The DxDxDG Motif for Calcium Binding: Multiple Structural Contexts and Implications for Evolution

ARTICLE *in* JOURNAL OF MOLECULAR BIOLOGY · NOVEMBER 2004

Impact Factor: 4.33 · DOI: 10.1016/j.jmb.2004.08.077 · Source: PubMed

2 AUTHORS:

Daniel J Rigden
University of Liverpool
181 PUBLICATIONS **4,205** CITATIONS
SEE PROFILE

Michael Galperin
National Institutes of Health
207 PUBLICATIONS **12,534** CITATIONS
SEE PROFILE

doi:10.1016/j.jmb.2004.08.077 J. Mol. Biol. (2004) 343, 971–984

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The DxDxDG Motif for Calcium Binding: Multiple Structural Contexts and Implications for Evolution

Daniel J. Rigden[1]* and Michael Y. Galperin[2]

[1]*School of Biological Sciences University of Liverpool, Crown St., Liverpool L69 7ZB, UK*

[2]*National Center for Biotechnology Information National Library of Medicine National Institutes of Health Bethesda, MD 20894, USA*

**Corresponding author*

Calcium ions regulate many cellular processes and have important structural roles in living organisms. Despite the great variety of calcium-binding proteins (CaBPs), many of them contain the same Ca^{2+}-binding helix-loop-helix structure, referred to as the EF-hand. In the canonical EF-hand, the loop contains three calcium-binding aspartic acid residues, which form the DxDxDG sequence motif, and is flanked by two α-helices. Recently, other CaBPs containing the same motif, but lacking one or both helices, have been described. Here, structural motif searches were used to analyse the full diversity of structural context in the known set of DxDxDG-containing CaBPs, including those where the structural resemblance of a given DxDxDG motif to that of EF-hands had not been noted. The results obtained indicate that the EF-hand represents but one, among many, structural context for the DxDxDG-like Ca^{2+}-binding loops. While the structural similarity of the binuclear calcium-binding sites in anthrax protective antigen and human thrombospondin suggests that they are homologous, evolutionary relationships for mononuclear sites are harder to discern. The possible scenarios for the evolution of DxDxDG motif-containing calcium-binding loops in a variety of non-homologous proteins suggested loop transplant as a mechanism perhaps responsible for much of the diversity in structural contexts of present day DxDxDG-type CaBPs. Additionally, while it can be shown that existence of a DxDxDG sequence is not enough to confer a conformation suitable for calcium binding, local convergent evolution may still have a role. The analysis presented here has consequences for the prediction of calcium binding from sequence alone.

Keywords: calcium; EF-hand; protein structure; structure motifs; evolution

Introduction

Calcium ions are key cell regulators, controlling a variety of cellular processes, such as motility, chemotaxis, cell division, differentiation and programmed cell death.[1–3] The effects of calcium ions are mediated by Ca^{2+}-binding proteins (CaBPs), whose affinity to Ca^{2+} ranges from submicromolar (which is comparable to the intracellular Ca^{2+} levels) to millimolar levels. Although several different Ca^{2+}-binding motifs are known,[4,5] by far the most common such motif is a 30-residue helix-loop-helix structure, referred to as the EF-hand.[6–8] In the canonical EF-hand, Ca^{2+} coordination is provided by oxygen atoms from the side-chains of the first, third, and fifth residues from the loop, a backbone atom of the seventh loop residue, a water molecule coordinated by the side-chain of the ninth loop residue, and the side-chain of an acidic (usually Glu) residue in the 12th position in the beginning of the second helix.[7–10] The Ca^{2+}-coordinating residues of the loop are most commonly aspartates forming a DxDxD pattern. However, other residues can also be seen in the third and fifth loop positions, which is reflected in the fairly complicated PROSITE pattern PS00018[11,12] for the EF-hand: D-x-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-x(2)-[DE]-[LIVMFYW]. A sequence logo view (Figure 1) of the EF-hand domains listed in the Pfam database[13] (Pfam entry PF00036) reveals an additional conserved residue, glycine, at the sixth position of the loop, and two conserved hydrophobic residues that form a hydrophobic staple.[11] Although there is a significant sequence

Abbreviations used: CaBPs, calcium-binding proteins; PDB, Protein Data Bank.

E-mail address of the corresponding author: drigden@liverpool.ac.uk

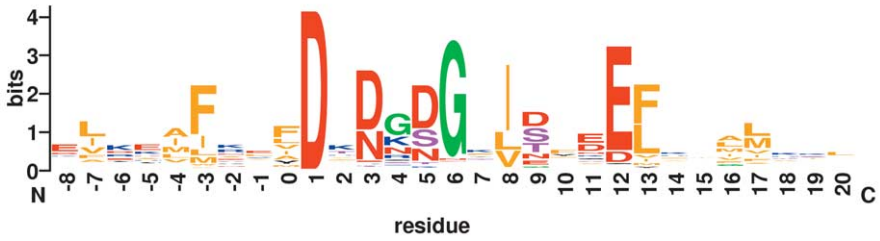

Figure 1. A sequence logo of the EF-hand domain in the Pfam database[13] (Pfam entry PF00036). The Figure was generated from the Pfam seed alignment using the WebLogo[62] tool. The residues are coloured as follows: D, E, N, red; K, R, blue; S, T, purple; G, green; hydrophobic, yellow.

variation in the Ca^{2+}-binding region (Figure 1), for simplicity we refer to the Ca^{2+}-binding motif throughout this article as DxDxDG. Sequence variation is also observed in the motifs found in other structural contexts but a preference for Asp at the first, third and fifth positions is always evident.

While the canonical EF-hand domain appears to provide the highest affinity of Ca^{2+} binding ($K_d \sim 0.1$–1 mM), certain CaBPs display only one of its two characteristic features, either the helix-loop-helix structural motif or the DxDxDG sequence motif. The first variant, which is found in calbindin, calcyclin, and S100 proteins, is often referred to as a pseudo EF-hand motif.[9,10] It consists of two helices connected by a longer Ca^{2+}-binding loop, so that Ca^{2+} coordination is carried out by main chain oxygen atoms of residues 1, 4, 6 and 9, the water molecule is coordinated by residue 11, and the bidentate ligand (Glu or Asp) is moved to the 14th position, compared to the 12th position in the canonical EF-hand. Because of the involvement of main chain oxygen atoms, there is much less constraint on the amino acid sequence of the Ca^{2+}-binding loop, which makes the pseudo EF-hand motif difficult to identify in standard sequence comparisons. Nevertheless, pseudo EF-hands, just like canonical EF-hands, are typically found in pairs, forming a distinct structural motif and leaving little doubt that these two variants have a common origin.[9] In contrast, the other hallmark of the canonical EF-hand motif, the DxDxDG-containing loop can often be found in non-EF-hand structural context; in some of such cases this loop still retains the capacity to bind calcium ions. Examples include the periplasmic galactose-binding protein from *Salmonella typhimurium*, where a Ca^{2+}-binding loop, nearly superimposable on those of EF-hands, is in a helix-β-turn-loop-strand structural context[14,15] and the dockerin domain, where the N-terminal helix (E-helix), preceding the Ca^{2+}-binding loop, is missing, whereas the C-terminal helix (F-helix) follows the loop, just as in typical EF-hands.[16] In a series of elegant experiments, insertion of the 12-residue Ca^{2+}-binding loop from calmodulin flanked by glycine residues into a loop connecting two β-strands of rat CD2 protein was sufficient to create a new Ca^{2+} and La^{3+}- binding site.[17,18]

Our interest in the Ca^{2+}-binding properties of the DxDxDG-containing loop structure was ignited by the discovery of a conserved protein domain, named Excalibur, in several bacterial proteins.[19] Presence of the conserved DxDxDG sequence motif, predicted to form a C-terminal loop, flanked by a putative disulphide bridge in all Excalibur sequences, suggested that this domain could bind calcium ions, a hypothesis supported by the Excalibur domain distribution. This prompted us to take a closer look at the DxDxDG-containing loops in bacterial CaBPs that were summarily referred to as EF-hand loops.[20] This study revealed a surprising diversity of the structural contexts where such a loop could be found.[21] Here, we expand our preliminary observations on the structural organisation of the DxDxDG-containing loops in bacterial CaBPs and reveal a dramatic variety of structural contexts for the DxDxDG loop across all kingdoms. We also show that the presence of such a loop in a protein by no means guarantees that it is a CaBP.

Results and Discussion

The set of DxDxDG loops in known structures

The structural motif searches using SPASM[22] and PINTS[23] produced a list of no fewer than 12 different structural contexts in which locally structurally similar DxDxDG loops bind calcium (Table 1; Figures 2 and 3(a)). In one further case (Table 1), strong structural resemblance is seen but no Ca^{2+} is bound; instead a water molecule occupies the corresponding position. In many of these cases the structural resemblance to the EF-hand loop has not previously been noted. Superposition of all residues of the DxDxDG loops using an extended main chain definition (N, C^{α}, C^{β}, C, O) reveals the extent of their structural similarity to the calmodulin representative (Figures 2 and 3(a)) producing r.m.s. deviations of 27–28 atoms (depending on number of glycine residues) of 0.18–1.08 Å (mean 0.45 Å). For comparison, r.m.s. deviations upon comparison to the other DxDxDG loops in the *Paramecium* calmodulin are 0.00–0.42 Å. This value is equal to or higher than those obtained for eight of our alternative structural contexts. For comparison, r.m.s. deviations from a set of non-calcium binding DxDxDG sequences located in

Table 1. Families containing DxDxDG loop proteins of known structure

Representative	Calcium binding status	SCOP class; fold of domain containing DxDxDG loop	PDB code, reference	Residue number of DxDxDG start	Distribution of proteins containing DxDxDG loop	Frequency of DxDxDG loop in homologous proteins[a]	R.m.s. fit of extended main chain atoms of DxDxDG to the first calmodulin motif	Number of residues separating DxDxDG and later calcium ligand(s) (D or E)	Function of bound calcium
Paramecium tetraurelia calmodulin	Confirmed	All α; EF hand-like	1exr[37]	20; 56; 93; 129	Mainly eukaryotic, few bacterial, very few archaeal	Almost 100% of variable number of sites (4 in representative)	0.00–0.42	5 (when present)	Regulation[64] or structural[24] or buffering/transport[65]
Salmonella typhimurium periplasmic galactose-binding protein	Confirmed	α/β; Periplasmic binding protein-like I	1gcg[66]	134	Bacterial	Approx. 3%	0.27	65	Structural[66]
Sphingomonas sp. periplasmic alginate-binding protein	Confirmed	α/β; Periplasmic binding protein-like II	1kwh[25]	171	Bacterial	Approx. 1%	0.36	2/3	Potentially sensor[25]
Clostridium thermocellum dockerin	Confirmed	All α; type I dockerin domain	1daq[16]	8; 40	Bacterial, very few archaeal	100% of 2 sites	0.98–1.12	5	Structural[16]
Escherichia coli soluble lytic transglycosylase Slt35	Confirmed	α+β; Lysozyme-like	1qut[35]	237	Bacterial	100%	0.42	8	Not known
Bacillus anthracis protective antigen	Confirmed	Membrane and cell-surface; anthrax protective antigen	1acc[38]	177	Bacterial	100%	0.31	2[b]/5	Structural[67]
Thermotoga maritima 4-α-glucanotransferase	Confirmed	α/β; TIM β/α-barrel	1lwj[27]	13	Eukaryotic and bacterial	Approx. 35%	0.71	2	Not known
Human integrin αVβ3	Confirmed	All β; 7-bladed β-propeller	1jv2[30]	284; 349; 413	Eukaryotic	100% of 3 sites	0.61–1.07	2	Potentially regulatory[30]

(*continued on next page*)

Table 1 (*continued*)

Representative	Calcium binding status	SCOP class; fold of domain containing DxDxDG loop	PDB code, reference	Residue number of DxDxDG start	Distribution of proteins containing DxDxDG loop	Frequency of DxDxDG loop in homologous proteins[a]	R.m.s. fit of extended main chain atoms of DxDxDG to the first calmodulin motif	Number of residues separating DxDxDG and later calcium ligand(s) (D or E)	Function of bound calcium
Pseudomonas "Tac II 18" alkaline protease	Confirmed	α+β; Zincin-like	1h71[26]	49	*Pseudomonas*, *serratia*	Approx. 35%	0.18	2/59	Not known
Pseudomonas aeruginosa alkaline protease	Confirmed	All β; β-roll	1kap[5]	446	*Pseudomonas aeruginosa*	Approx. 15%	0.37	2	Not known
Human thrombospondin-1	Confirmed	Not yet assigned	1ux6[29]	828; 843; 866; 879; 915; 902	Metazoan	100% of 5 binuclear sites	0.18–0.26	2[b]/5	Regulation of binding[68] or mobilization of deposits[50]
						100% of 1 mononuclear site	0.24	2[b]	
Human transglutaminase 3	Confirmed	α+β; Cysteine proteinases[c]	1vjj[31]	301	Eukaryotic	Approx. 20%	0.42	18	Regulation of catalytic activity[31]
Thermoascus aurantiacus cellulase	Predicted	α/β; TIM β/α-barrel	1gzj[34]	287	Bacterial and yeast	Approx. 10%	0.72	No later D or E predicted to bind calcium	Not known

[a] As defined by PFAM, SMART or by full-length matches in PSI-BLAST (*e*-value of 0.0001) run until convergence.
[b] Through water interaction with calcium.
[c] Entry for a homologous protein.

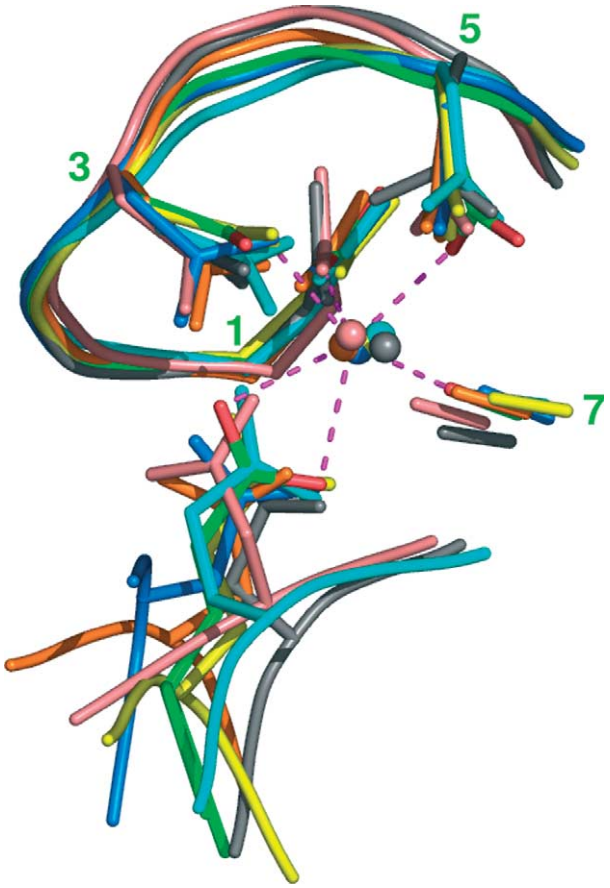

Figure 2. Comparison of mononuclear DxDxDG calcium-binding motifs with three D or N calcium ligands determined by X-ray crystallography. The DxDxDG backbones superimpose very well as do the side-chains and carbonyl, which coordinate calcium (labelled 1, 3, 5 and 7). In contrast, additional interactions by later acidic residues (in some cases not all are shown) are highly variable. The representative calmodulin (for PDB codes see Table 1) is shown coloured by atom type, with green carbon and magenta bound calcium, while other structures and their bound calcium ions are uniformly coloured as follows: yellow for galactose-binding protein, cyan for alginate-binding protein, salmon for glucanotransferase, orange for transglutaminase, grey for integrin, blue for alkaline protease. Interactions with bound calcium are shown as magenta dotted lines. The Figure was made with PYMOL,[63] as were Figures 5(a) and 6(a)–(c).

the Protein Data Bank (PDB), discussed at greater length below, range from 1.87 Å to 3.11 Å.

The DxDxDG loops are found in domains of all structural classes, from all-α to all-β *via* α/β and α+β. Their phylogenetic distribution also varies widely. The EF-hand structural context, for example, is found across eukaryotes, bacteria and archaea, while the DxDxDG loop of the alkaline protease from *Pseudomonas aeruginosa*[5] is, in the current databases, unique to that species. Bacteria contain the largest number of known structural contexts for DxDxDG loops (ten), while the fewest families are seen in archaea (three) and generally in small numbers of sequences. The previously noted general division of functions between largely structural in bacteria to regulatory in eukaryotes[21] seems to remain true. The only known exceptions to the trend seem to be the EF-hand, whose occasional high affinity leads to a purely structural role in some eukaryotic cases,[24] and the bacterial alginate-binding protein, for which a sensor role has been proposed.[25] In some cases, the DxDxDG loop is present in all known proteins of a given protein family; in others the DxDxDG loop is found in only a small fraction of the sequence in the respective family. Curiously, there seems to be a tendency for Ca^{2+}-binding DxDxDG loops to be found in proteins containing additional Ca^{2+}-binding sites of different geometry. Thus, the alkaline proteases contain several calcium ions bound in their β-roll domains,[5,26] and the glucanotransferase,[27] like some related TIM barrels involved in carbohydrate processing, contains a catalytically important Ca^{2+}-binding site.[28] Thrombospondin,[29] integrin[30] and transglutaminase[31] also contain non-DxDxDG-like Ca^{2+}-binding sites. It may be that the additional Ca^{2+} binding sites help to raise local metal concentration thereby maintaining occupancy of the key structural or catalytic sites. Such a role has been suggested, for example, for the non-catalytic heavy metal sites in the metal-dependent enzyme phytochelatin synthase.[32]

In the DxDxDG loops of EF-hands ligation of the Ca^{2+} usually involves the side-chains of the D positions (occupied by D, N and occasionally S) along with the main chain carbonyl of the residue immediately following the G position.[10] Usually, the side-chain of a further glutamate, separated from the G position by five residues, forms additional bidentate interactions with the bound Ca^{2+}, although the penta-EF-hand structures provide examples in which these interactions are absent.[33] The data in Table 1 show just how variable the Ca^{2+} interactions from outside the DxDxDG loop are. In the case of the predicted Ca^{2+}-binding site in *Thermoascus aurantiacus* cellulase,[34] there is no Asp or Glu following the loop that would interact with Ca^{2+}. Instead, the side-chain of the Thr residue following the G position is well positioned to form a hydrogen bond to the Ca^{2+}. Uniquely, therefore, binding by this DxDxDG loop would not involve the main chain of the residue following the G position. Apart from this cellulase, side-chains of acidic residues outside the loop invariably contribute to Ca^{2+} binding, usually through direct interaction but in the cases of the binuclear Ca^{2+} sites in anthrax protective antigen and integrin, *via* water (Table 1). In nine cases a single additional acidic residue is involved, in four cases two additional acidic residues contribute to binding (note that the thrombospondin structure contains distinct modes of Ca^{2+} ligation at its binuclear and mononuclear sites). The separation of DxDxDG loop and later acidic residues contributing to Ca^{2+} binding varies even more widely than previously supposed[21] from two residues to 65 residues. Curiously, there is no example in which

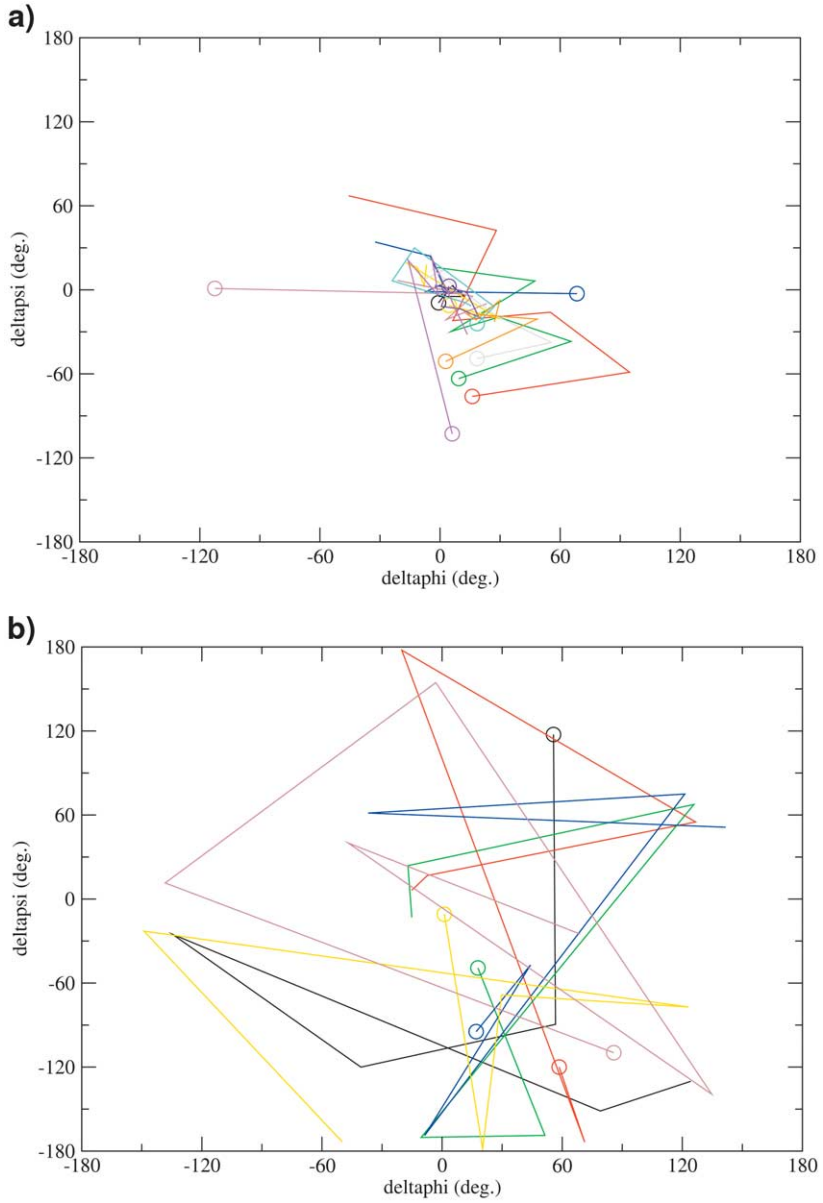

Figure 3. Difference Ramachandran (δ(phi)/δ(psi)) plots show (a) local structural homogeneity for a set of 12 DxDxDG sequences with the EF-hand-type local conformation and (b) wide local structural diversity for a set of six DxDxDG sequences not adopting the EF-hand loop-like structure in a 25% non-sequence redundant selection from the PDB. Differences are calculated using the first DxDxDG sequence of *Paramecium* calmodulin 1exr as a reference (Table 1). For each example, δ(phi)/δ(psi) points for consecutive residues are joined by a line with the first point marked with a circle. The DxDxDG sequences in (a) are contained in PDB entries 1gcg (starting residue 134; black), 1daq (40; red), 1lwj (13; green), 1acc (177; blue), 1qut (237; yellow), 1kwh (171; brown), 1uok (21; grey), 1h71 (49; violet), 1kap (446; cyan), 1gzj (202; magenta) and 1vjj (301; orange). The DxDxDG sequences in (b) are contained in PDB entries 1f96 (starting residue 9; black), 1ir0 (26; red), 1dg4 (477; green), 1ir6 (80; blue), 1ksi (451; yellow) and 1bgw (527; brown).

the additional acidic residue precedes the DxDxDG motif, although there seems no reason why such a configuration should not exist.

Different structural contexts for the DxDxDG loop have previously been typically perceived as EF-hands of different length.[20,35] Similarly, the dockerin structure was described as an EF-hand variant lacking the E helix.[16] Based on the dramatic variety among post-DxDxDG calcium interaction, we propose that the most useful core definition should relate to the DxDxDG motif alone. The presence of this motif in EF-hands could then be considered simply as one among a number of diverse contexts. This conclusion is reinforced by consideration of Figure 4, which shows the secondary structural context of the DxDxDG-type loops in the structures listed in Table 1. It is apparent that α-helices are not required, or even common, in the polypeptide chain leading up to the DxDxDG loop, nor in the following regions. The region immediately N-terminal of the DxDxDG loops may contain, as nearest neighbouring elements of regular

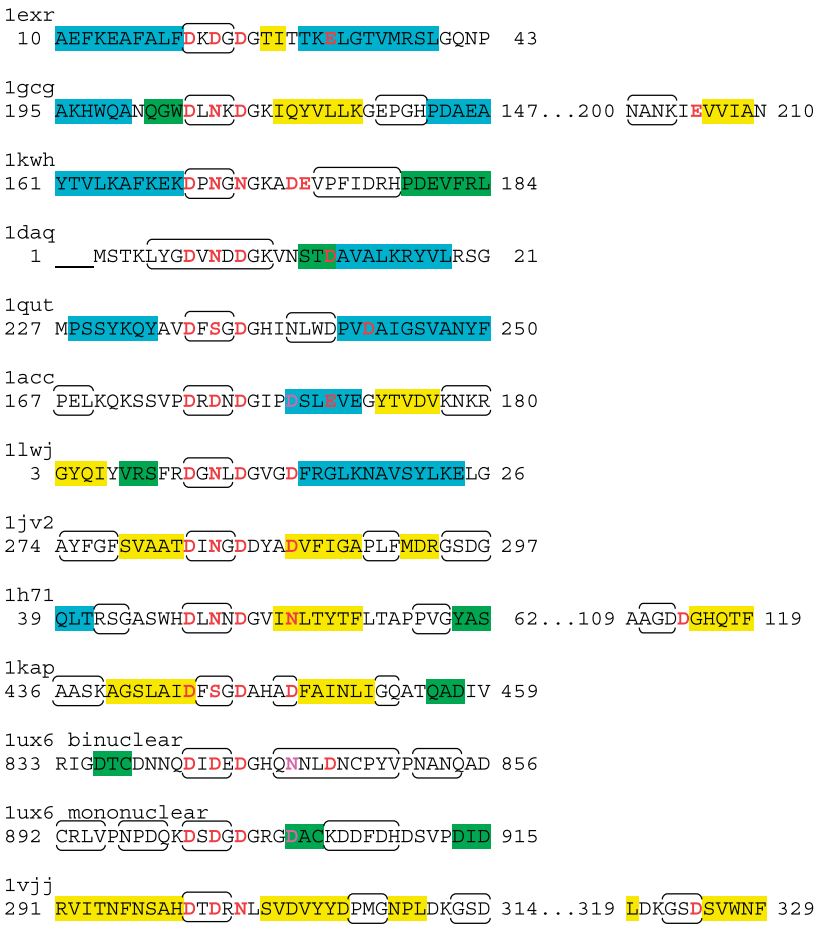

Figure 4. Secondary structure context of the DxDxDG loops and other calcium-coordinating residues for the confirmed calcium-binding proteins are shown in Table 1. Residues binding using side-chains are shown in red (direct interaction with calcium) or purple (through-water interaction). Secondary structure was defined with STRIDE[36] and is indicated as follows: α-helices, blue shading; β-strands, yellow shading; 3_{10} helices, green shading; turns, brackets.

secondary structure, α-helices (four cases), 3_{10}-helices (three cases) or β-strands (three cases). In two further cases the ten preceding residues contain no regular secondary structure and in the case of dockerin only seven residues lie N-terminal of the first DxDxDG loop. Similarly, immediately following the DxDxDG loop the conformation may be helical, extended or irregular. In fact, according to STRIDE,[36] the EF-hand of *Paramecium tetraurelia* calmodulin,[37] used as a representative EF-hand in Table 1, contains a short, two-residue β-strand before the F-helix. Interestingly, in the three cases in which an additional interaction is contributed by an acidic residue not within 15 residues of the DxDxDG loop, that residue invariably immediately precedes a β-strand. However, even among these three cases there is a difference. In transglutaminase, the strand following the additional residue lies antiparallel with the strand preceding the DxDxDG loop. In periplasmic galactose-binding protein and in *Pseudomonas* "tac II 18" alkaline protease the strand following the additional residue lies parallel with the strand following the DxDxDG loop.

Evolution of the binuclear calcium-binding sites

When the DxDxDG loops of the binuclear sites in anthrax protective antigen[38] and human thrombospodin[29] were overlaid, a remarkable structural similarity was observed throughout the whole binuclear Ca^{2+}-binding site (Figure 5). Forty extended main chain atoms of the Ca-coordinating residues shown in Figure 5 could be fitted with an r.m.s. deviation of 1.13 Å. Not only do both Ca^{2+} superimpose well, but so does a buried bound water molecule. In addition, most side-chain conformations are strongly similar. These similarities are so high that it is difficult to imagine that they could be due to convergent evolution, which suggests a shared evolutionary ancestry of these regions. This conclusion is supported by the good match between the region of close structural correspondence (residues 175–237 of protective antigen correspond to residues 841–863 of thrombospondin; protective antigen contains two insertions: see Figure 5) and the C-type repeat motif defined by structural and sequence analysis of thrombospondin.[29] The C-type repeat motif was

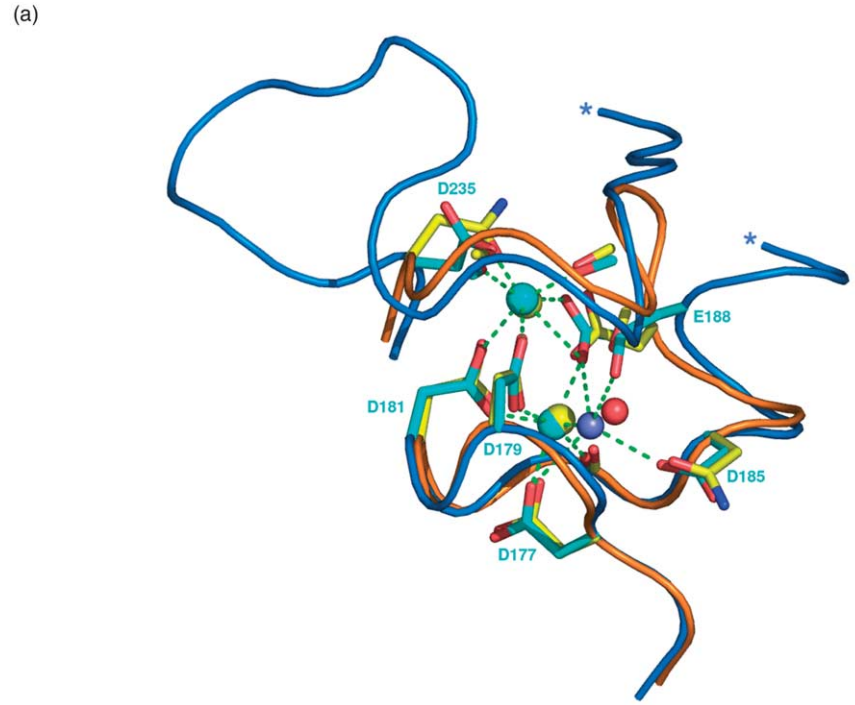

Figure 5. Comparison of the binuclear calcium-binding centres of anthrax protective antigen (PDB code 1acc[38]) and human thrombospondin (PDB code 1ux6[29]). (a) Structural comparison of protective antigen (shades of blue) and thrombospondin (shades of red and yellow) along with bound calcium (two larger spheres) and water (smaller sphere). Protective antigen residues making side-chain contacts to bound calcium or water are labelled. Green broken lines represent interactions with bound calcium and water. Asterisks mark the limits of the larger protective anthrax insertion, which is not shown for clarity. (b) Structure-based sequence alignment. Residues making contacts with either calcium or the bound water are shown in bold; those making contacts through their side-chain are coloured purple.

defined as encompassing, in this case, residues 843–864, matching closely the region, residues 841–863, for which structural similarity to protective antigen was observed. This region therefore has the characteristics of a domain, consistent with its duplication in thrombospondin. Most interestingly, the portion of protective antigen lying immediately prior to the binuclear Ca^{2+}-binding site has been shown to be representative of a module, the PA14 domain, widely distributed in bacteria and eukaryotes, with likely carbohydrate-binding function.[39] Hence, it is possible that *Bacillus anthracis* obtained both the presumed carbohydrate-binding domain and the Ca^{2+}-binding domain of its protective antigen from a eukaryotic source, perhaps even its human host.

Evolution of the mononuclear calcium-binding sites

While the two binuclear sites seem to be homologous, deciphering the possible relationships between the more numerous mononuclear sites is much more difficult. The possibility that the domains in which they are found all descend, in their entirety, from a common ancestor may be ruled out due to the dramatic size and structural differences seen between them. One exception may be the EF-hands and the dockerins, which each typically contain paired DxDxDG loops, are mainly helical and are comparably sized. For the remaining relationships two scenarios can be envisaged. In the first, since the DxDxDG motif is relatively simple, convergent evolution may have acted independently on loops located in various structural contexts to produce closely structurally similar Ca^{2+}-binding loops. Alternatively, the DxDxDG loop might represent an evolutionarily mobile element that has been spliced into several unrelated folds.

In support of the first hypothesis may be cited a recent search for examples of structurally similar triads in unrelated folds which yielded many examples of metal-binding sites.[40] While the examples found show how easily convergent evolution may produce sets of residues that superimpose reasonably well, the cases cited differ from the DxDxDG loop in important ways. Firstly, the degree of structural similarity between the DxDxDG loops is higher. Secondly, the unrelated positions along the chain of pairs of convergently evolved triads rule out an explanation involving binding site transplant. The abundance of mirror image sites[40] leads to the same conclusion.

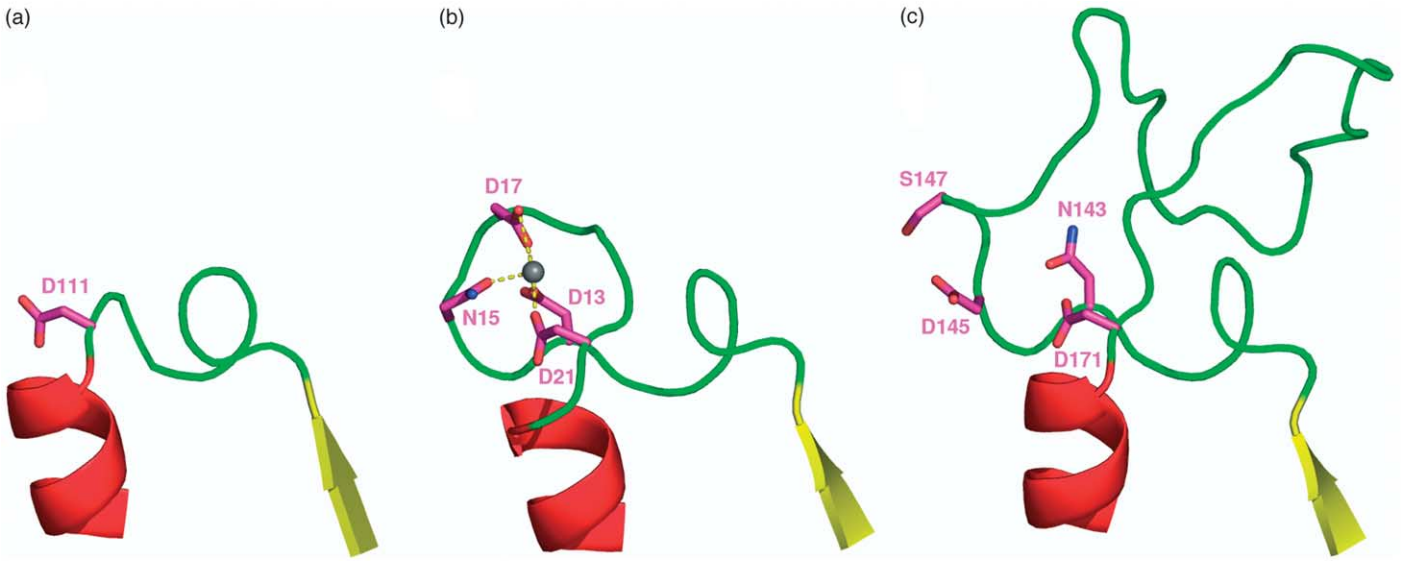

Figure 6. Comparison of the DxDxDG loop-containing *Thermotoga maritima* glucanotransferase (b) with *Neisseria polysaccharea* amylosucrase (a) and *Thermoactinomyces vulgaris* R-47 alpha-amylase II (c) in the vicinity of the loop. In (b) the grey sphere represents the bound calcium and dotted lines are used for calcium interactions. (d) A structure-based sequence alignment of the enzymes for the same region. The residues shown in (a)–(c) are indicated in bold and purple.

In contrast, the DxDxDG loop examples in Table 1 have the three D positions, close together in the chain, as by far the most important contributor to Ca^{2+} binding. Thus, the idea of loop transplant, by a mechanism as yet unknown, becomes viable. It is also important to remember that while pairs of convergently evolved sites may be readily found, we seek an explanation for at least 12, probably more, distinct structural contexts of superimposable DxDxDG loops.

Convergent evolution would be favoured if the motif DxDxDG were to have particular properties that strongly favoured a certain local conformation. In general, studies have shown that the strength of influence of protein sequence on local structure varies. On the one hand, large peptides of identical sequence (chameleon sequences) may adopt different structures in different, unrelated proteins.[41,42] Even within the same protein, a fragment may sometimes undergo dramatic local structural changes, as has been shown, for example, in the serpins.[43] On the other hand, clusters of similar protein structure fragments from unrelated proteins sometimes display significant sequence tendencies.[44,45] What can be said, with regard to the DxDxDG motif, is that its presence in a protein structure is insufficient to confer the EF-hand loop-like local structure. In a sequence search of a 25% sequence redundancy cut of the PDB for the sequence DxDxDG, six structures were returned in which the motif does not adopt the EF-hand loop-like structure (Figure 3(b)), compared to 12 in which it does (Figure 3(a)). Furthermore, it is not the case that the non-EF-hand-like motifs are necessarily buried, in contrast to the generally strong solvent exposure of EF-hand-like motifs. The mean solvent-exposed surface area of the 12 EF-hand-like DxDxDG motifs is 338 $Å^2$, a value exceeded by three of the six non-EF-hand-like motifs (1f96 476 $Å^2$; 1ir0 519 $Å^2$; 1bgw 420 $Å^2$). Thus, simple solvent exposure of the sequence DxDxDG is insufficient to confer an EF-hand-like local structure.

Additional evidence in favour of the idea of transplant of the DxDxDG loop between folds, are several examples in which structural comparison highlights pairs of structures, in one of which the DxDxDG loop appears as a structural insertion. We have already illustrated this phenomenon in both periplasmic binding protein cases.[21] The further DxDxDG loop contexts described here provide further examples. For example, as shown in Figure 6, *Neisseria polysaccharea* amylosucrase (PDB code 1g5a[46]) is shorter than the *Thermotoga maritima* glucanotransferase[27] by seven residues, corresponding to the DxDxDG loop plus a single further residue.

In fact, among the new DxDxDG loop examples that we highlight here, the case of glucanotransferase is particularly interesting since the broad phylogenetic range and the range of catalytic activities born by homologous DxDxDG loop-containing enzymes (maltase, amylase, trehalose synthase, trehalose-6-phosphate hydrolase, sucrase, glucodextranase) suggest that the DxDxDG loop was already present in a distant ancestor. Furthermore, available structures are consistent with an evolutionary scenario in which the DxDxDG was acquired but then lost function through mutation, although other possibilities can be imagined. As mentioned, the amylosucrase structure (Figure 6(a)) lacks the DxDxDG loop and could therefore, in this respect, resemble a putative ancestor that later received the transplant of the DxDxDG loop. In contrast, *Thermoactinomyces vulgaris* R-47 alpha-amylase II (PDB code 1bvz[47]) has a longer corresponding loop (Figure 6(c)). Sequence similarity to the DxDxDG loop of the glucanotransferase is still discernible in the longer loop (Figure 6(d)) but the extra residues have led to distortion such that the Ca^{2+} cannot bind. Of course, the same structures could equally have arisen from mutation leading to the DxDxDG loop in the ancestor of a proportion of extant enzymes. Nevertheless, the set of structures shown in Figure 6, combined with the truly remarkable ease of superposition of the DxDxDG loop examples (Table 1; Figure 2) means that the idea of transplant of loop between folds should be taken seriously.

Although the transplant hypothesis is plausible, at least for many of the loop contexts, our searches did uncover a related example that appears more readily explained by convergent evolution. When searches were carried out using DxDxD, a match was obtained with rabbit muscle phosphoglucomutase (PDB code 3pmg[48]). The extended main chain atoms of the five DxDxD residues Asp287–Asp291 superimpose with r.m.s. deviation of 0.73 Å on the first DxDxDG loop of our reference calmodulin. This DxDxD loop binds the catalytic divalent cation in the sugar mutase enzymes. Although bearing structural and functional similarity to the cases in Table 1, several factors suggest that it may be a chance resemblance, arising through convergent evolution, rather than a result of loop transplant. First, the structural similarity of the examples in Table 1 extends to the G position (whether occupied by a glycine or not): it is the positive phi angle at the G position that enables ligation of the Ca^{2+} by the following residue. In contrast, in the mutase enzymes the polypeptide chain adopts a different conformation following the DxDxD motif. Secondly, the mutase site has a preference for magnesium over Ca^{2+}, while Ca^{2+} preference is established for most examples in Table 1. Thirdly, the loop, located in the second of three ancient duplicated domains has contacts with both flanking domains. Thus, a putative three domain structure ready to receive a loop transplant might not maintain the defined relative interdomain positioning that leads to the formation of the catalytic site cleft. Such a putative acceptor might therefore be considered less likely to have a catalytic activity. While none of these factors is conclusive, taken together they suggest that convergent evolution might also have had a hand in producing the

present day diversity of DxDxD-containing metal-binding proteins.

DxDxDG families of unknown structures

As shown in Table 2, there are at least four further families of DxDxDG families of unknown structure. In the cases of FrpC,[49] a member of the RTX protein family, *E. granulosus* Ca^{2+}-binding protein[50] and caleosin,[51] Ca^{2+} binding has been experimentally observed. In the case of the Excalibur domain, Ca^{2+} binding has not yet been demonstrated, but the combinations of several Ca^{2+}-dependent or Ca^{2+}-regulated domains with Excalibur provide circumstantial support for binding.[19] In each case, given the foregoing, it is reasonable to assume that the conserved DxDxDG motifs in each are responsible for Ca^{2+} binding. In the case of caleosin, an unusual single EF-hand structure has been proposed and is supported by the typical EF-hand separation of DxDxDG loop and later, presumed Ca^{2+}-coordinating Glu and our own fold recognition analysis (not shown). Each of the other families has some interesting features. In the case of Excalibur, conserved Cys residues, presumed to form a disulphide bridge, flank the DxDxDG motif and some proteins terminate at the downstream, additional Ca^{2+}-coordinating Glu.[19] The *E. granulosus* Ca^{2+}-binding domain consists mainly of 15 copies of a 12–14 residues repeat in which a DxDxDG motif is separated by two residues from a later, invariant Asp.[50] In the present database, only another tapeworm and *Drosophila* seem to possess homologues with the same characteristics. In FrpC, five DxDxDG motifs can be identified, most interestingly with diverse separations of DxDxDG loops and putative additional Ca^{2+} ligands (Table 2). Mutation of acidic residues in the third, fourth and fifth of these motifs affects Ca^{2+}-dependent self-processing of FrpC.[49] Unfortunately, since the selection of residues to mutate was based on the expectation of a typical EF-hand separation of motif and later Ca^{2+} ligand of five residues, positions in the first two DxDxDG motifs were not subjected to mutagenesis so that their contributions to activation remain unknown. Indeed, one of the main conclusions from our survey must be that prediction of Ca^{2+} binding must consider the impressive variability of motif-additional ligand spacing, from 2 to 65 residues (Tables 1 and 2).

Conclusions

Although by far the best studied, the EF-hand must now be considered as simply one structural context, among many, for the DxDxDG-like Ca^{2+}-binding motif. The DxDxDG motifs from the remarkable variety of structural motifs superimpose extremely well, in many cases as well as one EF-hand motif fits to another. It seems inconceivable that all the examples cited can share a single ancestor, given their structural diversity, although shared ancestry of EF-hand and dockerin domains can be imagined. Equally, the ease of superposition of the diverse Ca^{2+}-binding DxDxDG motifs, combined with our demonstration that the DxDxDG sequence itself may adopt diverse other conformations, argues against local convergent evolution as being responsible for all cases. Instead, it seems likely that the DxDxDG-containing loop motif has, during evolution, been transplanted, by means unknown, into multiple structural contexts. Such a scenario, although invoking the existence of a novel mechanism of loop transplant, is consistent with the existence of

Table 2. Families containing DxDxDG loop proteins of unknown structure

Representative	Calcium binding status	Distribution of proteins containing DxDxDG loop	Frequency of DxDxDG loop in homologous proteins[a]	Predicted number of residues separating DxDxDG and later potential calcium ligands (D or E)	Function of bound calcium
Neisseria meningitidis FrpC	Confirmed[49]	Bacteria	Motif 1 Approx. 70% Motif 2 Approx. 65% Motif 3 Approx. 90% Motif 4 Approx. 75% Motif 5 Approx. 75%	1 11, 12 2, 6 5 5	Regulation of self-cleaving/ligation [49]
Sesamum indicum caleosin	Confirmed[69]	Plants, one fungus	100%	5	Not known, possibly regulation of activity[51]
Echinococcus granulosus intracellular calcium-binding domain	Confirmed[50]	Metazoa	100%, although repeat number varies	2	Mobilisation of calcium reserves
Streptococcus pneumoniae predicted surface protein SP 0667	Predicted[19]	Bacteria	100%	3	Not known, possibly DNA binding[19]

[a] As defined by PFAM, SMART or by full-length matches in PSI-BLAST (*e*-value of 0.0001) run until convergence.

several cases of structural pairs, in one of which the DxDxDG loop appears as insertion. The binuclear Ca^{2+} centres of anthrax protective antigen and thrombospondin represent a special case in which detailed structural similarity over the whole domain are indicative of homology. The results presented here show that caution in prediction of Ca^{2+} binding should be exercised in two ways. First, the presence of a DxDxDG motif is not necessarily indicative of Ca^{2+} binding potential. Secondly, the separation of the core DxDxDG and additional ligands can vary widely and the latter may even be dispensed with in some cases.

Materials and Methods

Searches for DxDxDG loops in proteins of known structure were carried out using SPASM.[22] SPASM is able to locate motifs of similar 3D structure in different structural contexts using two pseudoatom positions to represent the main and side-chain, respectively, of each residue in the probe structure. The first DxDxDG motif (residues 20–25) of the high-resolution structure of *Paramecium tetraurelia* calmodulin (PDB code 1exr[37]) was used first. Only the D and G positions were used as constraints in the search. Further iterative searches using structures obtained in the first searches were later carried out. Initially, the only residue allowed to substitute for D was N and conservation of G was required. Later the search criteria were relaxed and S allowed at the D positions. Finally, the G position was omitted from the probe structure yielding much longer lists of potential hits with many false positives. All potential hits were screened visually using O[52] in order to locate only those of similar backbone and side-chain conformations to known Ca^{2+}-binding DxDxDG loops. Crystallographic evidence of Ca^{2+} binding was sought but lack of Ca^{2+} was not taken as proof of inability to bind Ca^{2+} since metal might not have been present in the crystallisation solution. The PINTS server†, employing an alternative method to locate local structural similarity,[23] was also used. In this case there was complete correspondence between the hits located by the two methods. The database used with SPASM was obtained by processing the PDB[53] in November 2003. In order to find structures containing DxDxDG loops determined since then literature searches in PubMed and motif searches in sequence database were carried out at the NPS@ facility.[54]

Analyses of the hit structures and sequence relationships in their respective families made use of the SCOP,[55] PFAM[13] and SMART[56] databases. For sequences lacking corresponding sequence database entries PSI-BLAST[57] was used to search for homologous sequences, employing an *e*-value cut-off of 0.0001 until convergence. Structural superpositions were obtained using the CE[58] and FSSP[59] databases, or locally with the program LSQMAN.[60] Secondary structure was assigned with STRIDE[36] and per-residue solvent-exposed surface area calculated with DSSP.[61]

† http://pints.embl.de

Acknowledgements

We thank Eugene Koonin and Wei Yang for helpful comments.

References

1. Smith, R. J. (1995). Calcium and bacteria. *Advan. Microb. Physiol.* **37**, 83–133.
2. Carafoli, E. & Klee, C. B., eds (1999). *Calcium as a Cellular Regulator*, Oxford University Press, New York.
3. Carafoli, E. (2002). Calcium signaling: a tale for all seasons. *Proc. Natl Acad. Sci. USA*, **99**, 1115–1122.
4. McPhalen, C. A., Strynadka, N. C. & James, M. N. (1991). Calcium-binding sites in proteins: a structural perspective. *Advan. Protein Chem.* **42**, 77–144.
5. Baumann, U., Wu, S., Flaherty, K. M. & McKay, D. B. (1993). Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel beta roll motif. *EMBO J.* **12**, 3357–3364.
6. Kretsinger, R. H. (1976). Calcium-binding proteins. *Annu. Rev. Biochem.* **45**, 239–266.
7. Strynadka, N. C. & James, M. N. (1989). Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biochem.* **58**, 951–998.
8. Lewit-Bentley, A. & Rety, S. (2000). EF-hand calcium-binding proteins. *Curr. Opin. Struct. Biol.* **10**, 637–643.
9. Kawasaki, H., Nakayama, S. & Kretsinger, R. H. (1998). Classification and evolution of EF-hand proteins. *Biometals*, **11**, 277–295.
10. Nelson, M. R. & Chazin, W. J. (1998). Structures of EF-hand Ca^{2+}-binding proteins: diversity in the organization, packing and response to Ca^{2+} binding. *Biometals*, **11**, 297–318.
11. Dragani, B. & Aceto, A. (1999). About the role of conserved amino acid residues in the calcium-binding site of proteins. *Arch. Biochem. Biophys.* **368**, 211–213.
12. Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M. *et al*. (2002). PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform.* **3**, 265–274.
13. Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S. *et al*. (2004). The Pfam protein families database. *Nucl. Acids Res.* **32**, D138–D141.
14. Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1987). A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. *Nature*, **327**, 635–638.
15. Zou, J. Y., Flocco, M. M. & Mowbray, S. L. (1993). The 1.7 Å refined X-ray structure of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J. Mol. Biol.* **233**, 739–752.
16. Lytle, B. L., Volkman, B. F., Westler, W. M., Heckman, M. P. & Wu, J. H. (2001). Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. *J. Mol. Biol.* **307**, 745–753.
17. Ye, Y., Lee, H. W., Yang, W., Shealy, S. J., Wilkins, A. L., Liu, Z. R. *et al*. (2001). Metal binding affinity and structural properties of an isolated EF-loop in a scaffold protein. *Protein Eng.* **14**, 1001–1013.
18. Ye, Y., Shealy, S., Lee, H. W., Torshin, I., Harrison, R. & Yang, J. J. (2003). A grafting approach to obtain site-specific metal-binding properties of EF-hand proteins. *Protein Eng.* **16**, 429–434.
19. Rigden, D. J., Jedrzejas, M. J. & Galperin, M. Y. (2003).

An extracellular calcium-binding domain in bacteria with a distant relationship to EF-hands. *FEMS Microbiol. Letters*, **221**, 103–110.
20. Michiels, J., Xi, C., Verhaert, J. & Vanderleyden, J. (2002). The functions of Ca^{2+} in bacteria: a role for EF-hand proteins? *Trends Microbiol.* **10**, 87–93.
21. Rigden, D. J., Jedrzejas, M. J., Moroz, O. V. & Galperin, M. Y. (2003). Structural diversity of calcium-binding proteins in bacteria: single-handed EF-hands? *Trends Microbiol.* **11**, 295–297.
22. Kleywegt, G. J. (1999). Recognition of spatial motifs in protein structures. *J. Mol. Biol.* **285**, 1887–1897.
23. Stark, A. & Russell, R. B. (2003). Annotation in three dimensions. PINTS: patterns in non-homologous tertiary structures. *Nucl. Acids Res.* **31**, 3341–3344.
24. Potter, J. D. & Gergely, J. (1975). The calcium and magnesium binding sites on troponin and their role in the regulation of myofibrillar adenosine triphosphatase. *J. Biol. Chem.* **250**, 4628–4633.
25. Momma, K., Mikami, B., Mishima, Y., Hashimoto, W. & Murata, K. (2002). Crystal structure of AlgQ2, a macromolecule (alginate)-binding protein of *Sphingomonas* sp. A1 at 2.0 Å resolution. *J. Mol. Biol.* **316**, 1051–1059.
26. Aghajari, N., Van Petegem, F., Villeret, V., Chessa, J. P., Gerday, C., Haser, R. & Van Beeumen, J. (2003). Crystal structures of a psychrophilic metalloprotease reveal new insights into catalysis by cold-adapted proteases. *Proteins: Struct. Funct. Genet.* **50**, 636–647.
27. Roujeinikova, A., Raasch, C., Sedelnikova, S., Liebl, W. & Rice, D. W. (2002). Crystal structure of *Thermotoga maritima* 4-alpha-glucanotransferase and its acarbose complex: implications for substrate specificity and catalysis. *J. Mol. Biol.* **321**, 149–162.
28. Buisson, G., Duee, E., Haser, R. & Payan, F. (1987). Three dimensional structure of porcine pancreatic alpha-amylase at 2.9 Å resolution. Role of calcium in structure and activity. *EMBO J.* **6**, 3909–3916.
29. Kvansakul, M., Adams, J. C. & Hohenester, E. (2004). Structure of a thrombospondin C-terminal fragment reveals a novel calcium core in the type 3 repeats. *EMBO J.* **23**, 1223–1233.
30. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L. *et al*. (2001). Crystal structure of the extracellular segment of integrin αVβ3. *Science*, **294**, 339–345.
31. Ahvazi, B., Boeshans, K. M., Idler, W., Baxa, U., Steinert, P. M. & Rastinejad, F. (2004). Structural basis for the coordinated regulation of transglutaminase 3 by guanine nucleotides and calcium/magnesium. *J. Biol. Chem.* **279**, 7180–7192.
32. Cobbett, C. S. (2000). Phytochelatins and their roles in heavy metal detoxification. *Plant Physiol.* **123**, 825–832.
33. Jia, J., Tarabykina, S., Hansen, C., Berchtold, M. & Cygler, M. (2001). Structure of apoptosis-linked protein ALG-2: insights into Ca^{2+}-induced changes in penta-EF-hand proteins. *Structure (Camb)*, **9**, 267–275.
34. Lo Leggio, L. & Larsen, S. (2002). The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Letters*, **523**, 103–108.
35. van Asselt, E. J., Dijkstra, A. J., Kalk, K. H., Takacs, B., Keck, W. & Dijkstra, B. W. (1999). Crystal structure of *Escherichia coli* lytic transglycosylase Slt35 reveals a lysozyme-like catalytic domain with an EF-hand. *Struct. Fold. Des.* **7**, 1167–1180.
36. Heinig, M. & Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucl. Acids Res.* **32**, W500–W502.
37. Wilson, M. A. & Brunger, A. T. (2000). The 1.0 Å crystal structure of Ca^{2+}-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. *J. Mol. Biol.* **301**, 1237–1256.
38. Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. & Liddington, R. C. (1997). Crystal structure of the anthrax toxin protective antigen. *Nature*, **385**, 833–838.
39. Rigden, D. J., Mello, L. V. & Galperin, M. Y. (2004). The PA14 domain, a conserved all-β domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem. Sci.* **29**, 335–339.
40. Hamelryck, T. (2003). Efficient identification of side-chain patterns using a multidimensional index tree. *Proteins: Struct. Funct. Genet.* **51**, 96–108.
41. Minor, D. L., Jr & Kim, P. S. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature*, **380**, 730–734.
42. Mezei, M. (1998). Chameleon sequences in the PDB. *Protein Eng.* **11**, 411–414.
43. Gettins, P. G. (2002). Serpin structure, mechanism, and function. *Chem. Rev.* **102**, 4751–4804.
44. de Brevern, A. G., Valadie, H., Hazout, S. & Etchebest, C. (2002). Extension of a local backbone description using a structural alphabet: a new approach to the sequence-structure relationship. *Protein Sci.* **11**, 2871–2886.
45. Yang, A. S. & Wang, L. Y. (2003). Local structure prediction with local structure-based sequence profiles. *Bioinformatics*, **19**, 1267–1274.
46. Skov, L. K., Mirza, O., Henriksen, A., De Montalk, G. P., Remaud-Simeon, M., Sarcabal, P. *et al*. (2001). Amylosucrase, a glucan-synthesizing enzyme from the alpha-amylase family. *J. Biol. Chem.* **276**, 25273–25278.
47. Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T. & Sakano, Y. (1999). Crystal structure of *Thermoactinomyces vulgaris* R-47 alpha-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. *J. Mol. Biol.* **287**, 907–921.
48. Liu, Y. W., Ray, W. J. & Baranidharan, S. (1997). Structure of rabbit muscle phosphoglucomutase refined at 2.4 angstrom resolution. *Acta Crystallog. sect. D*, **53**, 392–405.
49. Osicka, R., Prochazkova, K., Sulc, M., Linhartova, I. I., Havlicek, V. & Sebo, P. (2004). A novel "clip-and-link" activity of repeat in toxin (RTX) proteins from gram-negative pathogens: covalent protein cross-linking by an Asp-Lys isopeptide bond upon calcium-dependent processing at an Asp-Pro bond. *J. Biol. Chem.* **279**, 24944–24956.
50. Rodrigues, J. J., Ferreira, H. B., Farias, S. E. & Zaha, A. (1997). A protein with a novel calcium-binding domain associated with calcareous corpuscles in *Echinococcus granulosus*. *Biochem. Biophys. Res. Commun.* **237**, 451–456.
51. Murphy, D. J., Hernendez-Pinzon, I., Patel, K., Hope, R. G. & McLauchlan, J. (2000). New insights into the mechanisms of lipid-body biogenesis in plants and other organisms. *Biochem. Soc. Trans.* **28**, 710–711.
52. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.

53. Bourne, P. E., Addess, K. J., Bluhm, W. F., Chen, L., Deshpande, N., Feng, Z. *et al*. (2004). The distribution and query systems of the RCSB Protein Data Bank. *Nucl. Acids Res.* **32**, D223–D225.
54. Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**, 147–150.
55. Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J., Chothia, C. & Murzin, A. G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucl. Acids Res.* **32**, D226–D229.
56. Letunic, I., Copley, R. R., Schmidt, S., Ciccarelli, F. D., Doerks, T., Schultz, J. *et al*. (2004). SMART 4.0: towards genomic data integration. *Nucl. Acids Res.* **32**, D142–D144.
57. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zheng, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST—a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
58. Shindyalov, I. N. & Bourne, P. E. (1998). Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747.
59. Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucl. Acids Res.* **26**, 316–319.
60. Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallog. sect. D*, **52**, 842–857.
61. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
62. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
63. DeLano, W. L. (2002). The PyMOL Molecular Graphics System on World Wide Web. http://www.pymol.org.
64. Donato, R. (1999). Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type. *Biochim. Biophys. Acta*, **1450**, 191–231.
65. Roberts, W. M. (1993). Spatial calcium buffering in saccular hair cells. *Nature*, **363**, 74–76.
66. Vyas, M. N., Jacobson, B. L. & Quiocho, F. A. (1989). The calcium-binding site in the galactose chemoreceptor protein. Crystallographic and metal-binding studies. *J. Biol. Chem.* **264**, 20817–20821.
67. Gao-Sheridan, S., Zhang, S. & Collier, R. J. (2003). Exchange characteristics of calcium ions bound to anthrax protective antigen. *Biochem. Biophys. Res. Commun.* **300**, 61–64.
68. Lawler, J., Weinstein, R. & Hynes, R. O. (1988). Cell attachment to thrombospondin: the role of ARG-GLY-ASP, calcium, and integrin receptors. *J. Cell Biol.* **107**, 2351–2361.
69. Frandsen, G., Muller-Uri, F., Nielsen, M., Mundy, J. & Skriver, K. (1996). Novel plant Ca^{2+}-binding protein expressed in response to abscisic acid and osmotic stress. *J. Biol. Chem.* **271**, 343–348.

Edited by J. Thornton

(Received 30 June 2004; received in revised form 12 August 2004; accepted 25 August 2004)