), CaBPE63-1 (P48593), rat CaM (pdb code: 3cln)). The pattern PS00018 focuses solely on the loop region and does not reflect conserved information within the flanking regions. To improve the pattern PS00018, we incorporated the diverse features of the flanking structural contexts and developed patterns catering to different constraints on EF-hands. On the basis of multiple sequence alignment results on over 1000 canonical EF-hands from the SwissProt protein sequence database, constraints on both flanking helices and the 12-residue loop were well defined in separate patterns. As shown in Table II, three patterns were derived: (a) x-{DNQ}-x(2)-{GP}-{ENPQS}-x(2)-{DPQR} for the entering helix; (b) [DNS]-x-[DNS]-{ILVFYW}-[DNESTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[ED] for the $Ca^{2+}$-binding loop; and (c) [FLYMVIW]-x(2)-{NPS}-{DENQ}-X(3) for the exiting helix. As revealed by the sequence alignment, hydrophobic residues are favored at positions −1, −4, −5, and −8 in the entering helix and at positions 13, 16, and 17 in the exiting helix [Fig. 2(A)]. Hence, hydrophilic residues or residues tending to interrupt helical structure were excluded at these positions.

**TABLE II. Summary of Patterns Used to Predict EF-Hand Proteins**

| Pattern number | Pattern name | Conservation positions | Motif signature |
|---|---|---|---|
| *To predict Canonical EF-hand* | | | |
| 1 | eloopf  | Canonical EF-loop and both flanking helices | x-{DNQ}-x(2)-{GP}-{ENPQS}-x(2)-{DPQR}-[DNS]-x-[DNS]-{FLIVWY}-[DNESTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[ED]-[FLYMVIW]-x(2)-{NPS}-{DENQ}-x(3) |
| 2 | eloop  | Canonical EF-loop and the entering helix | x-{DNQ}-x(2)-{GP}-{ENPQS}-x(2)-{DPQR}-[DNS]-x-[DNS]-{FLIVWY}-[DNESTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[ED] |
| 3 | loopf  | Canonical EF-loop and the exiting helix | [DNS]-x-[DNS]-{FLIVWY}-[DNESTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[ED]-[FLYMVIW]-x(2)-{NPS}-{DENQ}-x(3) |
| 4 | loop  PS00018[a] | Canonical EF-loop | [DNS]-x-[DNS]-{FLIVWY}-[DNESTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[ED] |
| 5 |  | Canonical EF-loop | D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW] |
| *To predict Pseudo EF-hand* | | | |
| 6 | PC  | Both helices of the pseudo EF-motif, both helices and the loop of the paired canonical EF-motif | [LMVITNF]-[FY]-x(2)-[YHIVF]-[SAITV]-x(5,9)-[LIVM]-x(3)-[EDS]-[LFM]-[KRQLE]-x(20,28)-[LQKF]-[DNG]-x-[DNSC]-x-[DNK]-x(4)-[FY]-x-[EKS] |
| 7 | Pseudo  PS00303[b] | Both helices of the pseudo EF-motif | [LMVITNF]-[FY]-x(2)-[YHIVF]-[SAITV]-x(5,9)-[LIVM]-x(3)-[EDS]-[LFM]-[KRQLE] |
| 8 |  | Both helices and the loop of the paired canonical EF-motif | [LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x-[ES]-[FYVC]-x(2)-[LIVMFS]-[LIVMF] |
| *To predict EF-hand-like* | | | |
| 9 | Excalibur[c]  | The 10-residue loop | D-x-D-x-D-G-x(2)-C-E |
| 10 | EF-hand-like  | The loop | D-x-[DNS]-{ILVFYW}-[DEN]-G-{GP}-x(5, 6)-[DE] |

[a]http://us.expasy.org/cgi-bin/nicesite.pl?PS00018.
[b]http://au.expasy.org/cgi-bin/nicesite.pl?PS00303.
[c]Rigden et al.[20]

The prediction of loop only (b), e-loop (a+b), loop-f (b+c), and e-loop-f (a+b+c) can then be achieved using the patterns in combination. This strategy provides an alternative way to perform prediction of EF-hands as well as EF-hand-like sites with deviations at the flanking regions.

Figure 3(A) shows the statistical results of the prediction of canonical EF-hand motifs using the pattern eloopf. According to the Prosite documentation PDOC00018, the pattern PS00018 results in more than 2000 hits in the SwissProt database. To compare our patterns with PS00018, a total of 170 protein sequences were randomly selected from the SwissProt database. Of these, 119 are true canonical EF-hand proteins with experimental validation, while 51 are not. With these sequences as the testing database, prediction results show that patterns 1–5 have similar sensitivity while the precision and PPV of patterns 1–3 increased by 10–20% when compared with the pattern PS00018. Fewer false positive hits were detected when using the patterns 1–3. Hence, additional constraints on the flanking regions enhance the overall accuracy of prediction and the true positive predictions.

### Pseudo EF-Hand Motif

As listed in Table I, pseudo EF-hands are mostly found in the S100 protein family and among members of the "fused gene" family, such as trichohyalin, horenin, and repetin.[37–42] The small, acidic S100 protein, calbindin$_{D9K}$ [Fig. 1(B)], carries two distinct EF-hands: a canonical EF-hand at the C-terminus and a pseudo EF-hand motif at the N-terminus. The canonical EF-hands are highly conserved among all the S100 proteins (Supplementary Fig. 1). However, there is a significant sequence variation in the $Ca^{2+}$-binding loop of the pseudo EF-hand [Fig. 2(B)]. The pseudo EF-hand within the S100A10 even loses the capacity to bind $Ca^{2+}$ ion because of the lack of chelating

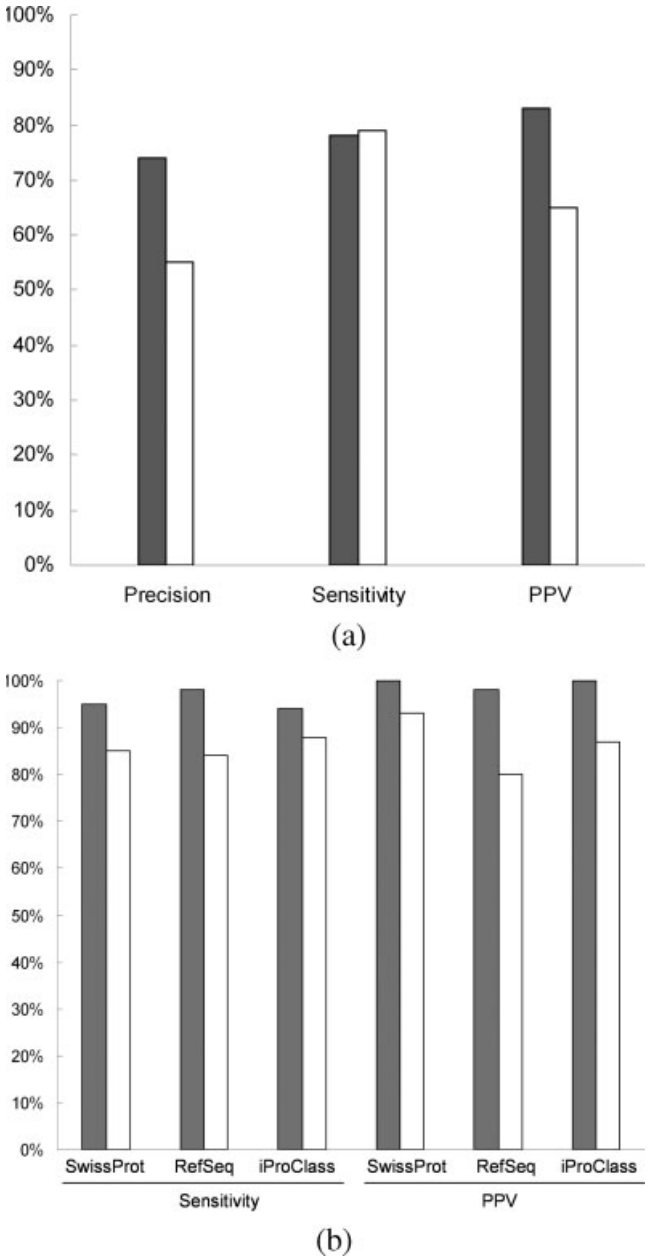



Fig. 3. (**A**) Prediction results using patterns eloopf (gray bar) and PS00018 (open bar). The patterns were applied to search for canonical EF-hand proteins in a test database containing 170 proteins, of which 119 are true EF-hand proteins with experimental validation and 51 are not. (**B**) Prediction results using the pseudo EF-hand pattern PC (gray bar) and PS00303 (open bar). Both patterns were used to search for potential pseudo EF-hand proteins against major protein sequence databases SwissProt, NCBI RefSeq, and iProClass.

ligands.[43] Therefore, one important concept that must be kept in mind is that the prediction of pseudo EF-hand does not presume the capability of binding $Ca^{2+}$ ion.

Using the multiple sequence alignment, we analyzed the $Ca^{2+}$-binding ligands of all pseudo EF-hands with known structures in Protein Data Bank (Table I). On the basis of the statistical results, a profile HMM and the resultant HMM logo were built [Fig. 2(B)]. Of the $Ca^{2+}$-binding ligands, Ser and Ala are preferred at loop position 1 (X); Glu dominates at both positions 4 (Y) and 14 (-Z); Gly and Leu are preferred at positions 5 and 10, respectively; Asp is most frequently found at position 6 (Z); and Thr and Lys reside with equal probability at position 9 (-Y). By integrating highly conserved residues located at the flanking helices (positions −1, −4, −5, 15, and 16), we generated a pattern (Table II, pattern pseudo) for the prediction of pseudo EF-hand $Ca^{2+}$-binding site. The pattern PC was further developed by incorporating the conserved signature in the downstream canonical EF-hand (positions −1, 1, 3, 5, 10, and 12).

To assess the performance of the developed patterns, we applied the patterns against major protein sequence databases such as SwissProt, iProClass (including PIR, trEMBL), and NCBI reference sequences (RefSeq). Figure 3(B) shows the comparison of the pattern PC and the pattern PS00303. A notable limitation of the pattern PS00303 is its failure to predict pseudo EF-hands within S100A13, S100A14, S100A16, S100A17, and S100P from some species because of stringent restraints at the C-terminal EF-hand motif. Moreover, since the prediction is based on the C-terminal canonical EF-hand, the prediction of PS00303 includes more false positive hits from the $Ca^{2+}$-binding proteins possessing only canonical EF-hands, such as calneuron 1 and $Ca^{2+}$-binding protein 7. In comparison with PS00303, 12.5% more true positive hits, on average, resulted with the pattern PC in the three databases. Meanwhile, the false positive hits were reduced by 10.5% on average. The average sensitivity and PPV of the pattern PC were 96% and 99%, respectively, which were 10% and 13% higher than those of the pattern PS00303 [Fig. 3(B)]. The pattern PC was able to identify the pseudo EF-hand motifs in at least three more subgroups of S100 proteins, including S100A13, S100A14, and S100A16 than the pattern PS00303 was. The first half of the pattern PC (or pattern pseudo) could be of great advantage in predicting S100-like proteins with deviations in the downstream canonical $Ca^{2+}$-binding loop or in predicting partially characterized incomplete hypothetical proteins. For instance, the pseudo EF-hand motifs in the novel $Ca^{2+}$-binding protein p26olf (named as the protein from frog olfactory epithelium)[44,45]could not be predicted by either PS00303 or PC since the C-terminal EF-hand contains an atypical EF-hand motif with a 4-residue insertion. However, without constraints on the C-terminus, the pattern pseudo (Table II) can easily detect them.

## Properties of EF-Hand-Like Motif

With the overall structural geometry of the $Ca^{2+}$ coordination remaining conserved, "EF-hand-like" motif refers to one containing the following deviations from the canonical EF-hand: (i) the length of the $Ca^{2+}$-binding loop is shorter or longer than 12-residues and/or (ii) the secondary structure elements of the flanking regions are not two helices. The first deviation can be represented in the motif signature by varying the length of the loop region (Patterns 9 and 10 in Table II). However, the structural devia-

**TABLE III. EF-Hand-like $Ca^{2+}$-Binding Proteins With Known Structure**

| PDB (resolution, Å ) and protein name | Deviation[a] | Sequence[b] (structure note) | Role of $Ca^{2+}$ | Organism |
|---|---|---|---|---|
| 1acc (2.10) Anthrax protective antigen | 3 | 168 STSAGPTVPDRDNDGIPDSLEVEGYTVDVKNKR 200 (loop–loop–helix) | Structural | *Bacillus anthracis* |
| 1daq Dockerin | 3 | 1 MSTKLYGDVNDDGKVNSTDAVALKRYVLR 29 (loop–loop–helix) | Structural | *Clostridium thermocellum* |
| | | 30 GISINTDNADLNEDGRVNSTDLGILKRY 58 (loop–loop–helix) | | |
| 1gcg (1.90) Glucose/galactose receptor | 4+5 | 125 KHWQANQGWDLNKDGKIQYVLLKGEP...E 205 (helix–loop–strand) | Structural | *Salmonella typhimurium* |
| 1h71(2.10) Alkaline protease | 4+5 | 49 QLTRSGASWHDLNNDGVINLTYTFLTA...D 114 (loop–loop–strand) | NA | *Psuedomonas sp.* |
| 1jv2 (3.10) Integrin αVβ3 | 1+3+4 | 275 YFGFSVAATDINGDDYADVFIGAPLFM 301 (strand–loop–strand) | Regulatory | *Homo sapiens* |
| 1kap (1.64) Alkaline protease | 3+4 | 446 ASKAGSLAIDFSGDAHADFAINLIGQA 472 (strand–loop–strand) | NA | *Psueomonas aeruginosa* |
| 1kwh (2.00) Alginate-binding protein | 1+4 | 162 TVLKAFKEKDPNGNGKADEVPFIDRHPDE 190 (helix–loop–loop) | Regulatory | *Sphingomona ssp.* |
| 1lwj (2.50) Glucanotransferase | 1 | 4 YQIYVRSFRDGNLDGVGDFRGLKNAVSYL 32 (helix–loop–helix) | NA | *Thermotoga maritima* |
| 1qut (2.44) Slt35 | 2 | 189 PSSYKQYAVDFSGDGHINLWDPVDAIGSVANY 220 (helix–loop–helix) | Structural | *Escherichia coli* |

[a]1 – shorter loop; 2 – longer loop; 3 – entering helix missing; 4 – exiting helix missing; 5 – distal located ligands.
[b]Ligand residues are underlined.

tion in the flanking regions can hardly be predicted merely from the sequences. Therefore, we conducted a retrospective search of EF-hand-like proteins in the PDB database (Table III). Four classes of EF-hand-like motifs are currently observed. The first class has a shorter loop (as seen in Excalibur) that contains a conserved 10-residue DxDxDGxxCE motif. The cysteine in the sequence may facilitate the orientation of the loop toward $Ca^{2+}$ binding by forming disulfide bonds.[20] The second class has a longer loop as seen in Slt35 (PDB code: 1qut), a soluble fragment of lytic transglycosylase B from *Escherichia coli* that has a 15-residue $Ca^{2+}$-binding loop flanked by two helices.[46,47] The third class lacks the entering helix as seen in protective antigen (PDB code: 1acc) from *Bacillus anthracis*[27] and dockerin from *Clostridium thermocellum* (PDB code: 1daq).[28] The fourth class lacks the exiting helix as seen in alginate-binding protein (PDB code: 1kwh) from *Sphingomonas sp*.[26] Some EF-hand-like proteins even infringe the EF-hand paradigm by possessing two or more types of deviations (Table III).

## EF-Hand Proteins in the Bacterial Genomes

To understand the roles of $Ca^{2+}$ in bacteria, we predicted putative EF-hand proteins in the bacteria genomes from the nonredundant REFerence protein database (NREF).[48] No pseudo-EF-hand motif was predicted using the pattern PC. A total of 467 EF-hand motifs in 397 entries of proteins were predicted using the pattern eloopf (Table II) for the canonical EF-hand motifs (supplementary Table I). There are 39 proteins that contain multiple EF-hand motifs ranging from 2 to 6. The other 358 proteins were predicted to contain mononuclear EF-hands. The roles of $Ca^{2+}$ in most of these proteins are yet to be characterized.

The 39 proteins with multiple EF-hand motifs, among which 16 proteins have been previously summarized,[23] are implicated in a variety of cellular activities, including $Ca^{2+}$ homeostasis,[49–51] chemotaxis,[22,52,53] scaffold protein binding,[54] resistance to acid stress[55,56] and so on. According to the sequence homology and assuming that they evolved from a common ancestor, they could be further classified into three major phylogenetic groups (Fig. 4). The first group includes calerythrin (*Saccharopolyspora erythrea*, P06495), calsymin (*Rhizobium etli,* AAG21376), putative glycosyl hydrolase (*Bacteroides fragilis*, NF02360737), a-xylosidase (*Bacteroides thetaiotaomicron*, NF01244792), and putative $Ca^{2+}$ binding proteins from the gram-positive bacterial genus streptomyces (*Streptomyces ambofaciens*, BAB19055; *Streptomyces coelicolor*, CAB76018, NP_628579,

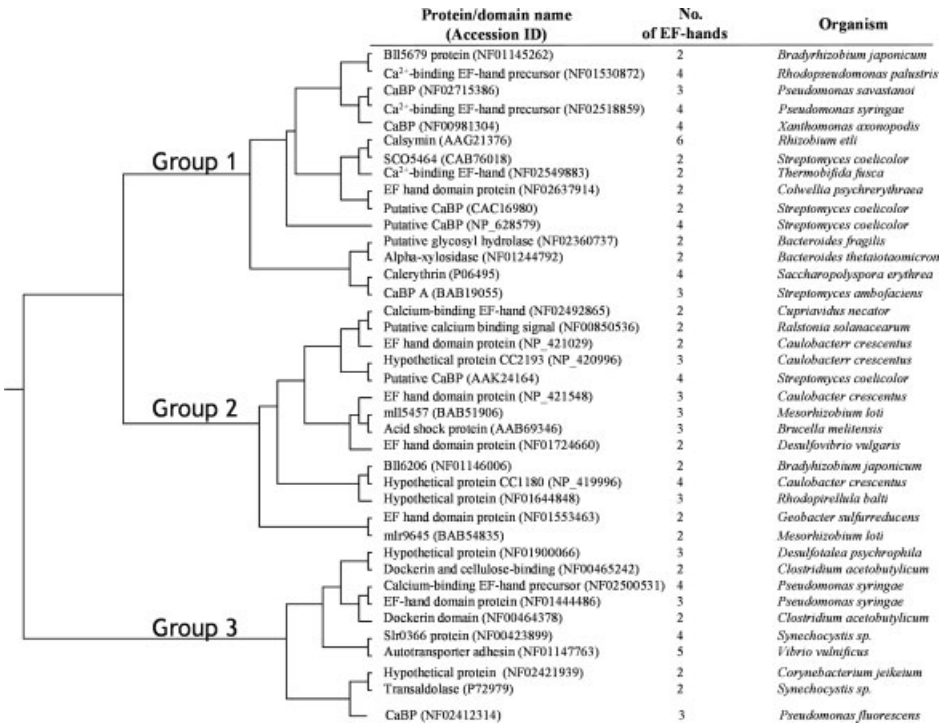



Fig. 4. Phylogenetic analysis of prokaryotic proteins containing multiple EF-hand motifs ranging from 6 to 2. The 39 proteins can be classified into three groups.

CAC16980). Calerythrin is the first characterized prokaryotic CaM-like protein possessing three canonical and an atypical EF-hand motif.[57,58] Two of the three high-affinity sites cooperatively chelate the metal ions and the apo protein adopts a molten global state conformation. It may function as either $Ca^{2+}$ buffer or transporter.[59] Being highly homologous to calerythrin, several members in this group (BAB19055, CAB76018, NP_628579, CAC16980, and NF02549883) are expected to adopt similar $Ca^{2+}$-dependent structural and biological behavior.[60] Another protein calsymin was implicated in symbiotic nitrogen fixation.[61] It contains three repeated homologous domains, each of which possesses two EF-hand motifs. Extracellular polysaccharide-degrading enzymes, putative glycosyl hydrolase, and α-xylosidase are involved in the metabolism of the bacterial wall. The $Ca^{2+}$ binding in the mesophilic xylanase and other members (NF02715683 and NF02518859) may protect the proteins from enzyme attack and thermal denaturation.[62]

The protein functions and the role of $Ca^{2+}$ in the second group are not well understood except for the acid shock protein (AAB69346) from the gram-negative bacterial genus *Brucella*. This protein is actively synthesized in response to the low pH to facilitate adaptation to acidic environments.[55] In addition, the putative EF-hand protein NF01724660 may link to the $Ca^{2+}$-induced aggregation of the sulfate-reducing bacteria *Desulfovibrio*.[63]

The third group encompasses dockerin (*Clostridium acetobutylicum*; NF00465242, NF00464378), bacterial transaldolase (*Synechocystis sp.*, P72797), adhesin (*Vibrio vulnificus*, NF01147763), and others with unknown functions. Dockerin is involved in the degradation of the plant cell wall by incorporating glycosyl hydrolase into the extracellular cellulose complex "cellulosome" via interaction with the cohesion domain.[54] $Ca^{2+}$ induces the folding of dockerin.[64] Bacterial transaldolase, like its eukaryotic homologues, is involved in the metabolism of glucose.[65] No explanation has thus far been offered for the unique presence of EF-hand motifs in this bacterial enzyme. Autotransporter adhesin, a prototype of the adhesin family, mediates the specific attachment of bacteria to target cells.[66] The binding of $Ca^{2+}$ would probably provoke the efficient interaction and facilitate the attachment.

## Possible Roles of Single EF-Hand Proteins

The 358 predicted proteins containing single EF-hand motifs are in 162 complete or incomplete bacterial genomes in the PIR-NREF database (http://pir.georgetown.edu/cgi-bin/nfspecies.pl). They spread in the majority of bacterial species (Supplementary Table I). These proteins are implicated in a wide range of cellular processes such as drug resistance (multiple drug resistance protein, multidrug efflux transporter), ion and nutrilites transporting ($K^+$-transporting ATPase B, $Na^+$:solute symporter, hemin ABC transporter, cation efflux system protein), nucleic acid modification and metabolism (tRNA synthetase, ribonuclease G, exodeoxyribonuclease V gamma chain, RNA polymerase beta subunit, ATP-dependent DNA helicase, DNA polymerase tau subunit, DNA gyrase subunit A, DNA methyltransferase), transcriptional regulation (transcriptional regulator), stress response (DnaK, acid shock

proteins, heat shock protein HspG), chemotaxis (CheV, histidine kinase HAMP region), energy and nutrilites metabolism (GTP-binding protein, AMP nucleosidase, aminotransferase, acetyltransferase), redox reaction (flavodoxin oxidoreductase, thio-disulfide isomerase, iron-sulfur cluster binding protein, thioredoxin reductase), and cell wall modification and degradation (chitinase C, glycosyl hydrolase, exopolysaccharide synthesis protein, probably secreted sialidase, putative surface anchored protein).

Among all of these matches, ATP-binding cassette (ABC) transporter and Shr are of particular interest considering their important roles in bacterial activities and the possible implication of $Ca^{2+}$ in the biological context. ABC transporter couples the hydrolysis of ATP to the transport of various molecules including sugars, ions, antibiotics, and peptides across the cell membrane.[67–69] Shr in *streptococcus* encodes a large hydrophilic protein (putative $Fe^{3+}$-siderophore transport) that has no significant homologues in bacteria but shares partial homology with eukaryotic receptors such as Toll and G-protein dependent receptors. A leucine-rich repeat domain, an EF-hand domain, and two NEAT domains are identified in Shr. Shr directly binds heme-proteins such as hemoglobin, myoglobin, heme-BSA and the hemoglobin-heptoglobin complex.[70,71] The presence of a nearly perfect EF-hand domain in Shr raises the possibility that $Ca^{2+}$ may modulate its activity and represent a new type of $Ca^{2+}$ regulated receptor involved in heme-protein binding and iron acquisition.

The single-handed EF-hand motifs were also observed in *Arabidopsis*.[72] These observations raise the possibility that the ubiquitous EF-hand motif may function as an independent structural unit for $Ca^{2+}$ binding. To date, the majority of known EF-hand motifs are coupled through the hydrophobic interaction of the flanking helices.[18,19,73,74] Our work has shown that the isolated EF-loop III from CaM without the flanking helices in a host protein is able to bind $Ca^{2+}$ and remains monomer in solution.[75] The addition of flanking helices results in the dimer formation (unpublished results). Peptide fragments encompassing the EF-hand motifs were also shown to be dimers in solution.[76–79] We hope that our prediction will spur the exploration of the relationship between their function and $Ca^{2+}$ binding capability. The predicted proteins containing single EF-hand can be found in the supplementary Table I.

## Evolutionary Perspectives on EF-Hand Proteins

The classification and evolution of EF-hand proteins was first analyzed by Kretsinger and coworkers.[11,80–82] A dendrogram of the EF-hand proteins has been published previously. Since then, more EF-hand subfamilies, especially pseudo-EF-hand proteins, have been identified. To analyze the potential evolutionary scenario of the EF-hands, particularly pseudo EF-hands, phylogenetic analysis was carried out with updated pseudo EF-hand members and part of the canonical EF-hand proteins in this study on the basis of sequence alignments (Fig. 5).

The pseudo EF-hand N-J tree revealed three major groups assuming that they are evolved from a common ancestral protein. The largest group consists of two closely-related subgroups, one with S100A2, S100A3, S100A4, S100A5, and S100A6 and the other with S100A1, S100P, S100B, S100Z, and S100A10. It is interesting to note that S100A10, separating early from other members in its subgroup, loses the capacity to chelate $Ca^{2+}$ ion with mutations and deletions at the $Ca^{2+}$ liganding positions in both canonical and pseudo-EF-hand motifs. The small phylogenetic distance between S100A2, S100A3, S100A4, S100A5, and S100A6 is consistent with the clustered organization of these genes.[83] Additionally, S100A2, S100A3, S100A5, and S100A6 have been proposed to coordinate $Zn^{2+}$ with varying affinity.[84–87] The second major group is comprised of S100A8, S100A9, S100A12, trichohyalin, and MRP-126 from chicken. All of these proteins (except for trichohyalin) are excreted to the extracellular space, where $Ca^{2+}$ concentration is at the millimolar level.[88] Their common targets are cytoskeletal or cell membrane proteins. In addition, the proteins in this group are associated with pro-inflammatory functions by inducing chemotaxis or secretion of pro-inflammatory mediators. Interestingly, members in this group possess the $Zn^{2+}$-chelating motif His-x-x-x-His at the C-termini, with possible involvement of an upstream glutamate.[89] The third major group consists of S100A7, S100A11, S100A11P, S100A15, S100H, and repetin. Repetin contains an N-terminal S100-like domain and central tandem repeats of glutamine-rich sequence.[37] It is involved in epidermal differentiation. Repetin is separated early from other members in the group. The other members (S100A13 and S100A14; S100A16 and S100A17; and S100G and hornerin) form three minor groups. They are rather diverse and no valuable evolutionary clues can be inferred at present. Repetin, trichohyalin, and hornerin belong to the "fused gene" family. A proposed evolutionary pathway for hornerin involves the fusion of an S100-like $Ca^{2+}$-binding protein with an ancestral epidermal structural gene containing tandem repeats that reside in the same chromosomal locus 1q21.[37,90]

The canonical EF-hands are ubiquitously distributed across the eukaryota, bacteria, and archaea. The gene replication could also be tentatively used to explain the appearance of penta- (calpain subfamily) and hexa-EF-hands (calretinin and calbindin D28k), both of which have distinct phylogenetic pathways (Fig. 5).[91,92] The abundant single-handed EF-hand-like motifs found in the genomes of bacteria could provide clues for the origin of the prototypical EF-hand.[22,24,93,94] The evolutionary mobile single $Ca^{2+}$ binding loop first present in the ancestral protein could be "transplanted" to several locations of the host protein or to several host proteins.[19,95]

In contrast to the canonical EF-hands, our study shows that the pseudo EF-hands are exclusively found in vertebrates with tissue- and cell-specific expression profiles and are not predicted in the bacteria genomes. The lowest organism containing pseudo EF-hand reported thus far is the spiny dogfish (*Squalus acanthias*) with a pseudo EF-hand protein that is closely related to S100A1.[13] Thus, it

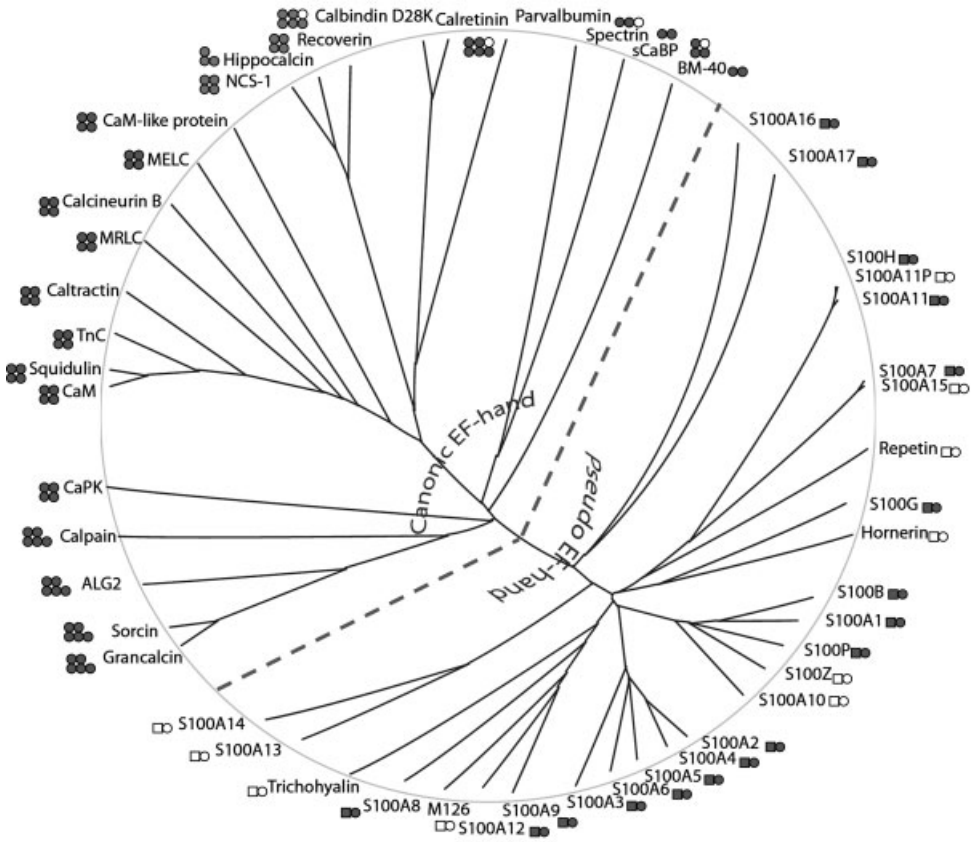



Fig. 5. Phylogenetic analysis of the EF-hand protein family. The unrooted N-J tree was generated on the basis of multiple sequence alignments of 27 typical proteins containing pseudo EF-hand motifs and 22 proteins with canonical EF hand motifs. (Circle: canonical EF-hand; Square: pseudo EF-hand; Solid: bind $Ca^{2+}$; Open: do not bind $Ca^{2+}$ or $Ca^{2+}$ binding capability is unknown).

is reasonable to postulate that pseudo EF-hands are phylogenetically younger and have a shorter history than canonical EF-hands. Although more evidence is required to confirm the postulation, the current observation in natural proteins that pseudo EF-hands always pair with canonical EF-hands but canonical EF-hands do not necessarily pair with pseudo EF-hands also indicate that pseudo EF-hands appear later than the canonical EF-hands. Genomic study on human and rat S100 proteins has also indicated the recent origin of the S100 subfamily.[13,83] It has been hypothesized that the evolution of pseudo EF-hands might be achieved by domain swapping through gene duplication or exon recombination from a CaM-type protein with subsequent loss of two of the four EF-hands.[13] Then, evolutionary divergence of EF-hands follows, thereby creating the sequence diversity of pseudo EF-hands. During this process, pseudo EF-hands become distant relatives of canonical EF-hand and a number of pseudo EF-hands (S100B, S100A2, S100A3, S100A5, S100A6, S100A7, S100A12) acquire the ability to bind other metal ions such as $Zn^{2+}$ or $Cu^{2+}$ to further adapt to the tissue-specific temporal-spatial requirement.[43] They evolved largely varied $Ca^{2+}$ affinity from nM to mM to meet the versatile requirements at various cellular compartments.[84,85,87,96,97]

## CONCLUSIONS

With an increasing number of EF-hand $Ca^{2+}$ binding proteins being discovered and characterized in bacteria, archaea, and eukaryotes, structural and functional knowledge of the EF-hand proteins has expanded steadily in recent years. The EF-hand-like proteins contain $Ca^{2+}$-binding sequences that closely resemble the canonical EF-hand motif yet with diversified flanking structural elements. An easy and straightforward searching method to identify both canonical and pseudo EF-hands has been established based on our modified patterns. In addition to being supplementary to the signatures PS00018 and PS00303, the newly developed patterns convey information on the flanking structural contents with higher accuracy and sensitivity. Screening of the prokaryotic genome information revealed 397 entries of putative EF-hand proteins (467 motifs) implicated in a variety of cellular activities. The results enable us to envision the possible scenarios of the evolutionary history of EF-hands. The pseudo EF-hands are likely to be phylogenetically younger than canonical EF-hand motifs. The prediction of $Ca^{2+}$ binding motifs in bacteria genomes is helpful for the exploration of the role of $Ca^{2+}$ and $Ca^{2+}$ binding proteins in bacteria. Moving toward our goal of advancing calciomics, we have also been carrying out the prediction studies on other

genomes, such as plants and viruses, with our prediction method. This will further enable us to better understand the role of $Ca^{2+}$ in diverse biological systems.

## ACKNOWLEDGMENTS

The authors appreciate the critical review by Dan Adams and other members in JJY's group.

## REFERENCES

1. Carafoli E. Calcium signaling: a tale for all seasons. Proc Natl Acad Sci USA 2002;99:1115–1122.
2. Ermak G, Davies KJ. Calcium and oxidative stress: from cell signaling to cell death. Mol Immunol 2002;38:713–721.
3. Orrenius S, Zhivotovsky B, Nicotera P. Regulation of cell death: the calcium-apoptosis link. Nat Rev Mol Cell Biol 2003;4:552–565.
4. Dowd DR. Calcium regulation of apoptosis. Adv Second Messenger Phosphoprotein Res 1995;30:255–280.
5. Means AR, Rasmussen CD. Calcium, calmodulin and cell proliferation. Cell Calcium 1988;9:313–319.
6. Tsien RW, Tsien RY. Calcium channels, stores, and oscillations. Annu Rev Cell Biol 1990;6:715–760.
7. Levine BA, Dalgarno DC, Esnouf MP, Klevit RE, Scott GM, Williams RJ. The mobility of calcium-trigger proteins and its function. Ciba Found Symp 1983;93:72–97.
8. Schroder B, Schlumbohm C, Kaune R, Breves G. Role of calbindin-D9k in buffering cytosolic free $Ca^{2+}$ ions in pig duodenal enterocytes. J Physiol 1996;492(Part 3):715–722.
9. Buchanan JD, Corbett RJ, Roche RS. The thermodynamics of calcium binding to thermolysin. Biophys Chem 1986;23:183–199.
10. Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, Hood L. Gene families: the taxonomy of protein paralogs and chimeras. Science 1997;278:609–614.
11. Kawasaki H, Nakayama S, Kretsinger RH. Classification and evolution of EF-hand proteins. Biometals 1998;11:277–295.
12. Kretsinger RH, Nockolds CE. Carp muscle calcium-binding protein. II. Structure determination and general description. J Biol Chem 1973;248:3313–3326.
13. Ravasi T, Hsu K, Goyette J, Schroder K, Yang Z, Rahimi F, Miranda LP, Alewood PF, Hume DA, Geczy C. Probing the S100 protein family through genomic and functional analysis. Genomics 2004;84:10–22.
14. Schuster-Bockler B, Schultz J, Rahmann S. HMM Logos for visualization of protein families. BMC Bioinformatics 2004;5:7.
15. Pal GP, Elce JS, Jia Z. Dissociation and aggregation of calpain in the presence of calcium. J Biol Chem 2001;276:47233–47238.
16. Raser KJ, Buroker-Kilgore M, Wang KK. Binding and aggregation of human mu-calpain by terbium ion. Biochim Biophys Acta 1996;1292:9–14.
17. Ravulapalli R, Diaz BG, Campbell RL, Davies PL. Homodimerization of calpain 3 penta-EF-hand domain. Biochem J 2005;388(Part 2): 585–591.
18. Hunter MJ, Chazin WJ. High level expression and dimer characterization of the S100 EF-hand proteins, migration inhibitory factor-related proteins 8 and 14. J Biol Chem 1998;273:12427–12435.
19. Yap KL, Ames JB, Swindells MB, Ikura M. Diversity of conformational states and changes within the EF-hand protein superfamily. Proteins 1999;37:499–507.
20. Rigden DJ, Jedrzejas MJ, Galperin MY. An extracellular calcium-binding domain in bacteria with a distant relationship to EF-hands. FEMS Microbiol Lett 2003;221:103–110.
21. Rigden DJ, Jedrzejas MJ, Moroz OV, Galperin MY. Structural diversity of calcium-binding proteins in bacteria: single-handed EF-hands? Trends Microbiol 2003;11:295–297.
22. Vyas NK, Vyas MN, Quiocho FA. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. Nature 1987;327:635–638.
23. Michiels J, Xi C, Verhaert J, Vanderleyden J. The functions of $Ca^{2+}$ in bacteria: a role for EF-hand proteins? Trends Microbiol 2002;10:87–93.
24. Rigden DJ, Galperin MY. The DxDxDG motif for calcium binding: multiple structural contexts and implications for evolution. J Mol Biol 2004;343:971–984.
25. Yang K. Prokaryotic calmodulins: recent developments and evolutionary implications. J Mol Microbiol Biotechnol 2001;3:457–459.
26. Momma K, Mikami B, Mishima Y, Hashimoto W, Murata K. Crystal structure of AlgQ2, a macromolecule (alginate)-binding protein of Sphingomonas sp. A1 at 2.0A resolution. J Mol Biol 2002; 316:1051–1059.
27. Petosa C, Collier RJ, Klimpel KR, Leppla SH, Liddington RC. Crystal structure of the anthrax toxin protective antigen. Nature 1997;385:833–838.
28. Lytle BL, Volkman BF, Westler WM, Heckman MP, Wu JH. Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. J Mol Biol 2001;307: 745–753.
29. Ye Y, Shealy S, Lee HW, Torshin I, Harrison R, Yang JJ. A grafting approach to obtain site-specific metal-binding properties of EF-hand proteins. Protein Eng 2003;16:429–434.
30. Ye Y, Lee HW, Yang W, Shealy S, Yang JJ. Probing site-specific calmodulin calcium and lanthanide affinity by grafting. J Am Chem Soc 2005;127:3743–3750.
31. Ye Y, Lee HW, Yang W, Yang JJ. Calcium and lanthanide affinity of the EF-loops from the C-terminal domain of calmodulin. J Inorg Biochem 2005;99:1376–1383.
32. Rashidi HH, Bauer M, Patterson J, Smith DW. Sequence motifs determine structure and $Ca^{++}$-binding by EF-hand proteins. J Mol Microbiol Biotechnol 1999;1:175–182.
33. Dragani B, Aceto A. About the role of conserved amino acid residues in the calcium-binding site of proteins. Arch Biochem Biophys 1999;368:211–213.
34. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994;22:4673–4680.
35. Page RD. TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 1996;12:357, 358.
36. Eddy SR. Profile hidden markov models. Bioinformatics 1998;14: 755–763.
37. Krieg P, Schuppler M, Koesters R, Mincheva A, Lichter P, Marks F. Repetin (Rptn), a new member of the "fused gene" subgroup within the S100 gene family encoding a murine epidermal differentiation protein. Genomics 1997;43:339–348.
38. Makino T, Takaishi M, Morohashi M, Huh NH. Hornerin, a novel profilaggrin-like protein and differentiation-specific marker isolated from mouse skin. J Biol Chem 2001;276:47445–47452.
39. Huber M, Siegenthaler G, Mirancea N, Marenholz I, Nizetic D, Breitkreutz D, Mischke D, Hohl D. Isolation and characterization of human repetin, a member of the fused gene family of the epidermal differentiation complex. J Invest Dermatol 2005;124:998–1007.
40. Lee SC, Kim IG, Marekov LN, O'Keefe EJ, Parry DA, Steinert PM. 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 1993;268:12164–12176.
41. Donato R. Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. Biochim Biophys Acta 1999;1450: 191–231.
42. Donato R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. Int J Biochem Cell Biol 2001;33:637–668.
43. Heizmann CW, Cox JA. New perspectives on S100 proteins: a multi-functional $Ca^{2+}$-, $Zn^{2+}$- and $Cu^{2+}$-binding protein family. Biometals 1998;11:383–397.
44. Tanaka T, Miwa N, Kawamura S, Sohma H, Nitta K, Matsushima N. Molecular modeling of single polypeptide chain of calcium-binding protein p26olf from dimeric S100B($\beta\beta$). Protein Eng 1999; 12:395–405.
45. Miwa N, Kobayashi M, Takamatsu K, Kawamura S. Purification and molecular cloning of a novel calcium-binding protein, p26olf, in the frog olfactory epithelium. Biochem Biophys Res Commun 1998;251:860–867.
46. Leung AK, Duewel HS, Honek JF, Berghuis AM. Crystal structure of the lytic transglycosylase from bacteriophage lambda in complex with hexa-*N*-acetylchitohexaose. Biochemistry 2001;40: 5665–5673.
47. van Asselt EJ, Dijkstra BW. Binding of calcium in the EF-hand of *Escherichia coli* lytic transglycosylase Slt35 is important for stability. FEBS Lett 1999;458:429–435.

48. Wu CH, Yeh LS, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Kourtesis P, Ledley RS, Suzek BE, Vinayaka CR, Zhang J, Barker WC. The protein information resource. Nucleic Acids Res 2003;31:345–347.
49. Aitio H, Annila A, Heikkinen S, Thulin E, Drakenberg T, Kilpelainen I. NMR assignments, secondary structure, and global fold of calerythrin, an EF-hand calcium-binding protein from *Saccharopolyspora erythraea*. Protein Sci 1999;8:2580–2588.
50. Gangola P, Rosen BP. Maintenance of intracellular calcium in *Escherichia coli*. J Biol Chem 1987;262:12570–12574.
51. Herbaud ML, Guiseppi A, Denizot F, Haiech J, Kilhoffer MC. Calcium signalling in *Bacillus subtilis*. Biochim Biophys Acta 1998;1448:212–226.
52. Kalckar HM. The periplasmic galactose receptor protein of *Escherichia coli* in relation to galactose chemotaxis. Biochimie 1976;58: 81–85.
53. Tisa LS, Adler J. Calcium ions are involved in *Escherichia coli* chemotaxis. Proc Natl Acad Sci USA 1992;89:11804–11808.
54. Carvalho AL, Dias FM, Prates JA, Nagy T, Gilbert HJ, Davies GJ, Ferreira LM, Romao MJ, Fontes CM. Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. Proc Natl Acad Sci USA 2003;100:13809–13814.
55. Lin J, Ficht TA. Protein synthesis in *Brucella abortus* induced during macrophage infection. Infect Immun 1995;63:1409–1414.
56. Teixeira-Gomes AP, Cloeckaert A, Zygmunt MS. Characterization of heat, oxidative, and acid stress responses in *Brucella melitensis*. Infect Immun 2000;68:2954–2961.
57. Leadly PF, Roberts G, Walker JE. Isolation of a novel calcium-binding protein from *Streptomyces erythreus*. FEBS Lett 1984;178: 157–160.
58. Natsume M, Yasui K, Marumo S. Calcium ion regulates aerial mycelium formation in actinomycetes. J Antibiot (Tokyo) 1989;42: 440–447.
59. Tossavainen H, Permi P, Annila A, Kilpelainen I, Drakenberg T. NMR solution structure of calerythrin, an EF-hand calcium-binding protein from *Saccharopolyspora erythraea*. Eur J Biochem 2003;270:2505–2512.
60. Yonekawa T, Ohnishi Y, Horinouchi S. A calmodulin-like protein in the bacterial genus *Streptomyces*. FEMS Microbiol Lett 2005; 244:315–321.
61. Xi C, Schoeters E, Vanderleyden J, Michiels J. Symbiosis-specific expression of *Rhizobium etli* casA encoding a secreted calmodulin-related protein. Proc Natl Acad Sci USA 2000;97:11114–11119.
62. Spurway TD, Morland C, Cooper A, Sumner I, Hazlewood GP, O'Donnell AG, Pickersgill RW, Gilbert HJ. Calcium protects a mesophilic xylanase from proteinase inactivation and thermal unfolding. J Biol Chem. 1997;272:17523–17530.
63. Chen L, Liu MY, Le Gall J. Calcium is required for the reduction of sulfite from hydrogen in a reconstituted electron transfer chain from the sulfate reducing bacterium, *Desulfovibrio gigas*. Biochem Biophys Res Commun 1991;180:238–242.
64. Lytle BL, Volkman BF, Westler WM, Wu JH. Secondary structure and calcium-induced folding of the *Clostridium thermocellum* dockerin domain determined by NMR spectroscopy. Arch Biochem Biophys 2000;379:237–244.
65. Kohler U, Cerff R, Brinkmann H. Transaldolase genes from the cyanobacteria *Anabaena variabilis* and *Synechocystis sp*. PCC 6803: comparison with other eubacterial and eukaryotic homologues. Plant Mol Biol 1996;30:213–218.
66. Laarmann S, Schmidt MA The *Escherichia coli* AIDA autotransporter adhesin recognizes an integral membrane glycoprotein as receptor. Microbiology 2003;149(Part 7):1871–1882.
67. Dean M, Allikmets R. Evolution of ATP-binding cassette transporter genes. Curr Opin Genet Dev 1995;5:779–785.
68. Higgins CF. ABC transporters: from microorganisms to man. Annu Rev Cell Biol 1992;8:67–113.
69. van Veen HW, Konings WN. Multidrug transporters from bacteria to man: similarities in structure and function. Semin Cancer Biol 1997;8:183–191.
70. Andrade MA, Ciccarelli FD, Perez-Iratxeta C, Bork P. NEAT: a domain duplicated in genes near the components of a putative $Fe^{3+}$ siderophore transporter from gram-positive pathogenic bacteria. Genome Biol 2002;3:research0047.1.
71. Bates CS, Montanez GE, Woods CR, Vincent RM, Eichenbaum Z. Identification and characterization of a *Streptococcus pyogenes* operon involved in binding of hemoproteins and acquisition of iron. Infect Immun 2003;71:1042–1055.
72. Day IS, Reddy VS, Shad Ali G, Reddy AS. Analysis of EF-hand-containing proteins in Arabidopsis. Genome Biol 2002;3: research0056.1.
73. Julenius K, Robblee J, Thulin E, Finn BE, Fairman R, Linse S. Coupling of ligand binding and dimerization of helix-loop-helix peptides: spectroscopic and sedimentation analyses of calbindin D9k EF-hands. Proteins 2002;47:323–333.
74. Kawasaki H, Kretsinger RH. Calcium-binding proteins 1: EF-hands. Protein Profile 1995;2:297–490.
75. Lee HW, Yang W, Ye Y, Liu ZR, Glushka J, Yang JJ. Isolated EF-loop III of calmodulin in a scaffold protein remains unpaired in solution using pulsed-field-gradient NMR spectroscopy. Biochim Biophys Acta 2002;1598:80–87.
76. Franchini PL, Reid RE. A model for circular dichroism monitored dimerization and calcium binding in an EF-hand synthetic peptide. J Theor Biol 1999;199:199–211.
77. Berggard T, Julenius K, Ogard A, Drakenberg T, Linse S. Fragment complementation studies of protein stabilization by hydrophobic core residues. Biochemistry 2001;40:1257–1264.
78. Wojcik J, Goral J, Pawlowski K, Bierzynski A. Isolated calcium-binding loops of EF-hand proteins can dimerize to form a native-like structure. Biochemistry 1997;36:680–687.
79. Berggard T, Thulin E, Akerfeldt KS, Linse S. Fragment complementation of calbindin D28k. Protein Sci 2000;9:2094–2108.
80. Kretsinger RH. Evolution and function of calcium-binding proteins. Int Rev Cytol 1976;46:323–393.
81. Kretsinger RH, Moncrief ND. Evolution of calcium modulated proteins. Va Explor 1989;5:7–9.
82. Nakayama S, Kretsinger RH. Evolution of the EF-hand family of proteins. Annu Rev Biophys Biomol Struct 1994;23:473–507.
83. Ridinger K, Ilg EC, Niggli FK, Heizmann CW, Schafer BW. Clustered organization of S100 genes in human and mouse. Biochim Biophys Acta 1998;1448:254–263.
84. Franz C, Durussel I, Cox JA, Schafer BW, Heizmann CW. Binding of $Ca^{2+}$ and $Zn^{2+}$ to human nuclear S100A2 and mutant proteins. J Biol Chem 1998;273:18826–18834.
85. Fohr UG, Heizmann CW, Engelkamp D, Schafer BW, Cox JA. Purification and cation binding properties of the recombinant human S100 calcium-binding protein A3, an EF-hand motif protein with high affinity for zinc. J Biol Chem 1995;270:21056–21061.
86. Schafer BW, Fritschy JM, Murmann P, Troxler H, Durussel I, Heizmann CW, Cox JA. Brain S100A5 is a novel calcium-, zinc-, and copper ion-binding protein of the EF-hand superfamily. J Biol Chem 2000;275:30623–30630.
87. Filipek A, Heizmann CW, Kuznicki J. Calcyclin is a calcium and zinc binding protein. FEBS Lett 1990;264:263–266.
88. Heizmann CW. The multifunctional S100 protein family. Methods Mol Biol 2002;172:69–80.
89. Clohessy PA, Golden BE. His-X-X-X-His motif in S100 protein, calprotectin: relation to microbiostatic activity. J Leukoc Biol 1996; 60:674.
90. Contzler R, Favre B, Huber M, Hohl D. Cornulin, a new member of the "fused gene" family, is expressed during epidermal differentiation. J Invest Dermatol 2005;124:990–997.
91. Maki M, Narayana SV, Hitomi K. A growing family of the $Ca^{2+}$-binding proteins with five EF-hand motifs. Biochem J 1997; 328(Part 2):718–720.
92. Palczewska M, Groves P, Batta G, Heise B, Kuznicki J. Calretinin and calbindin D28k have different domain organizations. Protein Sci 2003;12:180–184.
93. Swan DG, Hale RS, Dhillon N, Leadlay PF. A bacterial calcium-binding protein homologous to calmodulin. Nature 1987;329:84–85.
94. Yonekawa T, Ohnishi Y, Horinouchi S. A calcium-binding protein with four EF-hand motifs in *Streptomyces ambofaciens*. Biosci Biotechnol Biochem 2001;65:156–160.
95. Grishin NV. Fold change in evolution of protein structures. J Struct Biol 2001;134:167–185.
96. Fritz G, Heizmann CW, Kroneck PM. Probing the structure of the human $Ca^{2+}$- and $Zn^{2+}$-binding protein S100A3: spectroscopic investigations of its transition metal ion complexes, and three-dimensional structural model. Biochim Biophys Acta 1998; 1448:264–276.
97. Fritz G, Mittl PR, Vasak M, Grutter MG, Heizmann CW. The crystal structure of metal-free human EF-hand protein S100A3 at 1.7-A resolution. J Biol Chem 2002;277:33092–33098.