An extracellular calcium-binding domain in bacteria with a distant relationship to EF-hands

Daniel J. Rigden a, Mark J. Jedrzejas b, Michael Y. Galperin c,*

a National Center of Genetic Resources and Biotechnology, Cenargen/Embrapa, Brasilia D.F. 70770-900, Brazil
b Children’s Hospital Oakland Research Institute, Oakland, CA 94609, USA
c National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

Received 22 January 2003; received in revised form 19 February 2003; accepted 20 February 2003
First published online 12 March 2003

Abstract

Extracellular Ca^{2+}-dependent nuclease YokF from Bacillus subtilis and several other surface-exposed proteins from diverse bacteria are encoded in the genomes in two paralogous forms that differ by a ~45 amino acid fragment, which comprises a novel conserved domain. Sequence analysis of this domain revealed a conserved DxDxDGxxCE motif, which is strikingly similar to the Ca^{2+}-binding loop of the calmodulin-like EF-hand domains, suggesting an evolutionary relationship between them. Functions of many of the other proteins in which the novel domain, named Excalibur (extracellular calcium-binding region), is found, as well as a structural model of its conserved motif are consistent with the notion that the Excalibur domain binds calcium. This domain is but one more example of the diversity of structural contexts surrounding the EF-hand-like calcium-binding loop in bacteria. This loop is thus more widespread than hitherto recognized and the evolution of EF-hand-like domains is probably more complex than previously appreciated.

Keywords: Genome analysis; Calcium-binding site; Protein domain; Cell envelope; S-layer; Nuclease; Calmodulin; Evolution

1. Introduction

Many enzymes depend on metal cofactors for their activity. Several metal ions, such as copper and iron, feature dedicated binding domains that participate in their storage, transport and delivery to the target enzymes. Although the diversity of the metal-binding domains in prokaryotes appears to be much less than in eukaryotes, there are some well-characterized ones, such as the heavy metal-associated domain found in metal-transporting P-type ATPases, copper chaperones, and periplasmic mercury-binding proteins (reviewed in [1–3]) and the iron-binding ferritin domain (reviewed in [4]).

Calcium, which regulates a variety of regulatory processes in bacteria, such as motility, chemotaxis, cell division and differentiation [5,6], also has dedicated binding domains. The best studied of these are the parallel β-roll domain, characterized by a repeated GGxGxD sequence motif, found in a number of secreted proteins from Gram-negative bacteria [7] and the calmodulin-like EF-hand helix-loop-helix domain, consisting of two helices, E and F, connected by a calcium-binding loop with a DxDxDG consensus motif (reviewed in [8–10]). The EF-hand structure was found in a variety of eukaryotic Ca^{2+}-binding proteins, such as parvalbumin, calmodulin, aequorin, calbindin and S100 proteins [8–10]. In contrast, this structural motif has not been found so far in any bacterial protein with known three-dimensional structure. Nevertheless, bacterial Ca^{2+}-binding proteins with predicted EF-hands have been described, for example, calerythrin from the actinomycete Saccharopolyspora erythrea [11] and calysmin from Rhizobium etli [12]. The presence of the EF-hand helix-loop-helix structure in the former protein has been recently confirmed by NMR analysis [13]. Sequence comparisons revealed the presence of EF-hand-like motifs in a variety of proteins encoded in completely sequenced bacterial genomes [14,15]. In addition, Ca^{2+}-binding sites in several bacterial extracytoplasmic proteins were found to resemble the EF-hands. These include, for example, the soluble lytic transglycosylase from Escherichia coli [16], the periplasmic galactose-binding protein from Salmonella.

* Corresponding author. Tel.: +1 (301) 435 5910;
Fax: +1 (301) 435 7794;
E-mail address: galperin@ncbi.nlm.nih.gov (M.Y. Galperin).
typhimurium [17], Bacillus anthracis protective antigen [18] and the dockerin domain of the extracellular cellulase complex from Clostridium thermocellum [19]. While all these proteins deviate in one way or another from the typical EF-hands, they all share the conserved Ca\(^{2+}\)-binding motif Dx(D/N)xDG within the central loop of the EF-hand domain. The variety of Ca\(^{2+}\)-binding proteins in bacteria suggests the possibility of additional Ca\(^{2+}\)-binding structures. Here we describe one more bacterial extracellular domain that shares local but striking similarity with the EF-hand motif and is found in several Ca\(^{2+}\)-dependent surface proteins from both Gram-positive and Gram-negative bacteria.

2. Materials and methods

2.1. Database searches

Initial sequence similarity searches were performed against the non-redundant protein database and the database of finished and unfinished microbial genome sequences at the National Center for Biotechnology Information (NCBI, Bethesda, MD, USA) using PSI-BLAST [20] program run with default parameters. Database hits identified in these searches were used as queries for additional protein database PSI-BLAST searches with the relaxed inclusion threshold values of \(E = 0.018\) and for the searches against translated GenBank nucleotide database using the TBLASTn [20] program. The completeness of the search results was verified by direct pattern matching using the DxDxDGxxCE motif (Fig. 1B). The remarkable similarity of this motif to that defined for the Ca\(^{2+}\)-binding motif of the EF-hand calcium-binding domain [8–10] indicated that the novel domain should also bind Ca\(^{2+}\) and prompted us to name it accordingly as the Excalibur domain. Indeed, a comparison of Excalibur and EF-hand consensus sequences demonstrates conservation in the former of the three Ca\(^{2+}\)-binding Asp residues of the EF-hands and compatibility of the surrounding residues (shown in bold).

<table>
<thead>
<tr>
<th>Excalibur numbering</th>
<th>1 3 5 7 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% Excalibur consensus</td>
<td>bDRDsDGxhA--CE</td>
</tr>
<tr>
<td>50% EF-hand consensus</td>
<td>bDpDsDGxxE</td>
</tr>
</tbody>
</table>

where small letters indicate the following groups of residues: h – hydrophobic, p – polar, s – small, and x – any residue. The only significant difference between the two consensus motifs is the shorter distance between the conserved aspartates and the Glu residue in the Excalibur domain. Since in some EF-hand-like sequences this Glu residue can be shifted or even located elsewhere on the polypeptide chain [17], this comparison supports the prediction that the Excalibur domain is capable of binding calcium.

2.2. Secondary structure analysis

Signal peptides were predicted using the SignalIP [23] program. Secondary structure predictions were performed using PSI-PRED [24], SSpro [25], and several other programs available on the web (reviewed in [26]). The resulting predictions were manually reconciled based on the reliability figures for each particular residue, reported by each program. Consensus sequences were derived from alignments using Mview [27]; EF-hand sequences were obtained from the SMART database [21].

3. Results and discussion

3.1. Sequence and structure of the Excalibur (extracellular calcium-binding region) domain

In the course of a survey of the surface-exposed proteins of Streptococcus pneumoniae, the putative cell-wall- binding protein SP0667 (SPTREMBL accession Q97RW9) was found to contain an 80-aa fragment that was located between the predicted signal peptide and the cell-wall-binding domains and showed no statistically significant sequence similarity to any domain listed in SMART [21] or Pfam [22] domain databases. A PSI-BLAST search of the NCBI protein database using this fragment of SP0667 (residues 24–103) as the query converged after five iterations of PSI-BLAST retrieving 14 proteins from both low-GC and high-GC Gram-positive bacteria that contained a well-defined domain with two absolutely conserved Cys residues and three absolutely conserved Asp residues (Fig. 1). All these proteins had predicted signal peptides and invariably placed the new domain at the extreme N- or C-termini of the mature secreted proteins (Fig. 2). Using these proteins as queries in additional PSI-BLAST searches with a relaxed inclusion cut-off \(E\)-value retrieved proteins from Gram-negative bacteria that had the same sequence pattern. Several more instances of this domain, including proteins from Mycobacterium tuberculosis and Streptomyces exfoliatus, were found encoded in the DNA sequences in GenBank nucleotide database and in the database of unfinished microbial genome sequences (Table 1) but not translated into proteins, presumably because of their short size.

All the predicted proteins retrieved by these database searches contained a similar 40–45-aa domain, present in a stand-alone form in the Pasteurella multocida protein PM0157 and characterized by the conserved DxDxDGxxCE motif (Fig. 1B). The remarkable similarity of this motif to that defined for the Ca\(^{2+}\)-binding motif of the EF-hand calcium-binding domain [8–10] indicated that the novel domain should also bind Ca\(^{2+}\) and prompted us to name it accordingly as the Excalibur domain. Indeed, a comparison of Excalibur and EF-hand consensus sequences demonstrates conservation in the former of the three Ca\(^{2+}\)-binding Asp residues of the EF-hands and compatibility of the surrounding residues (shown in bold).

3.2. Diversity of the Excalibur-containing proteins

The domain organization of deduced proteins contain-
The Excalibur domain (Table 1 and Fig. 2) indicates that it is often found in association with domains that are either Ca\(^{2+}\)-dependent or Ca\(^{2+}\)-regulated. The most common of such domain structures is a combination of a C-terminal Excalibur domain with an N-terminal endonuclease domain, typified by the staphylococcal nuclease [28]. This domain combination is found in Gram-positive bacteria (Bacillus subtilis), cyanobacteria (Nostoc sp., Prochlorococcus marinus, Trichodesmium erythraeum), and in proteobacteria (Azotobacter vinelandii, Mesorhizobium loti). Remarkably, genomes of some of these organisms also encode paralogous forms of extracellular nuclease that lack the Excalibur domain (Fig. 2). In B. subtilis, for example, the core regions of yokF and yncB gene products are 92% identical. The YokF protein, however, contains an extra C-terminal 83-aa fragment [29], which consists of the Excalibur domain and a flexible linker (Fig. 2). A comparison of these two proteins revealed that YokF is

Table 1
A list of the Excalibur-containing proteins

<table>
<thead>
<tr>
<th>Organism</th>
<th>Protein name</th>
<th>Accession no. or gi no.</th>
<th>Length</th>
<th>Excalibur location</th>
<th>Other domains</th>
<th>Paralog without Excalibur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gram-positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. subtilis</td>
<td>YokF</td>
<td>AAC12979</td>
<td>296</td>
<td>C-terminal</td>
<td>SNase</td>
<td>YncB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>21400020</td>
<td>333</td>
<td>C-terminal</td>
<td>Coiled coil</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2182845</td>
<td>268</td>
<td>C-terminal</td>
<td>SLH</td>
<td>BA_1675</td>
</tr>
<tr>
<td>S. pneumoniae</td>
<td></td>
<td>AAK74812</td>
<td>332</td>
<td>N-terminal</td>
<td>Cell wall</td>
<td></td>
</tr>
<tr>
<td>Streptococcus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mitis</td>
<td></td>
<td>Unfinished</td>
<td>190</td>
<td>N-terminal</td>
<td>Cell wall</td>
<td></td>
</tr>
<tr>
<td>Staphylococcus</td>
<td></td>
<td>BAB57961</td>
<td>204</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>aureus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staphylococcus</td>
<td></td>
<td>AF548637 &gt;184</td>
<td></td>
<td>C-terminal</td>
<td>SNase</td>
<td></td>
</tr>
<tr>
<td>capsicum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Actinobacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corynebacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glutamicum</td>
<td>Cgl0341</td>
<td>BAB97734</td>
<td>191</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>Corynebacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glutamicum</td>
<td>Cgl0956</td>
<td>BAB98349</td>
<td>173</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>Corynebacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficiens</td>
<td>CE1921</td>
<td>BAC18731</td>
<td>211</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>Corynebacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>efficiens</td>
<td>CE2653</td>
<td>BAC19463</td>
<td>363</td>
<td>C-terminal</td>
<td>Rs2972c</td>
<td>CE1360</td>
</tr>
<tr>
<td>Streptomyces</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coelicolor</td>
<td>C7A8.31</td>
<td>CAB69780</td>
<td>157</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>S. exfoliatus</td>
<td></td>
<td>U97057</td>
<td>100</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>Thermophilicula</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fusca</td>
<td></td>
<td>Unfinished</td>
<td>142</td>
<td>C-terminal</td>
<td>Low comp.</td>
<td></td>
</tr>
<tr>
<td>Mycobacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>avium</td>
<td></td>
<td>Unfinished</td>
<td>69</td>
<td>Whole protein</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Mycobacterium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smegmatiscens</td>
<td></td>
<td>Unfinished</td>
<td>82</td>
<td>Whole protein</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>M. tuberculosis</td>
<td></td>
<td>NC_000962</td>
<td>71</td>
<td>Whole protein</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Cyanobacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nostoc sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCC7120alpha</td>
<td>Alr7333</td>
<td>BAB77091</td>
<td>175</td>
<td>C-terminal</td>
<td>SNase</td>
<td>All0265, All2317</td>
</tr>
<tr>
<td>Nostoc sp.</td>
<td>pCC7120alpha</td>
<td>Alr73814</td>
<td>128</td>
<td>C-terminal</td>
<td>SNase</td>
<td></td>
</tr>
<tr>
<td>Nostoc sp.</td>
<td>pCC7120alpha</td>
<td>Asr73434</td>
<td>69</td>
<td>Whole protein</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Prochlorococcus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>marinus</td>
<td>Pmnt0770</td>
<td>23131487</td>
<td>211</td>
<td>C-terminal</td>
<td>SNase</td>
<td></td>
</tr>
<tr>
<td>Trichodesmium</td>
<td></td>
<td>Tery3071</td>
<td>233</td>
<td>C-terminal</td>
<td>SNase</td>
<td></td>
</tr>
<tr>
<td>erythraeum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chloroflexi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chlorolexis</td>
<td>Chlo2608</td>
<td>22972742</td>
<td>448</td>
<td>C-terminal</td>
<td>ComEC</td>
<td></td>
</tr>
<tr>
<td>Proteobacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha subdivision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azotobacter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vinelandii</td>
<td>Avin3995</td>
<td>23105817</td>
<td>225</td>
<td>C-terminal</td>
<td>SNase</td>
<td>Avin0542</td>
</tr>
<tr>
<td>Mesorhizobium</td>
<td>teleti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>loti</td>
<td>Ms140</td>
<td>CAD31545</td>
<td>257</td>
<td>C-terminal</td>
<td>SNase</td>
<td>Mir0479</td>
</tr>
<tr>
<td>Novosphingobium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aromaticivorans</td>
<td>NA</td>
<td>Unfinished</td>
<td>129</td>
<td>C-terminal</td>
<td>–</td>
<td>MII8119</td>
</tr>
<tr>
<td>Rhizobium sp.</td>
<td>pNGR234alpha</td>
<td>NA</td>
<td>140</td>
<td>–</td>
<td>C-terminal</td>
<td></td>
</tr>
<tr>
<td>Rhodobacter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sphaeroides</td>
<td>Rsph0055</td>
<td>22956401</td>
<td>292</td>
<td>17</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Gamma subdivision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. multocida</td>
<td>PM0157</td>
<td>AAK02241</td>
<td>69</td>
<td>Whole protein</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Shevallenia</td>
<td>SO0733</td>
<td>AAN53811</td>
<td>203</td>
<td>–</td>
<td>Cold shock</td>
<td></td>
</tr>
<tr>
<td>oneidensis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vibrio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ruhfasics</td>
<td>VVI2552</td>
<td>AAO10888</td>
<td>155</td>
<td>–</td>
<td>Cold shock</td>
<td></td>
</tr>
</tbody>
</table>

aProtein names and accession numbers in the NCBI protein database; NA, not available (previously untranslated proteins or proteins from unfinished genomes). Bold accession numbers are from the GenBank nucleotide database, newly translated proteins have the following locations: AF548637: 1..555 Frame+1; U97057: 1482..1784 Frame−1; NC_000962: 2580036..2580251 Frame−1; AE000066: 1792..2214 Frame−1.

bDomain names and abbreviations are as in Fig. 2. Low comp. stands for low-complexity segments, a dash indicates absence of recognizable domains. See text for details.

cThe N-terminal fragment of this protein is missing.

dThis protein is a C-terminal fragment of an Alr7333-like nuclease, disrupted by a frameshift mutation.
responsible for the major part of the endonuclease activity in B. subtilis cells, while YncB represented a minor fraction of that activity [29]. Whether these differences are fully attributable to the presence of the Excalibur domain remains to be determined.

The B. anthracis gene BA_5470 encodes another notable domain combination, the Excalibur domain preceded by three N-terminal S-layer homology (SLH) domains (Fig. 2). SLH domains are capable of tight non-covalent binding to the peptidoglycan and often serve as adaptors for cell-wall anchoring of various enzymes [30]. Studies of peptidoglycan attachment by SLH domains suggested an involvement of metal ions, possibly Ca$^{2+}$, in this interaction [31]. Ca$^{2+}$ has been previously shown to be involved in S-layer assembly in Caulobacter crescentus [32], although this might be an unrelated process, as this organism's genome does not seem to encode SLH domains. Similarly to the case of B. subtilis endonucleases above, B. anthracis also encodes a close paralog of BA_5470 that is lacking the Excalibur domain (Fig. 3).

Fig. 1. Multiple alignment of the Excalibur domain. A: The alignment was constructed on the basis of PSI-BLAST [20] search results with minimal manual editing. The proteins are listed under their gene names and species names, abbreviated as follows: Avin, Azotobacter vinelandii; Bsub, B. subtilis; Bant, B. anthracis; Caur, Chloroflexus aurantiacus; Ceff, Corynebacterium efficiens; Cgлу, C. glutamicum; Mlot, Mesorhizobium loti; Nos, Nostoc (Anabaena) sp. PCC7120; Pmar, Prochlorococcus marinus; Pmul, Pasteurella multocida; Rsph, Rhodobacter sphaeroides; Saur, Staphylococcus aureus; Scoe, Streptomyces coelicolor; Spne, S. pneumoniae; Tery, Trichodesmium erythraeum; Vvul, Vibrio vulnificus. Missing amino acid residues at the end of the alignment indicate protein C-termini. The secondary structure prediction shown below the alignment represents a tentative consensus of six different prediction algorithms [26]. The NCBI protein database or Swiss-Prot accession numbers, where available, are shown in the right column. Conserved residues are shown in bold typeface and/or colored as follows: acidic - red; hydrophobic - yellow background; small (Gly, Ser, Ala) and Pro - green background. Cys residues that form predicted disulfide bond are shown in white letters on blue background. The yellow numbers on dark background indicate the lengths of variable inserts in the respective protein sequences. B: Consensus sequence of the Excalibur domain drawn in the SeqLogo format using the WebLogo tool (http://weblogo.berkeley.edu [44]). The height of each letter indicates the degree of its conservation, the total height of each column represents the statistical importance of the given position. The residue numbering is based on the sequence of the mature form of SP0667 (lacking the 26-aa signal peptide).
of the metallo-β-lactamase superfamily [33], closely related to the C-terminal part of B. subtilis competence protein ComEC [34]. Although the principal catalytic residues of β-lactamase are conserved in this protein (data not shown), and there is no doubt that it possesses a hydrolytic activity, its exact substrate specificity and the nature of the bound metal remain unclear; it could well be another extracellular endonuclease.

The Corynebacterium efficiens CE2653 gene product combines the Excalibur domain with a previously uncharacterized actinobacteria-specific globular domain that is distantly related to the periplasmic endonuclease I of E. coli [35] and can be predicted to have a similar activity. In other proteins from corynebacteria and streptomycetes, as well as from streptococci and staphylococci, the Excalibur domain is found in association with various low-complexity domains, which might be involved in adhesion and colonization processes. While immunogenic properties of these proteins have not been investigated so far, they might represent interesting vaccine candidates.

In two gamma-proteobacteria, Shewanella oneidensis and Vibrio vulnificus, the Excalibur domain is found associated with the cold-shock domain, known to bind RNA and single-stranded DNA [36,37]. This is the only context where Excalibur domain is predicted to be inside the cyto-
toplam, and its exact function there is as obscure as that of the cold-shock domain (see [38] for discussion). The absence of the Excalibur domain in all other γ-proteobacteria whose genomes have been completely sequenced to date probably indicates its relatively recent acquisition by S. oneidensis and V. vulnificus.

Finally, in several organisms, including M. tuberculosis, the Excalibur domain may be found on bacterial surface in the stand-alone form (Fig. 2). As this predicted protein has not been described before and was even missed in genome annotation of all M. tuberculosis strains sequenced so far, no data on its possible functions is currently available. Nevertheless, it seems plausible that the function of this domain in mycobacteria and in P. multocida could be related, respectively, to the maintenance of the stability of the cell envelope and capsular polysaccharide.

3.3. Structural aspects of Ca$^{2+}$ binding by EF-hands and the Excalibur domain

Classical EF-hands contain a helix–loop–helix structure with the Ca$^{2+}$ ion binding by the residues in the loop [8–10]. The Ca$^{2+}$ ion binds in a pentagonal bipyramidal fashion with the side chains of three turn residues, numbered 1, 3 and 5 in the corresponding PROSITE motif [39], typically aspartates, acting as monodentate ligands. A further side chain, contributed by residue 12, typically Glu, is a bidentate calcium ligand while the main chain carboxyl group of residue seven and water molecules complete the coordination of the Ca$^{2+}$ ion. Allowing for a different separation of aspartates and glutamate – four residues in the Excalibur domain, six residues for EF-hands – these characteristics appear to be entirely conserved in the former (Fig. 1). Examination of EF-hand structures and alignments [21] shows that other, non-ligating positions have characteristic residue distributions for structural reasons. Positions 4 and 6 lie, respectively, close to the core of the left-handed helical region of the Ramachandran plot, and at the very periphery of this region. These locations of the Ramachandran plot would be expected to define, respectively, a preference for turn-favoring residues and a near-exclusive preference for glycine residues. Indeed, exactly these tendencies are observed in the EF-hand consensus loop sequence. Once again, these residue preferences for positions 4 and 6 of the EF-hand are reflected exactly in the consensus sequence for the Excalibur domain (Fig. 1). These considerations favor an evolutionary relationship, or at least a structural similarity, between the Excalibur and EF-hand domains. Consistent with this idea, the lesser distance between DxDxD and conserved Glu in the novel domain is readily structurally accommodated (Fig. 2).

That the overall structures of the Excalibur and EF-hand domains differ is guaranteed by the fact that the novel domains terminate shortly after the conserved Glu while a further 12-residue helix is part of the EF-hand domain. Inconsistencies between secondary structure predictions for the Excalibur domain obtained by different methods do not allow for unequivocal determination of the presence or absence in it of the helix that precedes the Ca$^{2+}$-binding loop in EF-hands. In any case, shared structural context is neither required nor sufficient for the inference of homology, given the increasingly appreciated subtleties of protein structural evolution [40].

The two entirely conserved Cys residues in this extracellular protein family are suggestive of the existence of a disulfide bridge. It may be that this structural constraint substitutes for the larger domain framework of EF-hand containing proteins in orienting the ends of the Ca$^{2+}$-binding loop correctly for metal binding. Although simple, more generalized binding motifs can arise by convergent evolution [41], we believe that the extensive local similarities between EF-hands and the novel domain indicate a genuine evolutionary relationship.

3.4. Diversity and evolution of the EF-hand-like domains

Although we have compared the Excalibur domain with the relatively well-understood EF-hand, the Excalibur domain is, intriguingly, the fifth structural context in which the core EF-hand-like Ca$^{2+}$-binding loop has been identified. Structural determinations of the periplasmic galactose-binding protein from S. typhimurium (PDB entry 1gcg) revealed, surprisingly, a DxDxD calcium-binding loop, nearly superimposable on those of EF-hands, in a helix–loop–strand structural context [17]. Just as in EF-hands, the third monodentate side chain ligand (a carbon of an Asp residue) is followed there by glycine, which allows the interaction with the bound calcium of the main chain carbonyl of the residue in the seventh position. Such similarities again argue for an evolutionary relationship between the two loops, despite differences in the position of the glutamate acting as the bidentate Ca$^{2+}$ ligand.

In the dockerin domain, the E-helix in front of the Ca$^{2+}$-binding loop is missing, but the F-helix follows the loop, just as in typical EF-hands. The context of the Ca$^{2+}$-binding loop in dockerins may therefore be described as coil–turn–helix. The same separation of the DxDD calcium-binding loop and the conserved bidentate calcium ligand, in this case an aspartate, is present in the dockerin domain as in the EF-hand. Essentially the same organization of the Ca$^{2+}$-binding site is seen in the B. anthracis protective antigen. Again the characteristic stereochemistry of the Ca$^{2+}$-binding loop is reflected in similar residue preferences to the EF-hand, even outside the actual calcium ligands, providing further support for an evolutionary relationship.

Another notable example is the Ca$^{2+}$-binding protein from the flatworm Echinococcus granulosus [42], where the DxDD motif, followed by a conserved acidic residue at the typical EF-hand spacing, is present in 15 tandem repeats of 12–14 residues. Based on limited prediction of
β-structure for some of the repeats, a structural relationship to the β-roll Ca\(^{2+}\)-binding proteins was proposed [42]. There are several reasons to doubt this idea, not least the fact that the currently available programs predict very little regular secondary structure in the repeat region (analysis not shown). Most importantly, given the excellent match between the *E. granulosus* repeat consensus and the EF-hand consensus and the former’s complete lack of similarity with the β-roll consensus, the more conservative explanation is that the repeats of the *E. granulosus* protein bind calcium in loops of similar structure to those of EF-hands.

Based on the discovery of calmodulin-like proteins in bacteria, it has been argued that the EF-hands had originated in the prokaryotic kingdom [11,14]. While this could still be correct, one has to consider that the Ca\(^{2+}\)-binding loop is found in bacteria in a variety of diverse structural contexts: helix-loop-helix in calerythrin [13] and soluble lytic transglycosylase [16], helix-loop-strand in the periplasmic galactose-binding protein from *S. typhimurium* [17], coil-loop-helix in the dockerin domain [19] and in *B. anthracis* [17], coil-loop-helix in the dockerin domain [19] and in *S. typhimurium* [18], and as the C-terminal loop in the Excalibur domain (Fig. 1). This suggests that evolution of the EF-hand-like domains might be much more complex than previously appreciated and these domains could be much more widespread than was previously recognized, involving various modifications of the structural elements surrounding the Ca\(^{2+}\)-binding loop. The classical EF-hand is therefore just the most visible representative of the whole superfamily, not necessarily the most ancient one.

In conclusion, prokaryotic relatives of the Ca\(^{2+}\)-binding EF-hand domain clearly exhibit a significantly wider sequence and structural diversity than has been so far appreciated, with the Excalibur domain representing a further member of a superfamily of Ca\(^{2+}\)-binding domains. Five structural contexts for the core Ca\(^{2+}\)-binding loop motif, originally associated solely with EF-hands, are now known: EF-hands themselves, the periplasmic galactose-binding protein, the dockerin domain/protective antigen [18], and as the C-terminal loop in the Excalibur domain (Fig. 1). This suggests that evolution of the EF-hand-like domains might be much more complex than previously appreciated and these domains could be much more widespread than was previously recognized, involving various modifications of the structural elements surrounding the Ca\(^{2+}\)-binding loop. The classical EF-hand is therefore just the most visible representative of the whole superfamily, not necessarily the most ancient one.

In conclusion, prokaryotic relatives of the Ca\(^{2+}\)-binding EF-hand domain clearly exhibit a significantly wider sequence and structural diversity than has been so far appreciated, with the Excalibur domain representing a further member of a superfamily of Ca\(^{2+}\)-binding domains. Five structural contexts for the core Ca\(^{2+}\)-binding loop motif, originally associated solely with EF-hands, are now known: EF-hands themselves, the periplasmic galactose-binding protein, the dockerin domain/protective antigen, the Excalibur domain, and the repeated structure typified by a *E. granulosus* protein. This indicates a very complex evolutionary history, involving acquisition and/or loss of flanking helices and strands. Another possibility is that the Ca\(^{2+}\)-binding loop motif is an evolutionarily mobile unit that has become implanted into several different ‘host’ proteins (see [40] for the discussion of possible scenarios). Direct experiments will be needed to define the functions of the Excalibur domain in each of its diverse domain contexts and, indeed, to confirm experimentally the computational prediction of Ca\(^{2+}\)-binding ability.

It is relevant to note that the other classes of proteins containing these Ca\(^{2+}\)-binding loops bind calcium for a variety of reasons – buffering or regulation (EF-hands [8–10]), structural (lytic transglycosylase [16] and periplasmic galactose-binding protein [17]) and for calcium mobilization to and/or from deposits (the *E. granulosus* Ca\(^{2+}\)-binding protein [42]). Hence, it also remains to be seen whether binding and storing Ca\(^{2+}\) ions is the sole biological function of the Excalibur domain, as it has been proposed for other bacterial calmodulin-like proteins [43], or whether the metal-bound Excalibur domain has some additional function. Predicted nucleic acid-binding proteins are relatively abundant among those containing Excalibur domains, raising the possibility that this domain might bind to DNA or RNA.

**Acknowledgements**

Supported in part by NIH Grant AI44079 and DARPA Grant N66001-01-C-8013 (M.J.J.) and by CNPq and PRODETAB (D.J.R.).

**References**
