J Biol Inorg Chem (2001) 6:479–489
DOI 10.1007/s007750100214

MINIREVIEW

Elna Pidcock · Geoffrey R. Moore

Structural characteristics of protein binding sites for calcium and lanthanide ions

Received: 30 August 2000 / Accepted: 19 January 2001 / Published online: 30 May 2001

Abstract Surveys of X-ray structures of Ca^{2+}-containing and lanthanide ion-containing proteins and coordination complexes have been performed and structural features of the metal binding sites compared. A total of 515 structures of Ca^{2+}-containing proteins were considered, although the final data set contained only 44 structures and 60 Ca^{2+} binding sites with a total of 323 ligands. Eighteen protein structures containing lanthanide ions were considered with a final data set containing eight structures and 11 metal binding sites. Structural features analysed include coordination numbers of the metal ions, the identity of their ligands, the denticity of carboxylate ligands, and the type of secondary structure from which the ligands are derived. Three general types of calcium binding site were identified in the final data set: class I sites supply the Ca^{2+} ligands from a continuous short sequence of amino acids; class II sites have one ligand supplied by a part of the amino acid sequence far removed from the main binding sequence; and class III sites are created by amino acids remote from one another in the sequence. The abundant EF-hand type of Ca^{2+} binding site was under-represented in the data set of structures analysed as far as its biological distribution is concerned, but was adequately represented for the chemical survey undertaken. A turn or loop structure was found to provide the bulk of the ligands to Ca^{2+}, but helix and sheet secondary structures are slightly better providers of bidentate carboxylate ligation than turn or loop structures. The average coordination number for Ca^{2+} was 6.0, though for EF-hand sites it is 7. The average coordination number of a lanthanide ion in an intrinsic protein Ca^{2+} site was 7.2, but for the adventitious sites was only 4.4. A survey of the Cambridge Structural Database showed there are small-molecule lanthanide complexes with low coordination numbers but it is likely that water molecules, which do not appear in the electron density maps, are present for some lanthanide sites in proteins. A detailed comparison of the well-defined Ca^{2+} and lanthanide ion binding sites suggests that a reduction of hydrogen bonding associated with the ligating residues of the binding sites containing lanthanide ions may be a response to the additional positive charge of the lanthanide ion. Major structural differences between Ca^{2+} binding sites with weak and strong binding affinities were not obvious, a consequence of long-range electrostatic interactions and metal ion-induced protein conformational changes modulating affinities.

Keywords Calcium ions · Lanthanide ions · Protein structures

E. Pidcock · G.R. Moore (✉)
School of Chemical Sciences, University of East Anglia,
Norwich, NR4 7TJ, UK
E-mail: g.moore@uea.ac.uk
Phone: +44-1603-592697
Fax: +44-1603-592697

Introduction

Since proteins that bind metal ions as part of their normal physiological function may account for as many as 40% of all proteins, understanding the relationships between the structures of proteins and their ability to bind metal ions is an important aspect of protein science [1, 2, 3, 4]. Ca^{2+} binding has a variety of functional roles in proteins, including: behaving as a structure-forming switching control, as in calmodulins and many other Ca^{2+}-sensor proteins [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]; enhancing protein stability [13]; and acting as an electrostatic control on the redox potential of free radical species in heme-containing redox proteins such as peroxidases [14]. To date, comparative structural analyses of Ca^{2+} binding to proteins have usually concentrated on structural features of the protein itself rather than the inner coordination sphere

of the Ca^{2+}. Notable exceptions to this are the studies of Einspahr and Bugg [6] and McPhalen et al. [7], which considered both protein structural features and details of the coordination geometry of the Ca^{2+} ion site; the review of Glusker [4], which dealt with the coordination chemistry of Ca^{2+} within the context of its binding to proteins; and the analysis of Katz et al. [15], which compared the binding of divalent metal ions in proteins and small molecules. This last work showed that the average coordination number of Ca^{2+} ions in proteins was 7, a value higher than for protein-bound Mg^{2+} ions even though there were generally less water molecules bound to Ca^{2+} than to Mg^{2+}. Katz et al. [15] suggested that this was because there was a tendency for Ca^{2+} to be bound in sites with less solvent exposure than Mg^{2+} binding sites. There has been a significant increase in the number of Ca^{2+}-containing protein structures determined: prior to 1994 there were 100 such structures but since 1994 there have been more than 515 additional structures. As a recent detailed analysis of the coordination chemistry of Ca^{2+} in proteins has not been reported, we have carried out a detailed survey of structures deposited in the Protein Databank since 1994 to determine the preferences for particular amino acids to bind to Ca^{2+}, and to investigate whether monodentate or bidentate coordination of carboxylate groups is more prevalent. Our study complements and extends the earlier ones [6, 7, 15], although there are significant differences between them stemming from differences in the way the protein structural data sets used were constructed. While our main data set had only one EF-hand site to avoid biasing general conclusions concerning the coordination chemistry of protein-bound Ca^{2+}, the earlier studies had a high proportion of EF-hand Ca^{2+} binding sites [5] in their data sets, reflecting the exceptional abundance of this binding motif. This is illustrated by the *Drosphila melanogaster* genome [16] in which the EF-hand family is the 12th most abundant protein domain.

One of the reasons for identifying the key chemical features of Ca^{2+} binding sites is that they are able to bind lanthanide ions, and there has been a resurgence of interest in the interaction between proteins and lanthanide ions, including the construction of lanthanide ion binding sites in proteins previously lacking that capability [17, 18] and the investigation of lanthanideion binding to transition metal ion binding sites (e.g. [19]). This activity stems from the use of lanthanide ions in spectroscopic and crystallographic studies of biomolecules. The crystallographic application revolves around the use of a lanthanide ion to determine the phases of the diffracted X-rays, either through the conventional isomorphous replacement procedure or through anomalous dispersion [20, 21, 22, 23, 24, 25, 26]. Spectroscopic applications include the use of lanthanide ions with specific magnetic properties to obtain information to characterize protein structures by NMR spectroscopy [27, 28, 29, 30, 31], and the use of luminescent lanthanide ions in bioanalytical assays [17, 32, 33, 34, 35]. Since these studies may be assisted by analyses of the coordination chemistry of lanthanide ions in proteins, we have carried out a detailed survey of structures in the PDB that bind lanthanide ions in well-formed sites.

The structures we have analysed do not contain the unusual amino acids sometimes found in Ca^{2+} binding proteins, namely γ-carboxyglutamic acid (Gla), which contains two side-chain carboxylic acid groups, β-hydroxyaspartate and β-hydroxyasparagine. More than 10 Gla residues are often found in the first 40 amino acids of proteins involved in blood clotting and bone structure, and they are involved in binding an array of Ca^{2+} ions (often >5). The two carboxylate groups of the Gla residue can donate one oxygen of each carboxylate group to a single Ca^{2+}, with the other carboxylic oxygen atoms able to interact with adjacent Ca^{2+} ions. Thus Gla is able to be part of a site binding multiple metal ions in the form of a "Ca^{2+} cluster" rather than acting as a tridentate or tetradentate ligand in a single metal binding site. The X-ray structure of blood coagulation factor VIIa illustrates one such "Ca^{2+} cluster" [36]. β-Hydroxyaspartate and β-hydroxyasparagine residues are often found in EGF-like protein domains, but their functional role is not clear since Sunnerhagen et al. [37] found from NMR studies that the β-hydroxyl group is not essential for high-affinity Ca^{2+} binding and is not a ligand for Ca^{2+}. Thus, our survey has concentrated on well-characterized proteins containing mononuclear Ca^{2+} and/or lanthanide ion binding sites.

Materials and methods

The Cambridge Structural Database [38] was searched using QUEST to build a molecular fragment. The Protein Databank [39] was searched using the PDB Browser (http://pdb-browsers.ebi.ac.uk) and the header files of the desired crystallographic structures were downloaded. A program was written in MATLAB which used the PDB header file as input and returned the amino acid counts used in the statistical analysis: total number of aminoacids; number of amino acids in helices and sheets; number of aspartate and glutamate residues present in helices and sheets. Ligands to the metal ions were classified as any group (for example, carbonyl oxygen or carboxylate oxygen) within 3 Å of the metal ion, visualized using RASMOL [40] or MOLMOL [41]. Diagrams of protein structures were constructed with MOLMOL.

NMR structures were excluded from the statistical analyses because the number of water molecules coordinated to Ca^{2+} is not generally obtained in such studies; andalthough the identities of the amino acid ligands are usually established (e.g. [37, 42]), their denticity may not be. Care was taken when selecting the data set to ensure that the statistics were not biased by over-representation of Ca^{2+} binding sites whose ligands were supplied by an amino acid sequence common with many other proteins in the data set, for example EF-hand sites or trypsin sites. PROSITE [43] was used to identify structures which contained sequences that coded over the calcium binding ligands (hence identifying similar calcium binding sites in non-homologous proteins) and PDBSum [44] was used in the identification of families of homologous proteins. The initial data set was reduced by

removing structures whose Ca^{2+} binding sites contained ligands which were shared between metal binding sites or ligands other than those donated by the protein and water. Information contained within the header of the PDB files was used to determine whether the Ca^{2+} binding ligands were localized in stretches of regular secondary structure. In the final data set, aspartate residues accounted for 6.5% of the total amino acids and glutamate for 4.4%. These frequencies are different from those obtained for a larger number of unrelated proteins (1021 structures) [45], which contained 5.3% aspartate and 6.2% glutamate residues. The frequencies from the present data set were used for calculations in this study.

McPhalen et al. [7] discuss the general procedure for crystallographic structure analyses of Ca^{2+}-containing proteins, particularly the fact that interatomic distances and angles are not usually restrained during structure refinement. This provides confidence for the kind of analysis reported in this study. However, similar confidence cannot be attributed to analyses of lanthanide ion binding sites in proteins where the lanthanide ion does not occupy an intrinsic Ca^{2+} binding site. This is because such sites are often not refined during the structure determination. To minimize the impact of poorly defined lanthanide ions present in adventitious sites for the purpose of determining the phases of the diffracted X-rays, only protein structures in which the lanthanide ion had two or more protein derived ligands were included in the data set; sites where the lanthanide ion was bound by exogenous ligands or was held on the surface of the protein by one ligand were ignored. Protein structures containing lanthanide ions and with multiple entries in the PDB were also removed from the final data set to reduce bias caused by common structures.

Results

Small-molecule complexes containing Ca^{2+} or lanthanide ions

A survey of calcium-containing and lanthanide-containing coordination complexes within the Cambridge Structural Database [38] was undertaken to establish a reference for protein binding sites. A search was conducted for Ca^{2+}, Pr^{3+}, Tb^{3+} and Yb^{3+} bonded to a specified number of oxygen ligands via an unspecified type of bond (Table 1). A total of 162 Ca^{2+} sites was found, 38% of which had a coordination number of 6, 30% a coordination number of 7 and 22% a coordination number of 9. The average coordination number was found to be 6.9. The complexes which exhibited the higher coordination numbers often had polydentate ligands with 3–6 oxygen atoms donated by a single macrocyclic ligand. The database was searched for complexes of Pr^{3+}, Tb^{3+} and Yb^{3+} containing oxygen donor ligands and with coordination numbers of 3–10 (Table 1). The average coordination numbers calculated for Pr^{3+}, Tb^{3+} and Yb^{3+} are 8.3 (91 complexes), 8.0 (46 complexes) and 6.9 (95 complexes), respectively. This trend of decreasing coordination number is a result of the contraction in size of the lanthanide ions [46, 47]. In contrast to the results obtained for Ca^{2+}-containing molecules, the majority of ligands of the high coordination number lanthanide complexes were not polydentate ligands but combinations of bidentate and tridentate ligands. A further difference between the data setsis found for the low coordination number complexes. Less than 1% of the Ca^{2+} data set is represented by complexes with a coordination number of three or four. For Pr^{3+}, Tb^{3+} and Yb^{3+}, 3–10% of the data sets are complexes with a coordination number of three or four. The ligands employed in such low coordination complexes are generally sterically bulky, for example trisubstituted *tert*-butyl phenolic complexes.

Table 1 Coordination numbers (CN) of small-molecule complexes containing Ca^{2+} or lanthanide ions

	Number of sites	CN 10	CN 9	CN 8	CN 7	CN 6	CN 5, 4, 3
Ca^{2+}	162	6	5	36	48	62	5
Pr^{3+}	91	18	28	26	9	5	5
Tb^{3+}	46	0	20	19	0	4	3
Yb^{3+}	95	4	12	31	8	17	23

Protein Ca^{2+} binding sites

Interrogation of the PDB for the years 1994–1999 yielded a list of 515 fully normalized crystal structures of Ca^{2+}-containing proteins with a resolution between 1.0 and 2.5 Å. Forty of the 44 structures in the final data set (Table 2) had a resolution of 2.3 Å or better, similar to the cut-off value chosen by McPhalen et al. [7]. Three general types of calcium binding site were identified in the final data set (Table 2): those which supply the Ca^{2+} ligands from a continuous short sequence of amino acids, labelled class I in Table 1; class II sites, which have one ligand supplied by a part of the amino acid sequence far removed from the main binding sequence; and the smallest category, class III, where all the ligands aresupplied by amino acids remote from one another in the sequence, as exemplified by adamalysin [48]. Similar classification schemes were employed by McPhalen et al. [7] and Einspahr and Bugg [6].

Representations of the structures of three of the proteins considered in the statistical analysis are shown in Fig. 1: the class I lectin (Fig. 1a and b) [49]; a class II site of a copper amine oxidase (Fig. 1c and d) [50]; and the class III site of adamalysin (Fig. 1e and f) [48]. All three proteins have global folds with a high content of secondary structure, and in all three the Ca^{2+} binding sites are located to one side of the protein, close to the molecular surface. The figures show that all of the Ca^{2+} binding sites are located in regions of the structure where some of the backbone forms loops or turns. The Ca^{2+} binding site of TolB is also shown in Fig. 1g and h. This remarkable structure, which was reported [51, 52] after our statistical analysis had been completed, contains a single Ca^{2+} ion within the central hole of aβ-propeller subunit. The Ca^{2+} ion is coordinated by only one amino acid side chain, a monodentate aspartate. There is suffi-

Table 2 List of calcium-containing protein structures obtained from the Protein Databank. Entries marked with an "a" indicate the bound Ca^{2+} has only one protein-derived ligand. I, II and III indicate sites where: the Ca^{2+} ligands are from a continuous short sequence of amino acids; one ligand is supplied by a part of the amino acid sequence far removed from the main binding sequence; all the ligands are supplied by amino acids remote from one another in the sequence, respectively

PDB code	Structure	Resolution (Å)	Number of sites	Type of site
2tep	Peanut lectin	2.5	1	I
1bf2	*Pseudomonas* isoamylase	2.0	1	II
1bfd	Benzoylformate decarboxylase	1.6	1	a
1bjr	Complex:lactoferrin fragment and proteinase K	2.4	1	II
1egz	Cellulase	2.3	1	II
1ce5	β-Trypsin	1.9	1	I
1b9z	β-Amylase	2.1	1	II
1b9o	Human α-lactalbumin	1.2	1	I
1fzc	Fibrin	2.3	2	I/I
1kvx	Carboxylic ester hydrolase	1.9	1	II
1ag9	Flavodoxin	1.8	2	I/I
1ai4	Penicillin acylase	2.4	1	II
1aqm	α-Amylase	1.9	1	III
1ax0	Lectin	1.9	1	I
1ayo	α-2-Macroglobulin	1.9	1	I
1gsp	Endoribonuclease	2.2	1	a
1hei	Hepatitis c virus RNA helicase domain	2.1	2	I/I
1irb	Carboxylic ester hydrolase	1.7	1	II
1jda	Maltotetraose forming *exo*-amylase	2.2	2	II/II
5chy	Signal transduction protein	2.0	1	II
1tcm	Cyclodextrin glycosyl transferase	2.2	1	III
1oil	Lipase	2.1	1	II
1obr	Carboxypeptidase T	2.3	2	II/I
1nbc	Cellusomal scaffolding protein A	1.8	1	II
1kit	Hydrolase	2.3	2	II/II
1js4	Endoexocellulase E4	2.0	2	II/II
1kbc	Neutrophil collagenase	1.8	2	II/II
1sbf	Soybean agglutinin	2.4	1	I
1tn3	Tetranectin	2.0	2	II/III
1vsi	Integrase	2.2	1	III
4lip	Lipase	1.8	1	II
2fib	Fibrogen	2.1	1	I
1ala	Annexin v	2.3	3	III/III/III
1atl	Atrolysin C	1.8	1	III
1clx	Xylanase	1.8	1	I
1mmq	Matrilysin	1.9	2	I/II
1oac	Amine oxidase	2.0	2	II/II
1sbh	Subtilisin	1.8	1	II
1sra	Calcium binding protein	2.0	2	I/I
1cel	Cellulase	1.8	1	III
1esl	E-Selectin	2.0	2	I/II
1gcg	Galactose binding protein	1.9	1	II
1hyt	Thermolysin	1.7	2	I/I
1iag	Adamalysin (II)	2.0	1	III

cient electron density for at least three bound water molecules; andfrom the remaining electron density and stereochemical arrangement of the Ca^{2+} ion and identifed ligands, Carr [52] suggests that there are actually five bound water molecules to create an octahedral binding site for the Ca^{2+} ion. Though the dissociation constant for the Ca^{2+} ion binding to TolB has not been reported, Carr [52] makes the point that since Ca^{2+} was not added to the purification or crystallization media, TolB must have a relatively high affinity for Ca^{2+} (the dissociation constant is probably $<10^{-9}$ M) to have acquired it from the *Escherichia coli* cytoplasm.

The 44 structures listed in Table 2 provide 60 calcium binding sites that consist of a total of 323 ligands (Table 3). Side-chain carboxylates of aspartate and glutamate residues form the major class of ligand, with water as the next most abundant. Glycine is the largest contributor to the total backbone carbonyl ligands, which represents 22% of all non-carboxylate, protein-derived ligands. Oxygen atoms from amide-containing side chains make up 9% of protein-derived ligation, with asparagine occurring nine times more frequently than glutamine. The average calcium coordination site is described by 1.4 waters, 1.9 carboxylate

Table 3 Coordinating ligands at the Ca^{2+} sites listed in Table 2

Total ligands	Water	Side-chain carboxylates	Side-chain amide	Backbone carbonyl	Side-chain hydroxyl
323	86	113 Asp/Glu 80/33	21 Asn/Gln 19/2	99 Gly/other 22/77	4 Ser/Thr 1/3

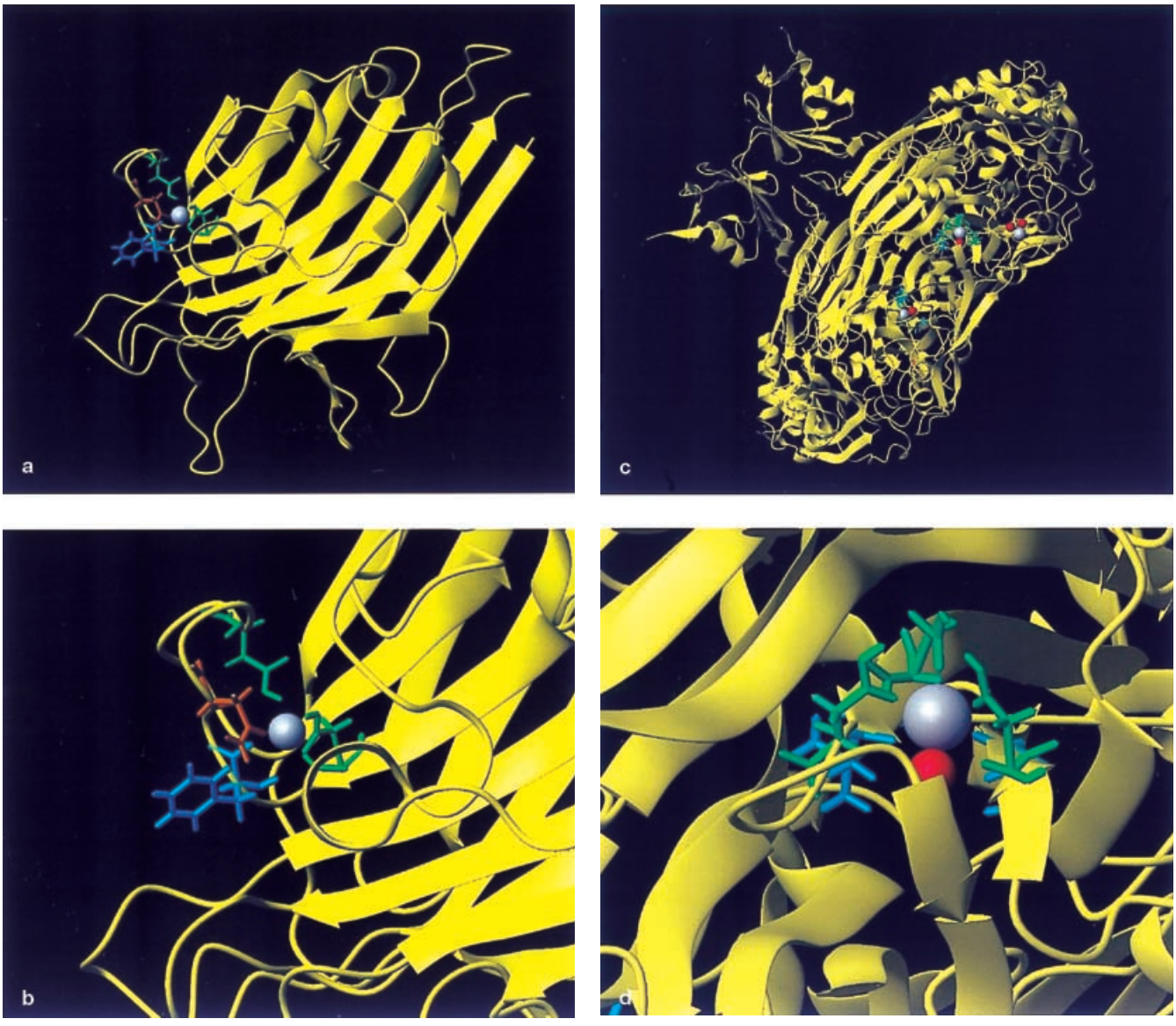
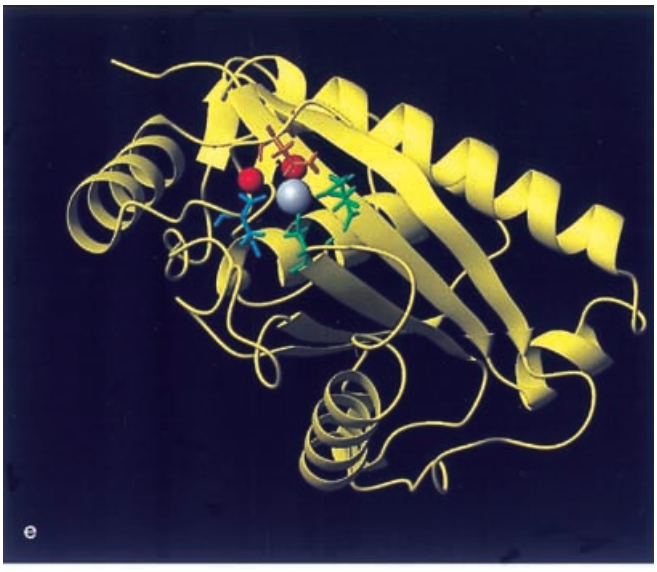


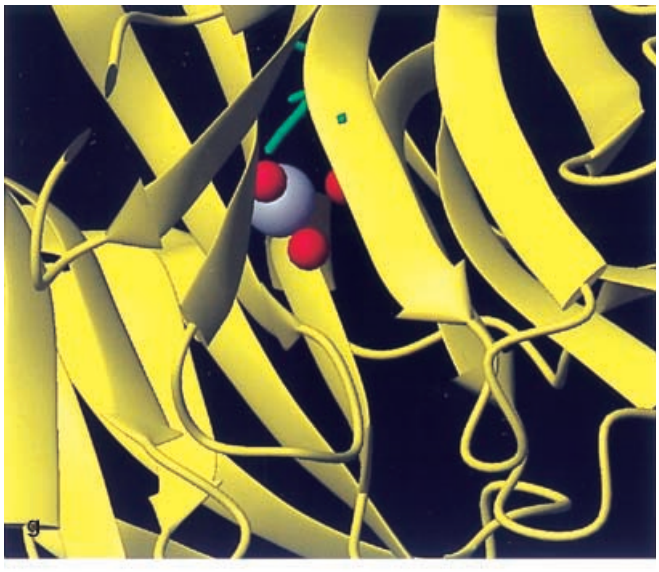
Fig. 1 Representations of the structures of (**a**, **b**) lectin (1ax0), (**c**, **d**) amine oxidase (10ac), (**e**, **f**) adamalysin (1iag), and (**g**, **h**) TolB ([51, 52], Carr S, Hemmings AM, personal communication), constructed with MOLMOL [41]. The figures of the class I, II and III sites show the overall fold of the protein (*top*) and a close-up view of the Ca^{2+} binding site. The figures for TolB show a close-up view of the Ca^{2+} binding site and a view of the entire molecule in which the molecular surface is shaded *grey*. The position of the Ca^{2+} ion in the central hole of the β-propeller domain can be seen in this figure. The colour coding is: *green* and *orange* for Asp/Glu and Asn/Gln residues whose side chains are ligands to Ca^{2+}; *cyan* for residues whose backbone carbonyls are ligands to Ca^{2+}; and, for **a**–**g**, *red* for the oxygen of water molecules bound to Ca^{2+}. The Ca^{2+} ions are shown in *grey* in **a**–**g** and in *red* in **h**. The ligands of the Ca^{2+} ions shown are: (**a**, **b**) lectin (1ax0): Asp129, Phe131 CO, Asn133, Asp136 and 2 H_2O molecules; (**c**, **d**) amine oxidase (10ac): Asp533, Leu534 CO, Asp535, Asp678, Ala679 CO and 1 H_2O molecule; (**e**, **f**) adamalysin (1iag): Glu9, Asp93, Cys197 CO, Asn200 and 1 H_2O molecule; (**g**, **h**) TolB: Asp337, Ala338 CO and 2–4 H_2O molecules

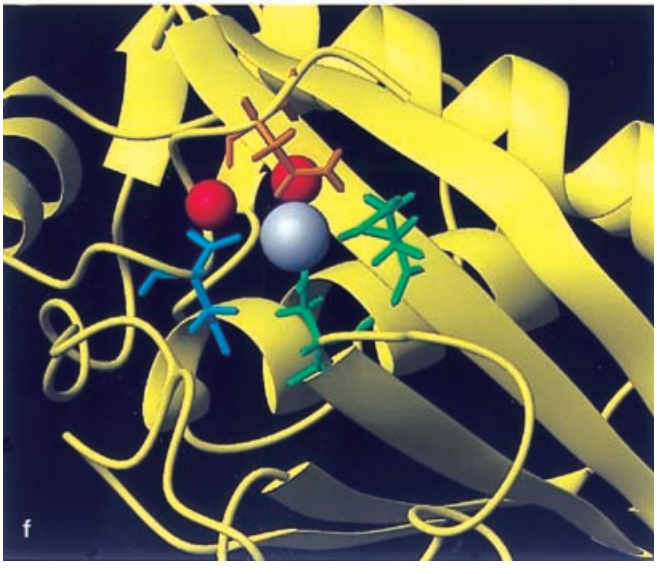
ligands (donating an average of 2.5 oxygens) and 2.1 oxygen atoms from carbonyl, amide or hydroxyl groups to give an overall, average coordination number for Ca^{2+} of 6.0.

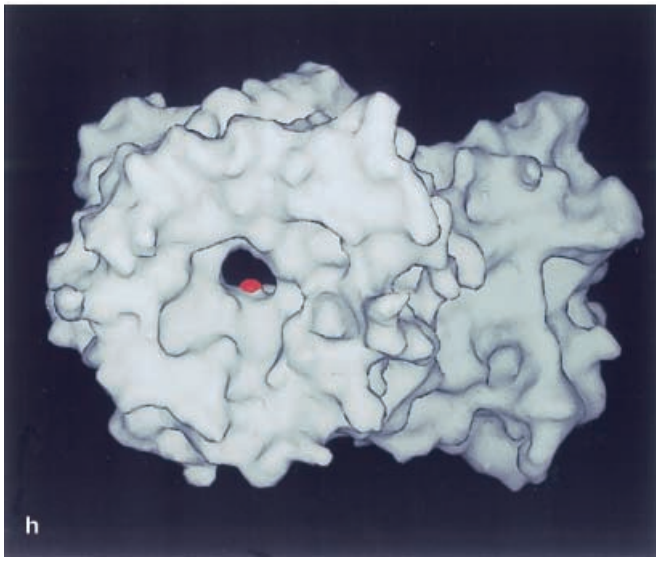
For 29 of the 60 Ca^{2+} sites the positive charge on the Ca^{2+} is balanced by the negative charges of its carboxylate ligands.In the remaining 31 cases the site is left with a residual positive (11 cases) or negative (19 cases) charge, after Ca^{2+} ion binding. Hydrogen bonding networks, which link the carboxylate oxygen atoms of calcium ligands and other polar residues located inthe calcium-binding loop to main-chain NH groups, may be involved in modulating the excess negative charge represented by the acidic residues [7, 9].

Of the Ca^{2+}-containing proteins studied (Table 2), 28% of the total of 13,598 amino acids were found in helices, 21% in sheets, leaving 51% in regions denoted as a turn/loop structure hereafter (Fig. 2). A total of 72.5% (172 of the 237) of the protein-derived Ca^{2+} lig-









ands (i.e. all ligands excluding water) are supplied by the turn/loop structure. Therefore the supply of ligands from the turn/loop structure is disproportionately large with respect to the occurrence of the turn/loop structure in the proteins of Table 2.

Of the 113 carboxylate ligands, 58% are supplied by turn/loop regions, 20% by helices and 22% by sheet structure, in reasonable agreement with the overall breakdown of secondary structure in the data set (Fig. 3). The ratio of aspartate to glutamate residues calculated from the 13,598 amino acids of the data set is 1.5:1. However, the observed ratio of aspartate to glutamate ligands is 2.4:1, a statistically significant deviation from the ratio expected on the basisof the population of aspartate and glutamate seen in our data set (i.e. 1.5:1). The expected distribution of aspartate and glutamate ligands through the different secondary structure types has been calculated using the values determined for the occurrence of these residues in the data set of Table 1, and their secondary structure propensities [45, 53]. For example, the expected number of ligands donated by turn/loop structures is 58 out of 113 (66 were observed), of which 39 are expected to be aspartates (50 were observed). Also, 64% of the aspartate and 72% of the glutamate ligands are monodentate. In terms of secondary structure, the ratio of monodentate to bidentate ligands in turn/loop structures is 2.5:1, helices 1.4:1 and sheets 1.5:1.

Protein lanthanide binding sites

Interrogation of the PDB yielded a list of 18 structures containing lanthanide ions, but after filtering the list as described in Materials and methods the final

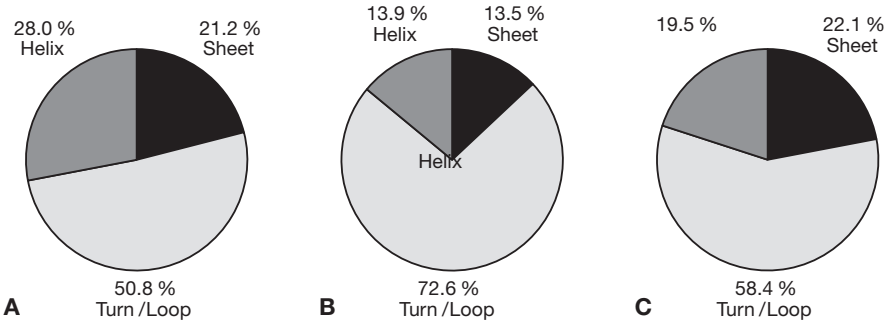


Fig. 2 Pie charts showing the secondary structure distribution of **A** all amino acids in the data set of Table 2; **B** all protein-derived ligands to Ca^{2+}; and **C** all carboxylate ligands to Ca^{2+} between the three secondary structural types

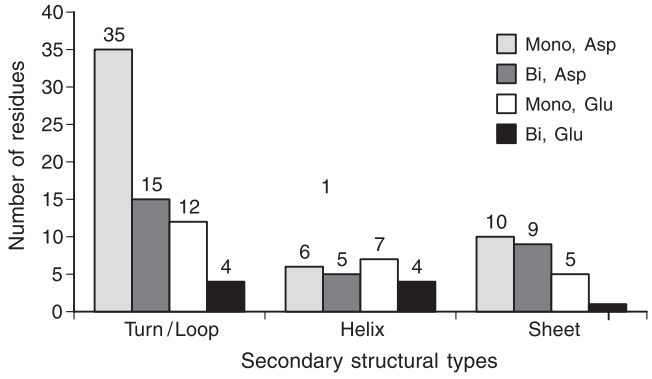


Fig. 3 Bar chart showing the distribution of carboxylate ligands to Ca^{2+} over the secondary structural types, obtained from proteins given in Table 2. Carboxylate ligands are divided on the basis of residue type and denticity of ligation

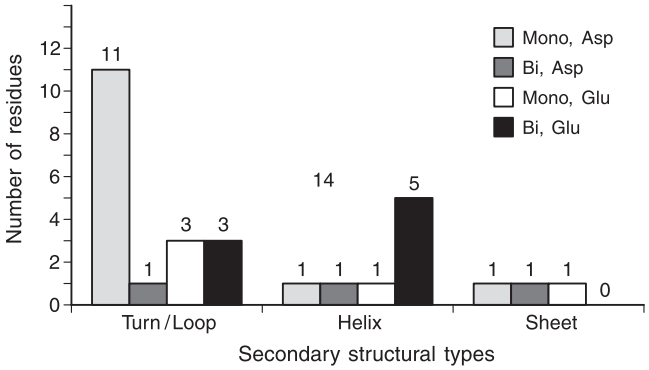


Fig. 4 Bar chart showing the distribution of carboxylate ligands to lanthanide ions over the secondary structural types, obtained from proteins in Table 4. Carboxylate ligands are divided on the basis of residue type and denticity of ligation

data set contained 8 structures and 11 metal binding sites (Table 4). In contrast to the data set for Ca^{2+} binding sites (Table 2), there is a high proportion of EF-hand sites in the lanthanide ion data set: 4 out of 11. The lanthanide data set (Table 4) also differs from the Ca^{2+} binding site data set (Table 2) in having a significant number of adventitious binding sites, 5 out of 11, with relatively few protein-derived ligands. The average lanthanide ion coordination site in this data set is described by 1.1 water ligands, 1.2 backbone carbonyl, amide or hydroxyl oxygen atoms and 2.6 carboxylate ligands supplying 3.6 carboxylate oxygen atoms to give an average coordination number of 5.9.

Of the 42 protein-derived ligands (excluding 12 water molecules) for the complete lanthanide ion data set (Table 4), 69% are carboxylate ligands. The majority of non-carboxylate ligands (9 out of 13) originate in turn/loop structures. The breakdown of the 29 carboxylate ligands which coordinate lanthanide ions in terms of secondary structural types is given in Fig. 4: 62% of carboxylate ligands are supplied by turn/loop structures and 38% from regular secondary structures. Similar to the data obtained for calcium sites, the single largest carboxylate oxygen donor is monodentate aspartate from turn/loop regions. The ratio of monodentate to bidentate ligands calculated from the 11lanthanide binding sites is 18:11 or 1.6:1.

Table 4 List of lanthanide-containing structures obtained from the Protein Databank. In column 5, A indicates the lanthanide ion site is adventitious, C that the lanthanide ion occupies an intrinsic Ca^{2+} binding site other than an EF-hand site, and E that the lanthanide ion binds to an EF-hand site

PDB code	Protein	Res. (Å)	Sites	Metal-substituted site
1a3c	Transcription regulation	1.6	1 Sm	A
1qat	Phospholipase C δ-1	3.0	2 Sm	A
2tcl	Fibroblast collagenase	2.2	1 Sm	C
1b9y	Phosducin complexed with transducin	3.0	1 Gd	A
1ncz	Troponin C	1.8	2 Tb	E
1nch	Cell adhesion protein	2.1	1 Yb	A
1ytt	Subtilisin fragment of mannose binding protein A	1.8	2 Yb	C
1psr	EF-hand protein	1.1	1 Ho	E

Discussion

Coordination numbers, types of ligands and secondary structure in protein Ca^{2+} binding sites

The average coordination number for the data set of Table 2 is 6.0, significantly lower than the 7 and 6.7 obtained in the analyses of Katz et al. [15] and McPhalen et al. [7], respectively. The difference is a result of the EF-hand class of protein, which contains seven-coordinate Ca^{2+} [5, 7, 8, 9, 12], being given less weight in our data set than in the earlier sets: there is only one EF-hand site in our set of 44 (Table 2) compared with 10 out of 27 for McPhalen et al. [7].

Helices and sheets provide relatively few of the ligands to Ca^{2+}, helices providing only 14% of all protein derived ligands and sheets 13.5%. Helices and sheets have well-defined structures and are not able to supply many ligands to a single site since only the side chain of every third or fourth amino acid in a helix will point in approximately the same direction (i.e. towards the calcium ion), while in a sheet the side chain of every other amino acid is orientated on the same side, but linearly spaced by approximately 6–7 Å. Conversely, the turn/loop structure is flexible and can readily supply three ligands from a sequence of five amino acids. Therefore the turn/loop structure provides the bulk of the ligands to Ca^{2+}.

Aspartate ligands are observed more frequently, and glutamate ligands less frequently, than expected. These results indicate that Ca^{2+} binding sites have a slight preference for aspartate residues. A contributing factor to this preference may be that aspartate residues are commonly involved in Asx turns, in which a side-chain oxygen of an aspartate (also of an asparagine, serine or threonine) in position *n* of a turn is hydrogen bonded to the main-chain NH of a residue at position *n*+2 [7, 54]. These hydrogen bonding networks are particularly prevalent in Ca^{2+} binding sites. Alternatively, a preference of Ca^{2+} binding sites for carboxylate ligands with a less bulky side chain may explain the greater than expected ratio of aspartate ligands to glutamate, consistent with asparagine being present to a greater degree than glutamine (Table 3).

Monodentate versus bidentate ligation in protein Ca^{2+} binding sites

In small complexes the denticity of a ligand is an important factor in determining binding affinities and coordination geometries. In proteins too the denticity of a ligand may be important. In order to determine if a particular residue or secondary structural type favours bidentate ligation, the expected ratio of monodentate to bidentate ligands has to be determined. The main consideration in this is the allowable combinations of monodentate and bidentate carboxylate ligands for a given coordination number. For example, the description of a Ca^{2+} binding site may include three carboxylate oxygen atoms, provided by either three monodentate carboxylates or by one monodentate and one bidentate carboxylate ligand. Therefore the expected ratio of monodentate to bidentate ligands observed for a sample of Ca^{2+} sites which contain three carboxylate oxygen atoms is 4:1. For four carboxylate oxygen atoms the ratio is 2:1. An analysis of this type for the data set given in Table 2 results in an expected ratio of monodentate to bidentate ligands of approximately 3:1, close to the observed values of 64% monodentate aspartates ligands and 72% monodentate glutamate ligands. Thus it appears that there is not a significant preference for aspartate or glutamate to bind in a monodentate or bidentate fashion.

The type of secondary structure donating the ligand seems to have more influence on the denticity of the ligand than the particular amino acid, with helix and sheet structures being slightly better providers of bidentate ligation than flexible turn/loop structures (Fig. 2). Though the data set is relatively small, this conclusion is consistent with the large number of structure studies of EF-hand [5, 6, 7, 12] and EGF-like [42, 55] motif binding sites, which show that a carefully placed ligand in a helix or as part of a sheet reliably provides bidentate coordination.

Lanthanide ion binding to proteins

The average coordination number for the data set of Table 4 is less than that calculated for the Ca^{2+} binding sites (Table 2), which is a reflection of the difference in composition of the data sets used. For the Ca^{2+} binding sites of Table 4 substituted by lanthanide ions the average coordination number is 7.2 and for the adventitious sites it is only 4.4. As described above, a search of the Cambridge Structural Database [38] established a precedent in small-molecule complexes for the low-coordination, adventitious lanthanide binding sites observed in the protein crystallographicdata set. However, three-coordinate complexes of lanthanide ions are susceptible to hydrolysis and therefore the presence of coordinated but disordered water molecules, which do not appear in the electron density maps for the low-coordinate lanthanide sites in proteins, cannot be discounted.

Bidentate ligation is present to a higher degree in the lanthanide ion binding sites (Table 4) than in Ca^{2+} sites (Table 2): of the 11 bidentate ligands, 7 (64%) are supplied by regular secondary structure, 5 by helices. The small size and nature of the data set precludes strong conclusions, but in agreement with the analysis of Ca^{2+} binding sites it appears that regular secondary structure may be a better provider of bidentate ligands than the turn/loop structure, which pro-

vides the majority of monodentate carboxylate ligation for the lanthanide ions.

In some cases, lanthanide ions bind to sites in proteins that do not bind Ca^{2+} strongly. *Bacillus subtilus* PyrR [56] is a good example. This is normally a hexameric protein but in the presence of lanthanide ions it becomes a dimer because the lanthanide ion binds to groups that would otherwise form the intersubunit contacts at the three-fold axis of the hexamer. In this site, Sm^{3+} is octahedrally coordinated to six side-chain carboxylates. Binding of Yb^{3+} to the cadherin NCD1 [19] is another example. In this structure, Yb^{3+} binds to two carboxylate side chains, both of which are bidentate, with two coordinated water molecules also visible. The Yb^{3+} has a high mobility, which presumably results from its attachment to only two amino acids. Lanthanide ions also bind at sites that are normally occupied by two divalent metal ions, thoughoften only one lanthanide ion will bind to such a site [20, 24]. The active site of the Klenow fragment of DNA polymerase I is a good example [57]. Here a single Eu^{3+} ion binds to carboxylate groups at the dinuclear metal ion binding site and occupies a position in between those taken by two divalent metal ions.

A common feature of the analysis above is an uncertainty about the precise coordination number of bound lanthanide ions, despite good quality X-ray structures. For some X-ray structures there is also an uncertainty concerning the location of the lanthanide ion, which may be relatively mobile compared to other parts of the structure. Both of these uncertainties are related to the fact thatthe inner coordination sphere of lanthanide ions can change rapidly, as Williams [47] noted for complexes of lanthanide ions with dipicolinic acid: the bound ligands interchanged position rapidly on the surface of the lanthanide ion andexchanged rapidly with ligands not bound to it. Even polydentate ligands bound to lanthanide ions fluctuate rapidly so that the co-existence of multiple isomeric forms is common (e.g. see [58, 59]).

Comparison of corresponding lanthanide ion and calcium bound structures

Where good quality crystal structures exist for both lanthanide ion and Ca^{2+} bound forms of a protein, a detailed comparison of their metal binding sites is revealing, as Matthews and Weaver [20] demonstrated for lanthanide ions binding to three Ca^{2+} binding sites in thermolysin. The data set for our study comprises three proteins: troponin C (1TOP, calcium; 1NCZ, terbium) [60, 61], mannose binding protein A (lectin domain) (2MSB, calcium; 1MSB, holmium) [21, 55] and parvalbumin (1B8R, calcium; 1PSR, holmium) [12, 25]. In all cases, neither the coordination number of the metal ion calculated from the PDB coordinates nor the ligating residues change between the structures with bound Ca^{2+} or lanthanide ions. These sites are carboxylate-rich, with at least three carboxylate ligands each, and hence no additional ligands are required to compensate for the additional charge of the trivalent lanthanide ion. Even though lanthanide ions favour high coordination numbers, no change in the denticity of the carboxylate ligands is induced and no extra water molecules are included in the lanthanide binding sites as defined by the X-ray structures. However, it is possible that the coordination number of the bound lanthanide ion is greater than that of Ca^{2+} in the same site. Horrocks [33] has used a luminescence spectroscopy procedure for determining the number of water molecules coordinated to Eu^{3+} and Tb^{3+} in proteins. With Eu^{3+} bound in the four calcium sites of troponin C, luminescence studies [62] show that two water molecules are bound per ion instead of the one seen in X-ray structures of the Ca^{2+}-containing troponin C. In general, Horrocks [33] suggests that lanthanide ions generally have a coordination number one greater than that of Ca^{2+} in well-formed protein sites.

Small differences between the Ca^{2+} and lanthanide sites were revealed by a survey of possible hydrogen bond donor groups to the carboxylate ligands of the metal ions. A backbone nitrogen atom within the range 2.7–3.2 Å of the coordinating carboxylate oxygen and with a C-O···N angle within the range 106–144° was considered to be a hydrogen bond donor [54]. Similar analyses were undertaken for hydrogen bond donors such as water, serine and threonine OH groups, and asparagine and glutamine NH_2 groups, using the appropriate distance and angle parameters given by Baker and Hubbard [54]. For the three pairs of Ca^{2+} and lanthanide ion binding sites considered, the lanthanide ion sites contained fewer groups capable of acting as donors of hydrogen bonds than the Ca^{2+} sites, as determined with the above criteria. Since hydrogen bonding networks may reduce the negative charge provided by the acidic side chains of the binding site [7, 9], a reduction of hydrogen bonding associated with the sites containing lanthanide ions may be a response to their additional positive charge. Matthews and Weaver [20] also noted small structural differences between thermolysin with Ca^{2+} and different lanthanide ions bound and they suggested that a shift in the position of some of the lanthanide ions occurred as a result of their difference in size.

Binding affinities for calcium and lanthanide ions to proteins

In their early study, Einspahr and Bugg [6] noted that there were no dramatic structural differences between calcium binding sites with weak and strong binding affinities, and that remains the case with the much greater database of structures. In particular, neither the coordination number of bound Ca^{2+} nor the

number of ligands supplied by the protein are correlated with binding affinity. This is shown by the presence of Ca^{2+} in the TolB structure (Fig. 1g, h) [51, 52] and by the variation in Ca^{2+} dissociation constants for a range of seven-coordinate EF-hand-containing proteins, which vary from 10^{-4} to 10^{-9} M [11]. This is because, in addition to inner-sphere effects such as those arising from the number and type of ligands coordinating the bound metal ions, there are two other general phenomena that will affect binding affinities: electrostatic interactions and metal ion-induced protein conformational changes. As statedearlier, less than half of the Ca^{2+}-containing binding sites in our data set are formally charge-balanced by the acidic carboxylate ligands, and of the remainder some have an excess of formal positive charge and some an excess of formal negative charge. There is a parallel here with the structural factors that control metalloprotein reduction potentials, which can be considered to be the energy changes associated with binding of an electron (see for example [63]). In either the reduced or oxidized state of most such proteins an additional charge buried within the protein has to be stabilized. Where it is an additional positive charge which is satisfactorily stabilized the reduction potential is generally low, but where anadditional positive charge is not well compensated the reduction potential is high. The compensating factors that affect the reduction potential include hydrogen bonding interactions at the metal site, electrostatic interactions with both buried and surface charges of the protein, some up to 20 Å away from the metal ion, and the electrostatic effects of protein and solvent dipoles [64]. Similar factors will influence the stability of Ca^{2+} and lanthanide ions in protein sites so that they do not need to be formally charge-balanced with the inner-sphere ligands alone. Conformational changes induced by the binding of Ca^{2+} ions can also have a major effect on apparent binding affinities, with large induced conformational changes utilizing some of the binding energy released by the Ca^{2+}-protein interaction and thereby reducing the observed binding affinity, as has been recognized by many authors (e.g. [7, 8, 9, 10, 11]). Because of these electrostatic and conformational phenomena, it is not possible to simply relate the structures listed in Tables 1 and 2 with reported binding affinities. However, EF-hand sites of proteins which are structured in their metal-free states have been extensively investigated by a combined protein engineering and metal ion binding approach, which has allowed structural features governing the binding affinities of these particular sites to be determined [9, 11, 65, 66, 67, 68].

Lanthanide ions bind more strongly to some carboxylate-rich sites than Ca^{2+} ions. For example, the single Ca^{2+} ion EF-hand site of the *E. coli* receptor for D-galactose and D-glucose (GGR) binds a Ca^{2+} ion with a dissociation constant (K_d) of 2.5×10^{-5} M and lanthanide ions such as Yb^{3+} and Lu^{3+} with K_d values of 0.8×10^{-6} M and 1.1×10^{-6} M, respectively [65]. Presumably this is a response to the higher positive charge of the lanthanide ions.

Choice of data sets and evolutionary selection of EF-hand sites

The protein sites chosen for this study were selected to inform on features of the coordination chemistry of protein-bound Ca^{2+} rather than evolutionary aspects of protein folds. For this reason we included no more than one representative of each type of homologous protein site in the final data set, unlike all other earlier surveys of proteins containing Ca^{2+}. Nevertheless, it is clear that the EF-hand type of binding site has a special significance as this is the most widespread type of protein Ca^{2+} binding motif in nature, as illustrated by its being the most abundant Ca^{2+} binding motif encoded by the *Drosophila* genome [16]. However, the natural selection of the EF-hand motif probably does not result from an especially high binding affinity for Ca^{2+} coordination since different proteins containing EF-hand sites have very different Ca^{2+} binding affinities [11]. The observation that the non-EF-hand binding sites in our data set (Table 2) have a similar Ca^{2+} coordination environment to Ca^{2+} in the EF-hand sites, albeit with a coordination number that is generally less than that of the EF-hand site, emphasizes that the EF-hand sites have been evolutionarily selected for other reasons.

Acknowledgements We thank the EPSRC and BBSRC for support via their Biomolecular Sciences Panel, the EU for support (contract no. BIO4-98-0156) and Kate Brown (London), Stephen Carr, Toni Georgiou, Andrew Hemmings, Colin Kleanthous (Norwich), Annie Powell (Karlsruhe), Zoe Pikramenou (Birmingham) and Ivano Bertini and Claudio Luchinat (Florence) for helpful discussions. We also thank Ruth Boetzel for constructing the diagrams in Fig. 1.

References

1. Holm RH, Solomon EI (eds) (1996) Chem Rev 96:2237–3042
2. Da Silva JJRF, Williams RJP (1991) The biological chemistry of the elements. Clarendon Press, Oxford
3. Lippard SJ, Berg JM (1994) Principles of bioinorganic chemistry. University Science Books, Sausalito, Calif
4. Glusker JP (1991) Adv Protein Chem 42:1–76
5. Kretsinger RH, Kockolds CE (1973) J Biol Chem 248:3313–3326
6. Einspahr H, Bugg CE (1984) In: Sigel H (ed) Metal ions in biological systems, vol 17. Dekker, Basel, pp 51–71
7. McPhalen CA, Strynadka NCJ, James MNG (1991) Adv Protien Chem 42:77–144
8. da Silva ACR, Reinach FC (1991) Trends Biochem Sci 16:53–57
9. Falke JJ, Drake SK, Hazard AL, Peersen OB (1994) Q Rev Biophys 27:219–290
10. Skelton NJ, Kördel J, Akke M, Forsén S, Chazin WJ (1994) Nat Struct Biol 1:239–245
11. Ikura M (1996) Trends Biochem Sci 21:14–17

12. Cates MS, Berry MB, Ho EL, Li Q, Potter JD, Phillips GN Jr (1999) Structure 7:1269–1278
13. Kuroki R, Taniyama Y, Seko C, Nakamura H, Kikuchi M, Ikehara M (1989) Proc Natl Acad Sci USA 86:6903–6907
14. Bonagura CA, Sundaramoorthy M, Pappa HS, Patterson WR, Poulos TL (1996) Biochemistry 35:6107–6115
15. Katz AK, Glusker JP, Beebe SA, Bock CW (1996) J Am Chem Soc 118:5752–5763
16. Rubin GM, Yandell MD, Wortmann JR, Miklos GLG, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LSB, Gong F, Guan P, Harris NL, Hay BA, Hoskins RA, Li J, Li Z, Hynes RO, Jones SJM, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng ZH, Zhong F, Zhong, W, Gibbs R, Venter JC, Adams MD, Lewis S (2000) Science 287:2204–2215
17. Clark ID, MacManus JP, Banville D, Szabo AG (1993) Anal Biochem 210:1–6
18. MacKenzie CR, Clark ID, Evans SV, Hill IE, MacManus JP, Dubuc G, Bundle DR, Narang SA, Young NM, Szabo AG (1995) Immunotechnology 1:139–150
19. Moore JD, Skinner MA, Swatman DR, Hawkins AR, Brown, KA (1998) J Am Chem Soc 120:7105–7106
20. Matthews BW, Weaver LH (1974) Biochemistry 13:1719–1725
21. Weis WI, Kahn R, Fourme R, Drickamer K, Hendrickson WA (1991) Science 254:1608–1615
22. Shapiro L, Fannon AM, Kwong PD, Thompson A, Lehmann MS, Grübel G, Legrand J-F, Als-Nielsen J, Colman DR, Hendrickson WA (1995) Nature 374:327–337
23. Burling FT, Weis WI, Flaherty KM, Brünger AT (1996) Science 271:72–77
24. Grobler JA, Essen L-O, Williams RL, Hurley JH (1996) Nat Struct Biol 3:788–795
25. Brodersen DE, Etzerodt M, Madsen P, Celis JE, Thøgersen HC, Nyborg J, Kjeldgaard M (1998) Structure 6:477–489
26. Boggon TJ, Shapiro L (2000) Structure 8:R143–R149
27. Dobson CM, Williams RJP (1977) In: Pullman H, Goldblum N (eds) Metal-ligand interactions in organic chemistry and biochemistry, vol 1. Reidel, Dordrecht, pp 255–282
28. Lee L, Sykes BD (1983) Biochemistry 22:4366–4373
29. Bentrop D, Bertini I, Cremonini MA, Forsén S, Luchinat C, Malmendal A (1997) Biochemistry 36:11605–11618
30. Biekofsky RR, Muskett FW, Schmidt JM, Martin SR, Browne JP, Bayley PM, Feeney J (1999) FEBS Lett 460:519–526
31. Allegrozi M, Bertini I, Janik MBL, Lee Y-M, Liu G, Luchinat C (2000) J Am Chem Soc 122:4154–4161
32. Yuan J, Matsumoto K (1998) Anal Chem 70:596–601
33. Horrocks WD Jr (1993) Methods Enzymol 226:495–538
34. Sabbatini N, Guardigli M, Lehn J-M (1993) Coord Chem Rev 123:201–228
35. Hertzberg RP, Pope AJ (2000) Curr Opin Chem Biol 4:445–451
36. Banner DW, D'Arcy A, Chène C, Winkler FK, Guha A, Konigsberg WH, Nemerson Y, Kirchhofer D (1996) Nature 380:41–46
37. Sunnerhagen MS, Persson E, Dahlqvist I, Drakenberg T, Stenflo J, Mayhew M, Robin M, Handford P, Tilley JW, Campbell ID, Brownlee GG (1993) J Biol Chem 268:23339–23344
38. Allen FH, Kennard O (1993) Chem Des Automation News 8:31–37
39. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112:535–542
40. Sayle R, Milner-White EJ (1995) Trends Biochem Sci 20:374–376
41. Koradi R, Billeter M, Wüthrich K (1996) J Mol Graphics 14:51–59
42. Handford PA, Mayhew M, Baron M, Winship PR, Campbell ID, Brownlee GG (1991) Nature 351:164–167
43. Hofmann K, Bucher P, Falquet L, Bairoch A (1999) Nucleic Acids Res 27:215–219
44. Laskowski RA, Hutchinson EG, Michie AD, Wallace AC, Jones ML, Thornton JM (1997) Trends Biochem Sci 22:488–490
45. Creighton TE (1993) Proteins. Freeman, New York, pp 255–257
46. Huheey JE, Keiter EA, Keiter RL (1993) Inorganic chemistry, 4th edn. HarperCollins, New York, pp 599–613
47. Williams RJP (1982) Struct Bonding 50:79–119
48. Gomis-Ruth F-X, Kress LF, Bode W (1993) EMBO J 12:4151–4157
49. Elgavish S, Shaanan B (1988) J Mol Biol 277:917–932
50. Parsons MR, Convery MA, Wilmot CM, Yadav KDS, Blakely V, Corner AS, Phillips SEV, McPherson MJ, Knowles PF (1995) Structure 3:1171–1184
51. Carr S, Penfold CN, Bamford V, James R, Hemmings AM (2000) Structure 8:57–66
52. Carr S (2000) PhD thesis, University of East Anglia, Norwich, UK
53. Shulz GE, Schirmer RH (1979) Principles of protein structure. Springer, Berlin Heidelberg New York, p 109
54. Baker EN, Hubbard RE (1984) Prog Biophys Mol Biol 44:97–179
55. Weis WI, Drickamer K, Hendrickson WA (1992) Nature 360:127–134
56. Tomchick DR, Turner RJ, Switzer RL, Smith JL (1998) Structure 6:337–350
57. Brautigan CA, Aschheim K, Steitz TA (1999) Chem Biol 6:901–908
58. Geraldes CFGC, Urbano AM, Hoefnagel MA, Peters JA (1993) Inorg Chem 32:2426–2432
59. Lammers H, Maton F, Pubanz D, van Laren MW, van Bekkum H, Merbach AE, Muller RN, Peters JA (1997) Inorg Chem 36:2527–2538
60. Satyshov KA, Pyzalska D, Greaser M, Rao ST, Sundaralingam M (1994) Acta Crystallogr Sect D 50:40–49
61. Rao ST, Satyshov KA, Greaser ML, Sundaralingam M (1996) Acta Crystallogr Sect D 52:916–922
62. Bruno J, Zauharm RJ, Horrocks WD Jr (1992) Biochemistry 31:7016–7026
63. Moore GR, Pettigrew GW, Rogers NK, Williams G (1986) In: Xavier AV (ed) Frontiers in bioinorganic chemistry. VCH, Weinheim, pp 494–506
64. Mauk AG, Moore GR (1997) JBIC 2:119–125
65. Snyder EE, Buoscio BW, Falke JJ (1990) Biochemistry 29:3937–3943
66. Falke JJ, Snyder EE, Thatcher KC, Voertler CS (1991) Biochemistry 30:8690–8697
67. Henzl MT, Hapak RC, Goodpasture EA (1996) Biochemistry 35:5856–5869
68. Drake SK, Zimmer MA, Kundrot C, Falke JJ (1997) J Gen Physiol 110:173–184