Gen. Physiol. Biophys. (2009), **28**, Focus Issue, F82–F88

Structural diversity of calcium binding sites

Daniel Bindreither and Peter Lackner

Department of Molecular Biology, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria

Abstract. Calcium Binding Proteins (CBPs) play a major role in many biological processes. The three dimensional (3D) structure of several CBPs has been resolved by means of X-ray crystallography and nuclear magnetic resonance. We consulted several databases to compile a collection of CBPs of known 3D structure. The analysis of these data shows, the CBP structures are distributed over many different functional families and fold types. The binding site itself is less frequently formed by a continuous sequence segment. In the majority of the cases Ca^{2+} ion coordination is spread over different secondary structure elements with considerable distance on the amino acid sequence. The sidechain of amino acids Asp and Glu are the major interaction partner for the ion. Less frequently it is the side chain of Asn, Gln, Ser and Thr. Often main chain oxygen contributes to the Ca^{2+} coordination. In addition, water molecules are frequently involved.

Key words: Calcium binding — Structural motif

Introduction

Calcium is a key regulator in many biological processes. Binding and release of Ca^{2+} ions changes structural properties of the involved calcium binding proteins (CBPs) such that they switch their state regarding interaction with other protein or performing enzymatic activities (Ikura 1996). By means of X-ray crystallography or nuclear magnetic resonance the 3D structure of a considerable number of CBPs has been resolved. For a subset, Ca^{2+} bound and Ca^{2+} free forms have been determined to investigate the molecular impact of the Ca^{2+} ion.

Here we focus on the three dimensional structure of Ca^{2+} binding sites. We have been interested in the different types and topologies of the secondary structural elements which host the residues involved in Ca^{2+} binding. Three databases, Pfam (Finn et al. 2008), PDB (Berman et al. 2000) and SCOP (Murzin et al. 1995) have been consulted for this purpose. Pfam-A is a manually curated collection of protein families which are represented by multiple alignments and annotated with biological information. PDB is the central public repository for structurally resolved proteins. SCOP, the Structural Classification of Proteins is a hierarchical, expert knowledge supported system, which groups protein domains with respect to their tertiary structure and evolutionary and functional relationships.

Correspondence to: Peter Lackner, Department of Molecular Biology, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria
E-mail: Peter.Lackner@sbg.ac.at

Results

We consulted the databases Pfam (version 23.0), PDB and SCOP (version 1.73) to retrieve structurally resolved proteins associated with Ca^{2+} binding. Pfam has been used to generate a list of families with assigned keywords 'calcium binding' and a link to a PDB entry. PDB and SCOP was used to determine the respective Ca^{2+} binding domain and their type of fold. This way, redundancies and false crosslinks arising from multidomain proteins have been removed. A representative structure was selected for each CBP family and its Ca^{2+} binding site was investigated in detail. Loosely following the SCOP hierarchy, we classify the sites into alpha-helix, beta-sheet and mixed alpha/beta based motifs. In addition we consulted the PROSITE database (Hulo et al. 2008) for relevant sequence pattern.

Table 1 provides a summary of the different types of binding sites including a representative structure and its classification in Pfam and SCOP. It is evident that Ca^{2+} binding sites are found in all SCOP secondary structural classes (a,b,c,d, and g), and in different fold types therein. Indicated by the number of Pfam domain architectures, about half of the CBP motifs are found within single domain proteins representing an integral component of the protein domain

Table 1. Classification of structurally known CBPs

Motif	SCOP code	Scop domain	Pfam ID	Pfam Arch.	Figure
Phospholipase A2	a.133.1.2	d1bp2a_	PF00068	3	2C
EF-hand	a.39.1.5	d1up5b1	PF00036	277	1
Phospholipase C	a.39.1.7	d1djwa1	PF09279	217	4A
Annexin	a.65.1.1	d2rana_	PF00191	18	2A/B
Peroxidase	a.93.1.1	d1hsra_	PF00141	17	4D/E
Carbohydrate binding module	b.18.1.10	d1gmma_	PF03422	135	3D
Collagen-binding domain	b.23.2.1	d1nqda_	PF04151	131	3F
Legume lectin	b.29.1.1	d2bqpa_	PF00139	45	3E
C2-domain	b.7.1.1	d1djwa2	PF00168	217	3H
Pectate lyase	b.80.1.1	d1o88a_	PF00544	18	3A
Hemolysin-type Ca-binding repeat	b.80.7.1	d1kapp1	PF00353	246	3B/C
CutC	c.1.30.1	d1x7ib1	PF03932	2	5B
Leucin rich repeat	c.10.2.6	d1jl5a_	PF00560	539	
Pancreatic lipase	c.69.1.19	d1lpbb2	PF01477	48	4B
Severin	d.109.1.1	d1svya_	PF00626	34	4C
C-type lectin domain	d.169.1.1	d2msba_	PF00059	144	3G
Lactalbumin	d.2.1.2	d1b9oa_	PF00062	4	
Transglutaminase	d.3.1.4	d1ggub4	PF01841	55	5A
LDL repeat	g.12.1.1	d1ajja_	PF00057	203	4F
GLA domain	g.32.1.1	d1j34c_	PF00594	25	5C

Motifs are named according to the Pfam family or by the protein name. Pfam ID: Pfam database accession code. Pfam Arch.: Number of different domain architectures found for the CBP domain. A high numbers indicates that a domain is found in many different multi-domain proteins. SCOP code: indicates class, fold, superfamily and family. SCOP domain: domain identifier; character 2-5 correspond to the four letter PDB code, character 6 is the PDB chain identifier, character 7 the domain number. An underscore means the protein appears as single domain respective 3D structure.

function. The other half are part of multidomain proteins and combined with many other functional domains such as kinases, peptidases or receptors.

Below we describe the different motifs as well as possible repetitive arrangements in the respective CBP domain. Detailed molecular information was derived from PDBsum (Laskowski 2009). USCF Chimera (Pettersen et al. 2004) has been used to visualize and investigate the respective representative molecules and to prepare the corresponding figures.

Alpha-Helix Based Motifs

The EF-hand, a (E)helix-loop-(F)helix supersecondary structure motif, is presumably the best known Ca^{2+} binding site conformation (Finn and Forsen 1995). The helices are arranged in an orthogonal manner. The loop has usually a size of 12 residues (Fig. 1A). The Ca^{2+} ion is predominantly coordinated by residues located in the loop region and one additional residue in the F-Helix. PROSITE defines a sequence pattern which covers a single EF-hand (pattern PS00018). Numerous EF-hand containing proteins have been structurally resolved and could be associated with a typical sequence pattern. EF-hand motifs predominantly appear in an even number of copies. An EF-hand subdomain consists of two of such motifs (Fig. 1B), as found for example in calbindin, S100 proteins or polcalcin. Osteonectin contains one additional helix, parvalbumin contains two additional helices. Calmodulin-like domains represent a duplicated EF-hand subdomain and thus host four EF-hand motifs (Fig. 1C). The SCOP classification currently lists 23 functionally diverse calmodulin-like domains such as calmodulin, troponin C or calcium-dependent protein kinase sk5. SCOP also describes a penta-EF-hand protein family including the proteins calpain and sorcin.

Several single domain EF-hand proteins appear as multimeres. Dimers are found for S100 proteins or polcalcin Phl p 7. Polcalcin Che a 3 is a tetramer. The symmetry and repetitive nature of EF-hand subdomains occasionally leads to domain swapping (Carey et al. 2007). An EF-hand subdomain is built by a continuous region of the two different chains. This has been observed for example in polcalcins or S100 proteins (Fig. 1D/E) (Wopfner et al. 2007).

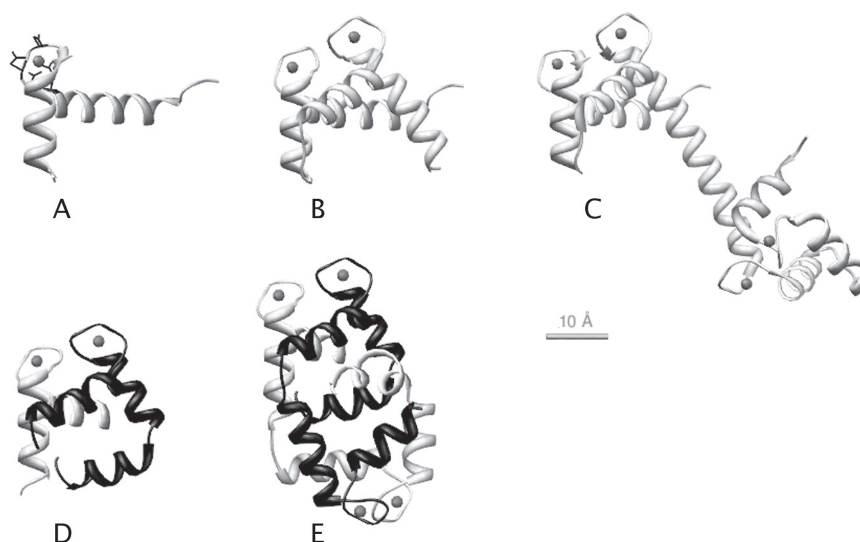

Figure 1. EF-hands and their arrangements in domains. (A) shows a single EF-hand. The Ca^{2+} interacting residues are highlighted in dark grey and the corresponding side chains are drawn sticks. The Ca^{2+} ion is the large grey sphere. Small spheres represent the oxygen of water molecules. (B) Two EF-hands build an EF-hand subdomain. (C) Calmodulin combines two EF-hand subdomains. (D) Domain swapping in Phl p 7: The two differnet chains are in light grey and dark grey respectively. Phl p 7 has two EF-hand subdomains, each composed of EF-hands which come from a different chain. In (E) the complete Phl p 7 is shown.

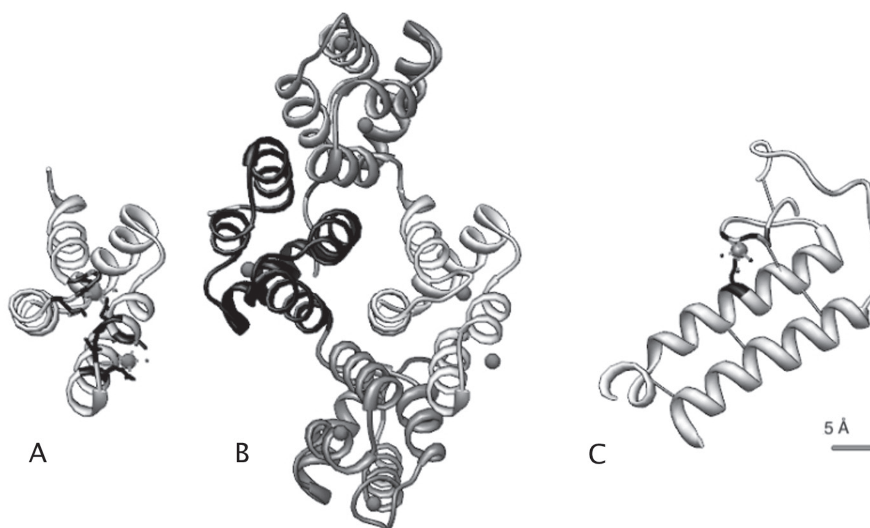

Figure 2. Other alpha helix based motifs. (B) Annexin is composed of four annexin repeats show in different shades of grey. (A) A single annexin repeat binds two Ca^{2+} ions. (C) Ca^{2+} ion binding in phospholipase A2. The complex loop and the helices are fixed by several disulphide bonds (indicated by solid lines).

A similar secondary structure pattern is the basis for the annexin repeat, which consists of a five helix folded leaf (Fig. 2A). The Ca^{2+} ions are bound in the loop region between the first and the second helix and the fourth and fifth helix. The Ca^{2+} binding site loop has about seven residues and therefore is shorter compared to the one in EF-hands, the helices are almost parallel. An annexin repeat is covered by the PROSITE pattern PS00223. The annexin core domain is made up of four repeats (Fig. 2B). In addition to the core domain animal and fungal annexins also have variable amino-terminal domains (Moss and Morgan 2004). Finally lactalbumin has a helix-loop-helix motif for Ca^{2+} binding

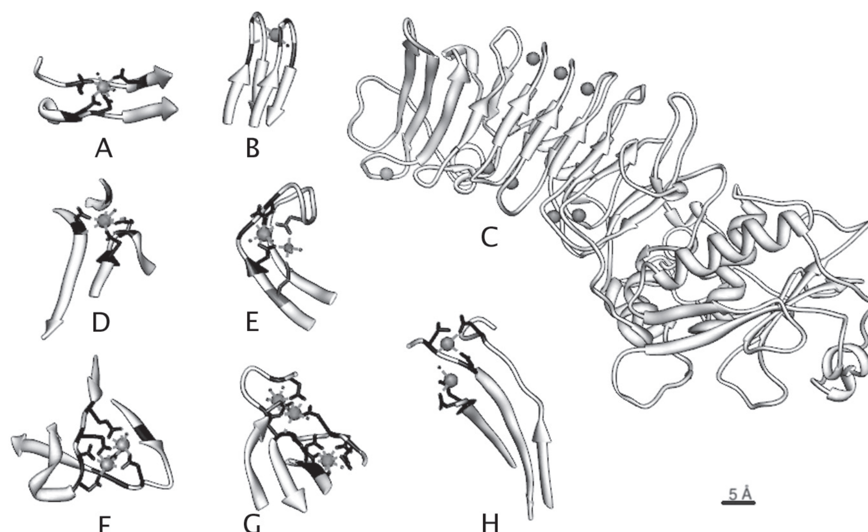

Figure 3. Beta sheet based motifs. (A) Motif from Pectate lyase C. (B) Hemolysin-type repeat of a metalloprotease. (C) Complete metalloprotease with several Ca^{2+} ions bound between consecutive loops. (D) Ca^{2+} binding site in Carbohydrate Binding Module CBM6. (E) In legume lectins there is Ca^{2+} site (larger sphere) near a Mn^{2+} site (middle sized spheres). Other motifs are from the collagen binding domain (F) with two ions bound, from C-lectin (G) which binds three Ca^{2+} ions and from the C2 domain (H).

(Acharya et al. 1989) with the associated PROSITE pattern PS00128.

In contrast to the continuous helix-loop-helix motif, phospholipase A2 and related proteins form a motif where a helix and a large complex loop is involved (Fig. 2C). A disulphide bond connects loop and helix, a second disulphide bond is found within the loop. The loop is rich of glycine residues such that the main chain can form tight turns. Several main chain oxygen atoms establish close contacts to the Ca^{2+} ion. The ion is required for activity in phospholipase A2 (Dennis 1994).

Beta-sheet based motifs

Beta-helix folds are formed by the association of beta strands into a helical topology. In SCOP, two Ca^{2+} binding protein families are found in this fold category. The pectate lyase C protein has a single binding site composed of four Asp/Glu residues located in neighbouring beta strands (Fig. 3A), and resembles the most simple beta-sheet based motif (Herron et al. 2003). In the metalloproteases the binding sites are mainly located between two loops of hemolysin-type calcium-binding repeats (Baumann et al. 1993). The loops themselves are glycine rich and thus rather tight, exposing main chain oxygen atoms (Fig. 3B). Due to highly repetitive nature of sequence and structure up to eight Ca^{2+} ions are bound by a single protein (Fig. 3C). A third class of highly repetitive beta domains is composed of leucin rich repeats (Bella et al. 2008). Several Ca^{2+} ions are bound in the loop regions by Asp and Asn residues.

More complex architecture of the Ca^{2+} binding site can be observed in other beta proteins. The Carbohydrate Binding Module CBM6 from xylanase (Czjzek et al. 2001), which belongs to galactose-binding domain-like fold class, comprise a binding motif, where three different strands and a loop are involved in Ca^{2+} coordination (Fig. 3D). Legume lectins (Sharon and Lis 1990) own a binding site, where two strands and a considerably large loop is involved (Fig. 3E). In addition a manganese ion is bound close to the Ca^{2+}.

In several other beta-sheet domains the Ca^{2+} binding site is located in a loop region. Two Ca^{2+} ions are coordinated by the collagen binding domain of class 1 collagenase (Wilson et al. 2003). The domain is described as CUB-like fold (Fig. 3F). The C2 domain is found in many proteins involved in signal transduction or membrane trafficking (Brose et al. 1995). It binds two Ca^{2+} ions in the loop region, although the topology is different compared to CUB-like folds (Fig. 3H). The C-lectin domain from the mannose-binding protein (Weis et al. 1992) binds three Ca^{2+} ions close to the carbonhydrate binding site (Fig. 3G).

Alpha/Beta based motifs

Phospholipase C (Rhee and Choi 1992) is composed of three domains. We already described the C2 domain in the previous section. The catalytic domain has a TIM-Barrel fold, which is found in many enzymes. In addition to the Ca^{2+} ions in the C2 domain there is also one Ca^{2+} ion located in the catalytic domain (Fig. 4A). It is in close proximity to the

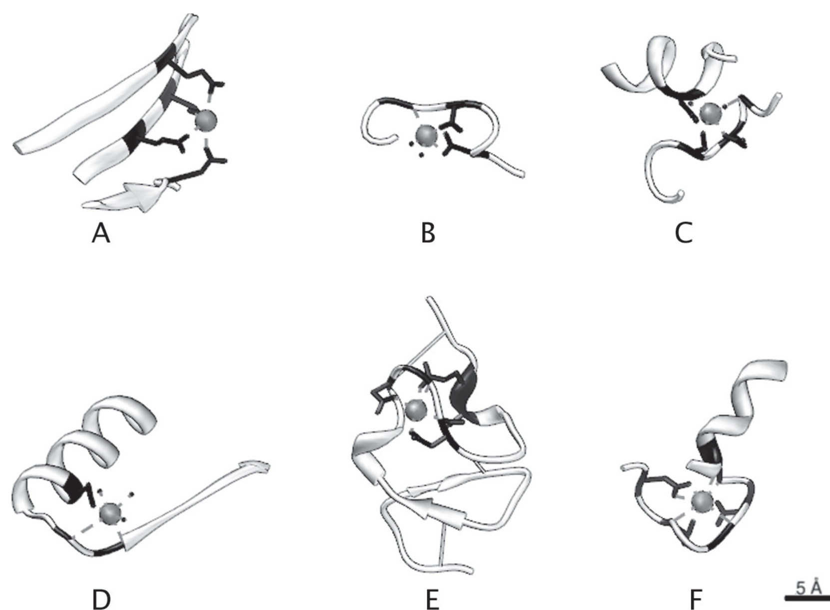

Figure 4. Alpha/beta based motifs. Ca^{2+} binding pattern in phospholipase C catalytic domain (A), N-terminal domain of pancreatic lipase (B), severin (C), heme dependent peroxidase (D and E). In (F) a complete LDL receptor repeat is shown. Disulphide bonds are indicated by solid lines.

catalytic site known to be involved in the reaction (Essen et al. 1997). Interestingly, also the third domain contains potential Ca^{2+} binding sites in form of four EF-hand motifs, although there are no Ca^{2+} ions bound in the X-ray structure. Both, the C2 and EF-hand domain possibly act as feedback regulator for the enzymatic domain.

The N-terminal domain of pancreatic lipase, a alpha/beta-hydrolase fold, contains one Ca^{2+} binding site in a loop region (Egloff et al. 1995) (Fig. 4B). In the severin domain 2 (Fig. 4C), classified in SCOP as gelsolin-like (three layer alpha/beta/alpha) fold, the Ca^{2+} binding site is located between a helix and strand, close to the N-terminal end of the strand (Puius et al. 2000). Heme dependent peroxidase from *Arthromyces ramosus* contains two Ca^{2+} binding sites in different buried regions of the protein (Fig. 4D/E). The structure of the two sites is rather dissimilar(Itakura et al. 1997).

The LDL receptor hosts six Ca^{2+} binding modules at the N-terminus (Fig. 4F). Each module has approximately 40 residues and forms three disulphide bonds. Several of these repeats have been investigated by X-ray(Fass et al. 1997) and NMR (Daly et al. 1995).

Out of the ordinary

In transglutaminase, only a single main chain oxygen from an alanine residues is involved in the Ca^{2+} coordination (Fig. 5A). Apart from that water molecules perform that task (Fox et al. 1999). Solely by main chain O atoms a Ca^{2+} ion is coordinated in the copper homeostasis protein CutC (Gupta et al. 1995) (Fig. 5B). Note that CutC has been structurally determined in a structural genomics project. Although annotated as copper ion binding, such a ion is not found in the different crystal structures.

Several coagulation factors are composed of two domains. The Gla domain is named after the modified amino acid gamma-carboxy-glutamat. The additional carboxy group enables high-affinity binding of calcium ions (Fig. 5C). The Gla fold has low secondary structure content. It undergoes an considerable conformational change on Ca^{2+} binding to form a compact fold, as show by NMR determination of the metal free form (Freedman et al. 1995). The C-type lectin-like domain also binds Ca^{2+}. The involded Glu residues are unmodified.

A different type of Ca^{2+} binding is found in the moderate affinity but high capacity Ca^{2+} binding of Calsequestrin. The protein consists of three identical domains which contain numerous acidic side chains. This yields to a highly negatively charged surface which can accommodate 40-50 Ca^{2+} ions. Binding of Ca^{2+} ions increases the secondary structure content of Calsequestrin (Wang et al. 1998).

Prediction of Ca^{2+} binding sites

It would be desirable to predict potential Ca^{2+} binding sites in proteins. As seen from the data above, the Ca^{2+} binding

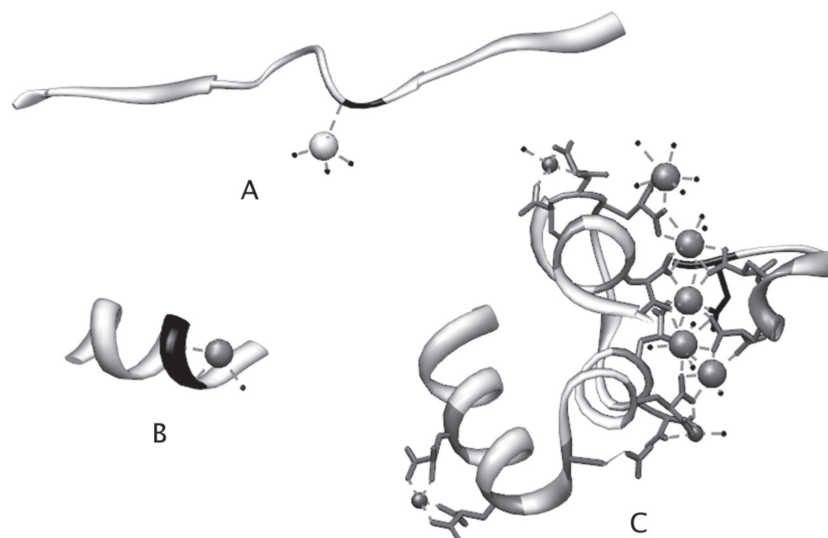

Figure 5. Unusual binding. Only coordinated by main chain oxygen is the Ca^{2+} in transglutaminase (A) and CutC (B). Modified Glu residues (gamma-carboxy-glutamate) bind the Ca^{2+} ion in Gla domains (C), the middle sized spheres are magnesium ion.

is rarely local in sequence. Furthermore, main chain oxygen atoms of different amino acids are involved which impedes to deduce strong sequence signals. Therefore sequence based methods hardly succeed. PROSITE defines at least a few pattern which match the Ca^{2+} binding region, e.g. for EF-hand and lactalbumin.

Therefore the 3D structure seems to be the better target for prediction methods. A few methods to predict potential Ca^{2+} binding sites in known 3D structures have been implemented. They are based on geometrical rules combined with atom types or charge distributions which describe potential binding cavities (Nayal and Di Cera 1994; Deng et al. 2006). These approaches should be applicable regardless the Ca^{2+} ion is bound by side chain or main chain oxygen. But however, this will also run into troubles if water molecules are involved.

Summary

Ca^{2+} ion binding sites are structurally diverse. It does not appear that a certain type or arrangement of secondary structure elements is preferred by nature. In several cases, a loop is involved. Especially if Ca^{2+} coordination is accomplished by the main chain oxygen atoms, glycine rich tight loops are incorporated. Water molecules may mediate Ca^{2+} coordination via H-bonds. Most of the binding sites are discontinuous regarding the involved residues in the proteins sequence. This complexity results in a challenge for the development of bioinformatics methods for binding site detection and prediction.

Acknowledgements. We thank O. Dissertori and C.Berbalk for the support in the manuscript preparation. This work was supported by the University of Salzburg.
The authors have no financial interests related to the material in the manuscript nor to the participation in the 2nd ECS Workshop.

References

Acharya K. R., Stuart D. I., Walker N. P., Lewis M., Phillips D.C. (1989): Refined structure of baboon alpha-lactalbumin at 1.7 A resolution. Comparison with C-type lysozyme. J. Mol. Biol. **208**, 99–127

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

Bella J., Hindle K. L., McEwan P. A., Lovell S. C. (2008): The leucine-rich repeat structure. Cell. Mol. Life Sci. **65,** 2307–2333

Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E. (2000): The protein data bank. Nucleic Acids Res. **28**, 235–242

Brose N., Hofmann K., Hata Y., Sudhof T. C. (1995): Mammalian homologues of *Caenorhabditis elegans* unc-13 gene define novel family of C2-domain proteins. J. Biol. Chem. **270**, 25273–25280

Carey J., Lindman S., Bauer M., Linse S. (2007): Protein reconstitution and three-dimensional domain swapping: benefits and constraints of covalency. Protein Sci. **16**, 2317–2333

Czjzek M., Bolam D. N., Mosbah A., Allouch J., Fontes C. M., Ferreira L. M., Bornet O., Zamboni V., Darbon H., Smith N. L., Black G. W., Henrissat B., Gilbert H. J. (2001): The

location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved. J. Biol. Chem. **276**, 48580–48587
Daly N. L., Scanlon M. J., Djordjevic J. T., Kroon P. A., Smith R. (1995): Three-dimensional structure of a cysteine-rich repeat from the low-density lipoprotein receptor. Proc. Natl. Acad. Sci. USA **92**, 6334–6338
Deng H., Chen G., Yang W., Yang J. J. (2006): Predicting calcium-binding sites in proteins - a graph theory and geometry approach. Proteins **64**, 34–42
Dennis E. A. (1994): Diversity of group types, regulation, and function of phospholipase A2. J. Biol. Chem. **269**, 13057–13060
Egloff M. P., Marguet F., Buono G., Verger R., Cambillau C., van Tilbeurgh H. (1995): The 2.46 A resolution structure of the pancreatic lipase-colipase complex inhibited by a C11 alkyl phosphonate. Biochemistry **34**, 2751–2762
Essen L. O., Perisic O., Katan M., Wu Y., Roberts M. F., Williams R. L. (1997): Structural mapping of the catalytic mechanism for a mammalian phosphoinositide-specific phospholipase C. Biochemistry **36**, 1704–1718
Fass D., Blacklow S., Kim P. S., Berger J. M. (1997) Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. Nature **388,** 691–693
Finn B. E., Forsen S. (1995): The evolving model of calmodulin structure, function and activation. Structure **3**, 7–11
Finn R. D., Tate J., Mistry J., Coggill P. C., Sammut S. J., Hotz H. R., Ceric G., Forslund K., Eddy S. R., Sonnhammer E. L., Bateman A. (2008): The Pfam protein families database. Nucleic Acids Res. **36,** D281–D288
Fox B. A., Yee V. C., Pedersen L. C., Le Trong I., Bishop P. D., Stenkamp R. E., Teller D. C. (1999): Identification of the calcium binding site and a novel ytterbium site in blood coagulation factor XIII by x-ray crystallography. J. Biol. Chem. **274**, 4917–4923
Freedman S. J., Furie B. C., Furie B., Baleja J. D. (1995): Structure of the metal-free gamma-carboxyglutamic acid-rich membrane binding region of factor IX by two-dimensional NMR spectroscopy. J. Biol. Chem. **270**, 7980–7987
Gupta S. D., Lee B. T., Camakaris J., Wu H. C. (1995): Identification of cutC and cutF (nlpE) genes involved in copper tolerance in *Escherichia coli*. J. Bacteriol. **177**, 4207–4215
Herron S. R., Scavetta R. D., Garrett M., Legner M., Jurnak F. (2003): Characterization and implications of Ca^{2+} binding to pectate lyase C. J. Biol. Chem. **278**, 12271–12277
Hulo N., Bairoch A., Bulliard V., Cerutti L., Cuche B. A., de Castro E., Lachaize C., Langendijk-Genevaux P. S., Sigrist C. J. (2008): The 20 years of PROSITE. Nucleic Acids Res. **36,** D245–D249
Ikura M. (1996): Calcium binding and conformational response in EF-hand proteins. Trends Biochem. Sci. **21**, 14–17
Itakura H., Oda Y., Fukuyama K. (1997): Binding mode of benzhydroxamic acid to *Arthromyces ramosus* peroxidase shown by X-ray crystallographic analysis of the complex at 1.6 A resolution. FEBS Lett. **412**, 107–110
Laskowski R. A. (2009): PDBsum new things. Nucleic Acids Res **37,** D355–D359
Moss S. E., Morgan R. O. (2004): The annexins. Genome Biol. **5**, 219
Murzin A. G., Brenner S. E., Hubbard T., Chothia C. (1995): SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247**, 536–540
Nayal M., Di Cera E. (1994): Predicting Ca^{2+}-binding sites in proteins. Proc. Natl. Acad. Sci. USA **91**, 817–821
Pettersen E. F., Goddard T. D., Huang C. C., Couch G. S., Greenblatt D. M., Meng E. C., Ferrin T. E. (2004): UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. **25**, 1605–1612
Puius Y. A., Fedorov E. V., Eichinger L., Schleicher M., Almo S. C. (2000): Mapping the functional surface of domain 2 in the gelsolin superfamily. Biochemistry **39**, 5322–5331
Rhee S. G., Choi K. D. (1992): Regulation of inositol phospholipid-specific phospholipase C isozymes. J. Biol. Chem. **267**, 12393–12396
Sharon N., Lis H. (1990): Legume lectins–a large family of homologous proteins. FASEB J. **4**, 3198–3208
Wang S., Trumble W. R., Liao H., Wesson C. R., Dunker A. K., Kang C. H. (1998): Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. Nat. Struct. Biol. **5**, 476–483
Weis W. I., Drickamer K., Hendrickson W. A. (1992): Structure of a C-type mannose-binding protein complexed with an oligosaccharide. Nature **360,** 127–134
Wilson J. J., Matsushita,O., Okabe A., Sakon J. (2003): A bacterial collagen-binding domain with novel calcium-binding motif controls domain orientation. EMBO J. **22**, 1743–1752
Wopfner N., Dissertori O., Ferreira F., Lackner P. (2007): Calcium-binding proteins and their role in allergic diseases. Immunol. Allergy Clin. North. Am. **27**, 29–44

Received: May 4, 2009
Final version accepted: June 18, 2009