PROTEINS: Structure, Function, and Genetics 47:344–356 (2002)

# Structural Analysis, Identification, and Design of Calcium-Binding Sites in Proteins

**Wei Yang,[1] Hsiau-Wei Lee,[2] Homme Hellinga,[3] and Jenny J. Yang[2]***
[1]*Department of Biology Drug Design, Georgia State University, Atlanta, Georgia*
[2]*Department of Chemistry, Center of Drug Design, Georgia State University, Atlanta, Georgia*
[3]*Medical Center, Department of Biochemistry, Duke University, Durham, North Carolina*

***ABSTRACT*** **Assigning proteins with functions based on the 3-D structure requires high-speed techniques to make a systematic survey of protein structures. Calcium regulates many biological systems by binding numerous proteins in different biological environments. Despite the great diversity in the composition of ligand residues and bond angles and lengths of calcium-binding sites, our structural analysis of 11 calcium-binding sites in different classes of proteins has shown that common local structural parameters can be used to identify and design calcium-binding proteins. Natural calcium-binding sites in both EF-hand proteins and non-EF-hand proteins can be described with the smallest deviation from the geometry of an ideal pentagonal bipyramid. Further, two different magnesium-binding sites in parvalbumin and calbindin$_{D9K}$ can also be identified using an octahedral geometry. Using the established method, we have designed *de novo* calcium-binding sites into the scaffold of non-calcium-binding proteins CD2 and Rop. Our results suggest that it is possible to identify calcium- and magnesium-binding sites in proteins and design de novo metal-binding sites. Proteins 2002;47:344–356.**


**Key words: EF-hand motif; calcium-binding proteins; magnesium binding; de novo design; computer algorithm**

## INTRODUCTION

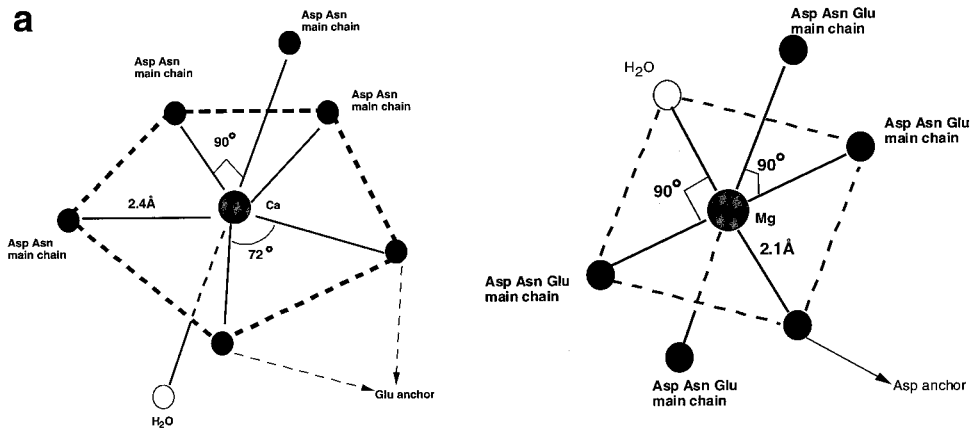
One important problem in current biology and chemistry is the ability to assign proteins with functions based on the 3-D structure. To catalog the common ways of protein folding and active site formation, structural genomics applies high-speed techniques to make a systematic survey of protein structures.[1,2] Although comparative analysis of primary sequences provides some important information, it would be useful to probe functions by chemical potentials based on the structure if the sequence is not conserved. This problem is directly related to the inverse protein design.[3] In this article, we demonstrate the ability to identify calcium- and magnesium-binding sites in proteins using an automated tool.

Calcium regulates many biological systems by interacting with proteins with different affinities in different biological environments.[4,5] The binding of calcium to proteins leads to an increase in stability and changes in conformation of the calcium-binding proteins. According to structural features of calcium-binding sites, proteins can be classified as EF-hand or non-EF-hand. The EF-hand calcium-binding proteins have more than 500 entries in the Protein Data Bank. They control many cellular processes and calcium levels in the cell.[6,7] Over 60 subfamilies have been identified.[8] These include parvalbumin (Parv), troponin C (TnC), calmodulin (CaM), sarcoplasmic calcium-binding protein, the essential and regulatory light chains of myosin, calbindin$_{D9K}$ (CBD), and the S100 and VIS subfamilies.[8] Although the EF-hand subfamilies vary greatly in sequence and in number of EF-domains, the canonical calcium-binding site (loop) remains narrowly defined. All of the EF-hand motifs that we know consist of a highly conserved loop flanked by two helices (helix–loop–helix)[9]; these can be further divided into classic EF-hand motifs and pseudo-EF-hand motifs (Fig. 1). Classic and pseudo-EF-hand motifs contain 12 and 14 residues in the calcium-binding loop, respectively (Fig. 1). For example, calmodulin contains four classic EF-hand calcium-binding sites. Calbindin$_{D9k}$ and S100A12 have a classic EF-hand motif in site 2 and a pseudo-EF-hand motif in site 1 of the protein. Parvalbumin has three classic EF-hand calcium-binding motifs, in which one motif lacks the calcium-binding ligands in the central loop.[10,11] EF-hand proteins not only have strong calcium affinities [about $10^{-6}$–$10^{-7}$ M ($K_d$)], but also display a strong selectivity (relative affinity $K_{Ca^{2+}}/K_{Mg^{2+}}$) for calcium over magnesium of up to $10^4$ because the intracellular $Mg^{2+}$concentration is about $10^4$-fold higher than that of $Ca^{2+}$ ($10^{-3}$ M for $Mg^{2+}$ and $10^{-7}$ M for $Ca^{2+}$ in the resting eukaryotic cell).[12] In contrast to the marked structural transitions induced by $Ca^{2+}$ binding, $Mg^{2+}$ binding causes only localized conformational changes within the four $Ca^{2+}$-binding loops of CaM.[13–15] For example, when magnesium binds to the EF-hand loop of parvalbumin it uses the same residues as the ligands used in calcium binding.[16,17] The unidentate

*Abbreviations*: CaM, calmodulin; TnC, troponin C; CBD, calbindin$_{D9k}$; Parv, parvalbumin; α-lac, α-lactalbumin; Ther, thermitase.

*Correspondence to: Jenny J. Yang, Department of Chemistry, Center of Drug Design, Georgia State University, Atlanta, GA 30303. E-mail: chejjy@panther.gsu.edu

Received 4 June 2001; Accepted 5 December 2001




**b**

Classic EF-hand

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| s | | s | | s | | m-c | | $H_2O$ | | | b-s |
| D (99%) | | D (76%)<br>N (23%) | | D (52%)<br>S (23%)<br>N (21%) | | T (23%)<br>F (16%)<br>K (12%) | | | | | E (92%) |

Pseudo EF-hand

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m-c | | | m-c | | m-c | | | m-c | | | | | b-s<br>E |

Fig. 1. (**a**). Calcium-binding site with the pentagonal bipyramidal geometry (left) and the magnesium-binding site with the octahedral geometry (right). The oxygen atom from water molecules as a calcium-binding ligand is shown as an empty circle. (**b**). Schematic diagrams of classic (top) and pseudo-EF-hand (bottom) calcium-binding motifs and their conserved ligand residues. S, side chain; b-s, bidentate side chain; and m-c, mainchain oxygen atoms.

ligand of Glu for magnesium at position 12 of the EF-loop is changed into a bidentate ligand for calcium, which results in the conversion of an octahedral geometry for magnesium into the pentagonal bipyramidal geometry for calcium (Fig.1). It is interesting to note that, in EF-hand proteins, calcium/magnesium exchange appears to be closely related to physiological processes that involve cell excitation and relaxation, such as muscle contraction.[6,11] Non-EF-hand proteins, such as cell adhesion molecules, receptors, and transmembrane proteins, often have Ca(II)-binding sites located at the linker regions between domains or at the exposed loop and turn regions between β-strands[18–20] and do not have conserved calcium-binding loops and flanked helices.[21]

Understanding the molecular basis of diseases caused by the overloading of calcium, the disruption of calcium-binding sites of proteins, and the mechanism of calcium-modulated signal transduction requires the establishment of the principles for calcium-binding affinity and calcium over magnesium selectivity of calcium-binding proteins.[8,22,23] To date, more than 100 structures of calcium-binding proteins have been solved. However, the process by which calcium-binding proteins specifically achieve their affinities for calcium is not clear despite the enormous efforts that have been devoted to the understanding and prediction of calcium-binding affinity and selectivity of proteins. As pointed out in many extensive reviews, calcium-binding sites in proteins are highly irregular with great variation in ligand type, length, and angle of the metal coordination shell. The prediction of calcium-binding sites at the residue level is challenging and has yet to be achieved.[24–26] In this article, we report our structural analysis of 11 calcium-binding sites in different classes of proteins with the aim of establishing structural parameters for identifying and designing calcium-binding proteins. In this work, we first use four calcium-binding sites in calmodulin as a training set to obtain geometric descriptions of calcium-binding sites. We then demonstrate that both classic and pseudo-EF-hand calcium-binding sites in three EF-hand proteins (calmodulin, parvalbumin, and calbindin$_{D9K}$) can be described and reidentified with the smallest deviation from the ideal pentagonal bipyramidal geometry. In addition, calcium-binding sites in non-EF-hand proteins, such as thermitase

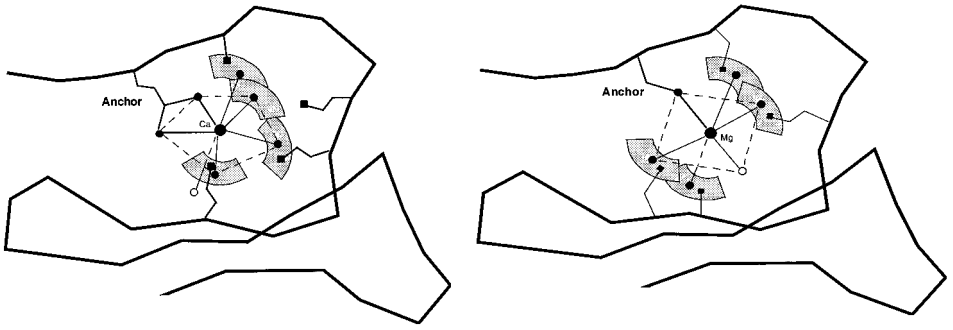



Fig. 2. Strategies for searching calcium- (left) and magnesium-binding sites (right). The first residue (bidentate Glu with the attachment of $Ca^{2+}$) located in the calculation (called anchor) defines the relative position of the calcium atom to the protein backbone and is used as a starting point to construct a calcium-binding site. Asp with the attachment of $Mg^{2+}$ is used as the anchor to search for the magnesium-binding site. After attaching the anchor residue to the backbone of the protein along the protein sequence, the metal-binding geometry or positions of other ligands are then defined around the anchor. A metal-binding site is constructed with the residue combination when all the oxygen ligand atoms satisfy the requirements. If any ligand atoms do not satisfy the geometric requirement, the residue combination is rejected and the program moves to the next anchor position in the primary sequence to test for the possibility of constructing the next metal-binding site.

and α-lactalbumin, can be identified despite their lack of homology in protein sequence and structure. We further demonstrate that two different magnesium-binding sites in metal-loaded forms of parvalbumin and calbindin$_{D9k}$ can also be described and identified using an octahedral geometry. In contrast, no calcium-binding sites were found in non-calcium-binding proteins with all natural residues. Finally, the design of novel calcium-binding sites has been carried out in non-calcium-binding proteins.

## MATERIALS AND METHODS

### Construction of Metal-Binding Sites

A geometric description of the ligands around a metal, the 3-D structure of the backbone of a protein, and a library of side-chain rotamers of amino acids (or atoms from the main-chain) were input into the Dezymer algorithm to identify a set of potential metal-binding sites using an SGI computer ($O_2$).[20,27]. Four calcium-binding sites in the high-resolution X-ray structure of calmodulin (3CLN) were used to establish the structural parameters of the calcium-binding site. Three EF-hand proteins S100A12 (1E8A),[28] parvalbumin (5CPV),[29]and calbindin$_{D9k}$ (4ICB)[30] containing both classic and pseudo-EF-hand calcium-binding sites, and two non-EF-hand proteins, α-lactalbumin (1ALC)[31] and thermitase (1THM),[32] were used to evaluate the structural parameters for describing calcium-binding sites. The natural magnesium-binding site (site 2) of calbindin$_{D9k}$ (5ICB)[33] was used for establishing geometric parameters of magnesium-binding sites in proteins. The crystal structure of the parvalbumin–magnesium complex (4PAL)[16,17] is then used to evaluate the structural parameters for magnesium-binding sites. All of the heteroatoms in these structural files, including metal ions and water, were deleted from pdb files. For a protein with about 100 amino acids, it takes about 2 h for the site-search step.

For the design of the calcium-binding sites in both classic and pseudo-EF-hand motifs, a bidentate ligand, Glu residue, was used as an anchor, which is the first residue placed on the protein frame. A calcium atom is placed on the same plane as two of the side-chain oxygen atoms and the Cδ atom with a distance of 2.4 Å between the calcium and oxygen atoms. The angle of O—Ca—O and torsion of Ca—O—Cδ—Cγ are set to 53.8 and 93.5°, respectively. This anchor residue (Glu—Ca) shares the same rotamer library of the Glu side-chain with 24 configurations.[27] For non-EF-hand calcium-binding sites, an anchor residue was generated by attaching a calcium atom to the Asp with an O—Ca distance of 2.4 Å and using a Ca—Oδ—Cγ angle of 139° and a Cγ—Oδ1—Oδ2—Ca torsion angle of 168°. For magnesium-binding sites, a pseudoresidue, aspartate with the attachment of a magnesium atom, was used as the anchor. The magnesium atom is placed 2.1 Å away from the side-chain oxygen atom of aspartate with a Mg—Oδ—Cγ angle of 141° and a Mg—Oδ—Cγ—Cβ dihedral angle of 62.5°.

After attaching the anchor residue to the backbone of the protein, the calcium-binding geometry or positions of other ligands are then defined around the anchor (Fig. 2). A pentagonal bipyramidal geometry, with five oxygen atoms on the same plane and the other two above and below the plane, is used to describe all calcium-binding sites in the natural calcium-binding proteins. The two oxygen atoms from the anchor residue Glu are placed on the plane. The other four ligand positions are further defined according to the relative position of each ligand residue to the anchor residue using oxygen atoms from the side-chain oxygens of Asp and Asn and the main-chain oxygen from the protein backbone. The distance between calcium and ligand oxygen is set to 1.5–3.5 Å for all of the ligands. The ranges for the angles of Oϵ—Ca—O (anchor Oϵ, Ca, and ligand O) are set to 30–150, 60–180, and 60–180° for the ligands in the same plane of the two side-chain oxygen atoms of the anchor because the ideal values for a pentagonal model are 72, 144, and 144°, respectively, with the dihedral angles of atom

Cδ—Oϵ—Ca—O (anchor Cδ, Oϵ, Ca, and ligand O) set to 0–30°. For the ligand out of the pentagonal plane, both these angles and the dihedral angles are set to 60–120°, as the ideal values are both 90°. For pseudo-EF-hand calcium-binding sites, (site 1) in calbindin$_{D9k}$ (4ICB)[30] and S100A12 (1E8A),[28] the oxygen—calcium distance is still set to 1.5–3.5 Å, while all of the angles and dihedral angles are set to −180–+180°. Except for the anchor Glu—Ca pseudoresidue in this calculation, only main-chain carbonyl oxygen atoms are used as ligands. For non-EF-hand calcium-binding sites, four non-anchor ligands from the protein were defined in the site-search step using a distance range of 1.5–3.5 Å and an angle range of 120°. For magnesium-binding sites, an octahedral geometry was used to define the magnesium-binding site (Fig. 1). The distance between the magnesium and the ligand oxygen is restricted to 1.0–3.0 Å for all four ligands. The ranges for angles of O—Mg—O are set to 30Å–140° because the ideal value for an octahedral geometry is 90°. The other angles and dihedral angles are not constrained. The remaining parameters for magnesium are identical to those for the EF-hand calcium-binding sites.

### Rank of Metal-Binding Sites

The configurations generated from the previous calculation are minimized using the Polak–Ribiere nonlinear conjugate gradient algorithm[34] with eq. (1).[27]

$$U(P_i) = \omega_L \sum_{j=1}^{N_L} \underbrace{(I/\sigma_{Lj}^2)(l_{ij} - L_j)^2}_{\text{(bond lengths)}} + \omega_\Omega \sum_{j=1}^{N_\Omega} \underbrace{(I/\sigma_{\Omega j}^2)(\omega_{ij} - \Omega_j)^2}_{\text{(bond angles)}} + \omega_\theta \sum_{j=1}^{N_\theta} \underbrace{(I/\sigma_{\theta j}^2)(\theta_{ij} - \Theta_j)^2}_{\text{(dihedral angles)}} + \omega_S \sum_{j=1}^{N_S} \sum_{k=1}^{3} \underbrace{(I/\sigma_{Sj}^2)(\chi_{ijk} - X_{jk})^2}_{\text{(shift constraints)}} + \omega_{\text{vdw}} \sum_{j=1}^{N_A} \sum_{k=1\,k\neq j}^{N_A} \times \left\{ \underbrace{\begin{array}{c} d_{jk} < (r_j + r_k),\ (I/\sigma_{\text{vdw}}^2)(d_{jk} - (r_j + r_k))^2 \\ d_{jk} > (r_j + r_k),\ 0 \end{array}}_{\text{(nonbonded contacts)}} \right\} \quad (1)$$

$P_i$ is the test configuration $i$ and $\omega_L$, $\omega_\Omega$, $\omega_{\Theta\mu}$, $\omega_\sigma$ , and $\omega_{\text{vdw}}$ are weights for bond lengths 1…$N_L$, bond angles l…$N_\Omega$, position shift constraints of amino acid backbone atoms 1…$N_s$, and van der Waals' contact between all atom pairs 1…$N_A$ in the test configuration. $\sigma_X$ is the standard deviation for parameter type $X$; $l_j$ and $L_j$ are the measured and target $j$th bond lengths, respectively. Similar definitions are given for bond angles $\omega_j$ and $\Omega_j$, dihedrals $\theta_j$ and $\Theta_j$, coordinates $x_j$ and $X_j$, and the distance $d_{jk}$ between nonbonded atoms $j$ and $k$ with the atomic hard-sphere radii $r_i$ and $r_j$, respectively.[27] The $L$ value used for all the calcium-binding sites is 2.4 Å and for all the magnesium-binding sites is 2.1 Å. The values of Ω and θ are the values from the ideal pentagonal bipyramidal or octahedral geometry. In the refinement step, the maximal cycles used were 250. For all the protein studies here, the average cycles used are 126.5. An average CPU time of 57.2 s is required to minimize one site.

## RESULTS AND DISCUSSION

### Common Features of Calcium-Binding Sites in Proteins

High-resolution X-ray structures of four EF-hand proteins, calmodulin, calbindin$_{D9k}$, S100A12, and parvalbumin, and two non-EF-hand proteins, α-lactalbumin and thermitase, were used to identify the common structural features of calcium- and magnesium-binding proteins. The bound forms of both classic and pseudo-EF-hand loops share a similar spatial arrangement despite their extremely diverse cellular functions and different response to calcium binding. For a classic (canonical) EF-hand motif, seven oxygen atoms from the side-chains of Asp, Asn, and Glu, the main chain, and water at the loop sequence positions of 1, 3, 5, 7, 9, and 12 coordinate the calcium ion in a pentagonal bipyramidal arrangement. In a typical geometry, position 1 of the calcium-binding loop has been shown to be Asp and the side-chain of Asp serves as a ligand on the $x$-axis. The $-x$-axis (position 9) is filled with a bridged water molecule connecting the side-chains of Asp, Ser, or Asn.[9,35] Axis $-z$ is shared by the two carboxyl oxygen atoms of a glutamate side-chain at position 12 that binds in a bidentate mode to Ca(II). Glu is used predominantly (92%) as a bidentate ligand for both classic and pseudo-EF-hand motifs in all intracellular calcium-binding proteins.[6,7,35] As shown in Figure 1, for the calcium-binding loop of a pseudo-EF-hand motif, four oxygen atoms from the main-chain carbonyl groups at sequence positions 1, 4, 6, and 9 provide ligands for the calcium ion with coordination geometry similar to that of a classic EF-hand motif.[30,36] As listed in Tables I and II, different types of oxygen ligands of natural calcium-binding sites in calmodulin, parvalbumin, and calbindin$_{D9k}$ have similar Ca—O distances with an average value of 2.4 ± 0.4 Å. On the other hand, the angles of O—Ca—O differ largely in the natural calcium-binding sites (Table I). The oxygen atoms out of the plane from the side-chain ligand (Asp or Asn) at sequence position 1 have O—Ca—O angles between 74 and 116° for natural calcium-binding sites.

For non-EF-hand proteins, the calcium-binding sites consist of fragments from several stretches of the protein sequence. Different numbers of oxygen from water molecules are used as ligand atoms. For example, site 2 of thermitase contains five ligand atoms from the carboxyl oxygens of the side-chains of Asp62 (bidentate), Asp57, and Gln66 and from the carbonyl oxygen atoms on the main-chain of Thr64. The ligand residue Asp57 is located on the β-strand rather than in the loop or helical regions as observed in EF-hand proteins. Although α-lactalbumin contains a continuous calcium-binding loop, it differs from an EF-hand motif in several respects. First, its calcium-binding loop is flanked by an α-helix and a $3^{10}$ helix instead of two EF helices. In addition, oxygens from two water molecules are used as calcium ligands. The other five ligands are from the side-chains of Asp82, 87, and 88 and from the main-chains of Lys79 and Asp84. While calcium-binding ligand residues are not conserved, Table II shows that the local geometry of non-EF-hand calcium-

**TABLE I. Ca—O Lengths and O—Ca—Oε Angles of the Natural and Reidentified Calcium-Binding Sites in CaM**

| Protein | Site | Ca—O lengths (Å) | | | | | | O—Ca—Oε angles (°) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CaM | 1 | E31/Oε1 | E31/Oε2 | D20/Oδ | D22/Oδ | D24/Oδ | T26/O | D20/Oδ[a] | D22/Oδ | D24/Oδ | T26/O |
| | Natural | 2.275 | 2.382 | 2.335 | 2.424 | 2.611 | 2.455 | 99.0 | 73.5 | 149.8 | 129.9 |
| | Identified | 2.249 | 2.441 | 2.446 | 2.430 | 2.401 | 2.347 | 110.9 | 74.3 | 145.1 | 133.0 |
| | 2 | E67/Oε1 | E67/Oε2 | D56/Oδ | D58/Oδ | N60/Oδ | T62/O | D56/Oδ[a] | D58/Oδ | N60/Oδ | T62/O |
| | Natural | 2.308 | 2.487 | 2.206 | 2.479 | 2.463 | 2.172 | 89.9 | 71.1 | 150.0 | 128.7 |
| | Identified | 2.375 | 2.469 | 2.407 | 2.367 | 2.388 | 2.340 | 106.5 | 78.0 | 149.3 | 138.1 |
| | 3 | E104/Oε1 | E104/Oε2 | D93/Oδ | D95/Oδ | N97/Oδ | Y99/O | D93/Oδ[a] | D95/Oδ | N97/Oδ | Y99/O |
| | Natural | 2.317 | 2.764 | 2.135 | 2.221 | 2.388 | 2.055 | 95.4 | 78.1 | 149.0 | 130.1 |
| | Identified | 2.156 | 2.452 | 2.425 | 2.180 | 2.423 | 2.079 | 110.4 | 76.5 | 148.5 | 137.4 |
| | 4 | E140/Oε1 | E140/Oε2 | N129/Oδ | D131/Oδ | D133/Oδ | Q135/O | N129/Oδ[a] | D131/Oδ | D133/Oδ | Q135/O |
| | Natural | 2.322 | 2.572 | 2.166 | 2.558 | 2.074 | 2.383 | 84.6 | 85.1 | 164.5 | 113.0 |
| | Identified | 2.312 | 2.309 | 2.371 | 2.391 | 2.338 | 2.345 | 96.1 | 75.6 | 146.0 | 137.4 |

[a]Oxygen atom is placed out of the plane.

**TABLE II. Ca—O Lengths and O—Ca—Oε Angles of the Natural and Reidentified Calcium-Binding Sites**

| Protein | Site | Ca—O lengths (Å) | | | | | | O—Ca—Oε angles (°) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Parv | 3 | E101/Oε1 | E101/Oε2 | D90/Oδ | D92/Oδ | D94/Oδ | K96/O | D90/Oδ[a] | D92/Oδ | D94/Oδ | K96/O |
| | Natural | 2.486 | 2.505 | 2.246 | 2.415 | 2.439 | 2.294 | 116.5 | 124.4 | 147.2 | 79.7 |
| | Identified | 2.124 | 2.560 | 2.575 | 2.445 | 2.200 | 2.140 | 131.9 | 129.0 | 146.3 | 84.3 |
| CBD | 2 | E65/Oε1 | E65/Oε2 | D54/Oδ | D56/Oδ | D58/Oδ | E60/O | D54/Oδ[a] | D56/Oδ | D58/Oδ | E60/O |
| | Natural | 2.534 | 2.538 | 2.412 | 2.336 | 2.387 | 2.387 | 112.3 | 125.2 | 152.9 | 76.8 |
| | Identified | 2.461 | 2.482 | 2.301 | 2.309 | 2.459 | 2.246 | 112.9 | 122.8 | 161.2 | 88.2 |
| | 1 | E27/Oε1 | E27/Oε2 | A14/O | E17/O | D19/O | Q22/O | A14/O[a] | E17/O | D19/O | Q22/O |
| | Natural | 2.124 | 2.560 | 2.575 | 2.445 | 2.200 | 2.140 | 98.7 | 110.0 | 161.8 | 82.1 |
| | Identified | 2.269 | 2.395 | 2.336 | 2.454 | 2.416 | 2.428 | 92.0 | 126.2 | 150.9 | 76.0 |
| S100A12 | 2 | E72/Oε1 | E72/Oε2 | D61/Oδ | N63/Oδ | D65/Oδ | Q67/O | D61/Oδ[a] | N63/Oδ | D65/Oδ | Q67/O |
| | Natural | 2.532 | 2.570 | 2.385 | 2.344 | 2.401 | 2.346 | 112.4 | 125.4 | 151.1 | 79.2 |
| | Identified | 2.204 | 2.385 | 2.156 | 2.383 | 2.575 | 2.476 | 142.1 | 123.6 | 116.8 | 118.6 |
| | 1 | E31/Oε1 | E31/Oε2 | S18/O | K21/O | H23/O | T26/O | S18/O[a] | K21/O | H23/O | T26 |
| | Natural | 2.583 | 2.416 | 2.302 | 2.378 | 2.436 | 2.402 | 74.7 | 72.6 | 140.9 | 122.4 |
| | Identified | 2.405 | 2.325 | 2.306 | 2.390 | 2.536 | 2.417 | 104.6 | 63.0 | 149.6 | 128.3 |
| Ther | 2 | D62/Oδ1 | D62/Oδ2 | D57/Oδ | Q66/Oε | T64/O | | D57/Oδ[a] | Q66/Oε | T64/O | |
| | Natural | 2.463 | 2.640 | 2.420 | 2.371 | 2.350 | | 93.0 | 146.7 | 79.3 | |
| | Identified | 2.339 | 2.378 | 2.440 | 2.709 | 2.392 | | 87.6 | 139.6 | 76.2 | |
| α-Lac | 1 | D82/Oδ | D84/O | D87/Oδ | D88/Oδ | K79/O | | D84/O[a] | D87/Oδ | D88/Oδ | K79/O[a] |
| | Natural | 2.366 | 2.269 | 2.319 | 2.293 | 2.222 | | 85.7 | 140.2 | 145.5 | 79.5 |
| | Identified | 2.529 | 2.526 | 2.380 | 2.372 | 2.389 | | 65.0 | 145.8 | 113.2 | 74.8 |

[a]Oxygen atom is placed out of the plane.

binding sites in α-lactalbumin and thermitase is similar to that of EF-hand proteins.

Based on the structural data of the coordination shell, all calcium-binding sites in natural calcium-binding proteins can be described with the common parameter set of the ideal pentagonal bipyramidal geometry. This geometry has Ca—O bond lengths of 2.4 Å with a relatively narrow range from 1.5–3.5 (2.5 ± 1.0) Å and a relatively open value with a range of 120° in the O—Ca—O angle to accommodate the great variations in different calcium-binding sites. We hypothesize that all the natural calcium-binding sites in these proteins can be reidentified with this set of structural parameters. To test this hypothesis, we carried out reidentifying calcium-binding sites in natural calcium-binding proteins after removing all of the calcium-binding sites (see Materials and Methods).

## Establishing Structural Description of Calcium-Binding Sites Using Calmodulin

All four of the natural calcium-binding sites in the calcium-bound form of calmodulin were first used for establishing and training our geometric parameters. They were successfully identified with the descriptions of the bond lengths of Ca—O from 1.5–3.5 Å (2.5 ± 1.0 Å) and a range of 120° for the O—Ca—O angles. The bond lengths of Ca—O and angles of O—Ca—O for the natural and reconstructed calcium-binding sites are listed in Table I. All of the reidentified natural calcium-binding sites use the expected ligand residues and corresponding oxygen atoms. As shown in Figure 3, the local geometries of the reidentified natural calcium-binding sites are similar to those of natural calcium-binding sites. The deviations of bond lengths of Ca—O and O—Ca—O angles are within 0.35 Å and 18°, respectively (Table I).

About 4000 potential sites were constructed in calmodulin. Their locations in the primary sequence are shown in Figure 4. The majority of the constructed sites were located at the four calcium-binding loops while only four bidentate Glu ligands (anchor) are placed out of the calcium-binding loops. Besides Glu at loop position 12, loop positions 2, 8, and especially 9 are also used to place the bidentate Glu (anchor). Because loop position 9 in the natural calcium-binding sites is often occupied by an oxygen ligand from a bridged water molecule,[5] in our calculation process water molecules were removed along with other heteromolecules before the computational con-

struction of metal-binding sites. This may cause a relatively larger space allowing for the placement of the oxygen atoms as the calcium-binding ligands from the longer side-chain of Glu.

The $U(p)$ numbers for all of the reconstructed calcium-binding sites in calmodulin are plotted in Figure 5. The highest number is 150.55 and the lowest is 5.02 with a mean of 58.21. The $U(p)$ number indicates the deviation of the geometry of a site from identity. For example, site 3091 has the highest $U(p)$ value of 150.55 with bond lengths that vary from 1.9–3.7 Å and O—Ca—O angles with a 50° deviation from the target values. Interestingly, all natural calcium-binding sites have small $U(p)$ numbers ($< 20$), suggesting small deviations from the ideal pentagonal bipyramid target model geometry.

Several similarities are shared by all of the constructed sites with $U(p)$ numbers close to those of the natural sites. First, they all show the correct bidentate Glu residues from positions 31, 67, 104, and 140 contained in the natural calcium-binding sites of calmodulin. Second, main-chain oxygen atoms from T26, T62, Y99, and Q135 in natural calcium-binding sites (sites 1–4) are always used. Third, residue numbers used for the remaining four ligand positions are the same as those in the natural calcium-binding sites with either Asp replaced by Asn or Asn replaced by Asp. These calcium-binding sites are introduced mainly because both Asn and Asp are used for four positions and both residues have similar side-chain rotamer configurations.

## Testing and Identification of Different Classes of Calcium-Binding Sites in Proteins

### *Classic EF-hand calcium-binding sites*

To evaluate the structural parameter set (Ca—O length of 2.5 ± 1.0 Å and O—Ca—O angle range of 120°), we carried out identifying three classic EF-hand calcium-binding sites from site 2 of calbindin$_{D9K}$, site 3 of parvalbumin, and site 2 of S100 A12. There were 399 potential sites constructed in calbindin$_{D9K}$ using the parameter set. Like calmodulin, the reidentified natural classic EF-hand site uses the expected ligand residues and corresponding oxygen atoms as shown in Figure 3 and Table I. In addition, the geometric properties, such as Ca—O bond lengths and O—Ca—O angles of the constructed calcium-binding sites, are similar to those in the natural calcium-binding sites. Among the 399 constructed calcium-binding sites in calbindin$_{D9K}$, 10% use the desired Glu65 at loop position 12 as the bidentate ligand (anchor) while the rest of the calcium-binding sites use Glu at loop position 62 (loop position 9) as the bidentate anchor. This situation is similar to that of calmodulin. The $U(p)$ value for the reidentified natural calcium-binding site is again the lowest among all the constructed sites of calbindin$_{D9K}$ (Fig. 5).

Site 3 of the EF-hand calcium-binding site (residues 90–101) in parvalbumin is also identified (Fig. 3 and Table II). The crystal structure 5CPV[29] was used to search the calcium-binding sites in parvalbumin using the same parameters as those used for calmodulin. To save time and disk space, the anchor positions were restricted to the residues after Gly80 to the C-terminal in parvalbumin. After removing the degenerate sites, 650 sites were constructed with an average $U(p)$ value of 54.42. The highest $U(p)$ value is 129.56 and the lowest is 10.06. The reidentified natural calcium-binding site in parvalbumin has a $U(p)$ of 19.54. Like calmodulin, about 52% of the 650 sites were constructed using the anchor located at residue 98, which is the ninth residue in the EF-loop. The other anchor positions used in the search were residues 81, 85, 91, and 101, which are located in the E-helix ($-9$, $-5$) and at the loop positions 2 and 12, respectively.

The classic EF-hand site 2 of S100A12 has also been reidentified. Among 327 calcium-binding sites constructed at residues in the EF-hand motif for site-search and refinement, the reidentified natural calcium-binding site with side-chain oxygen atoms from Asp61, Asn63, Asp65, Glu72, and the main-chain oxygen Gln67 has the pseudoenergy of 20.04. As shown in Table II, the Ca—O bond length and O—Ca—O angles of the reidentified natural calcium-binding site are similar to the natural calcium-binding site 2.

## Identification of Pseudo-EF-Hand Calcium-Binding Sites

Calbindin$_{D9k}$ and S100A12 have a pseudo-EF-hand calcium-binding site 1. Differing from their classic EF-hand calcium-binding site 2, four main-chain oxygen atoms at loop positions 1, 4, 6, and 9, rather than side-chains, are used as ligands except for the side-chain oxygen atoms at position 12, which remains as a bidentate ligand. To identify the pseudo-EF-hand site 1 in both proteins, the same Glu—Ca pseudoresidue used in the classic EF-hand sites is used as the anchor and the other ligands are all from main-chain carbonyl groups. When a main-chain carbonyl is used as a ligand in the calculation, only the anchor rotamer is used to search the positions that satisfy the requirements. In contrast, when the side-chain oxygen is used as a ligand both rotamers of the anchor and the tested residue type are used to search positions that satisfy the requirements. Therefore, to construct a calcium-binding site with the same parameters using the main-chain carbonyl groups as ligands is more difficult than using the side-chain carboxyl groups. To facilitate the calculation, only the oxygen—calcium distance is limited to 1.5–3.5 Å. All of the angles and dihedral angles are unconstrained ($-180$–$+180$°). However, the same target values are used for the minimization of the geometry, which eliminates the larger deviation caused by the loose restrictions for the construction of a site. Using the parameters with enlarged angle ranges, the identified natural pseudo-EF-hand site has an identical ligand type and ligand position to that of site 1 in calbindin$_{D9k}$. As shown in Figure 4 and Table II, the geometric values for both angles and both lengths for the identified pseudo-EF-hand sites from calbindin$_{D9k}$ and S100A12 are also similar to their corresponding natural calcium-binding site 1.

## Comparison of Natural EF-Hand Calcium-Binding Sites

The $U(p)$ number reflects the difference between the constructed metal-binding site and the defined target site

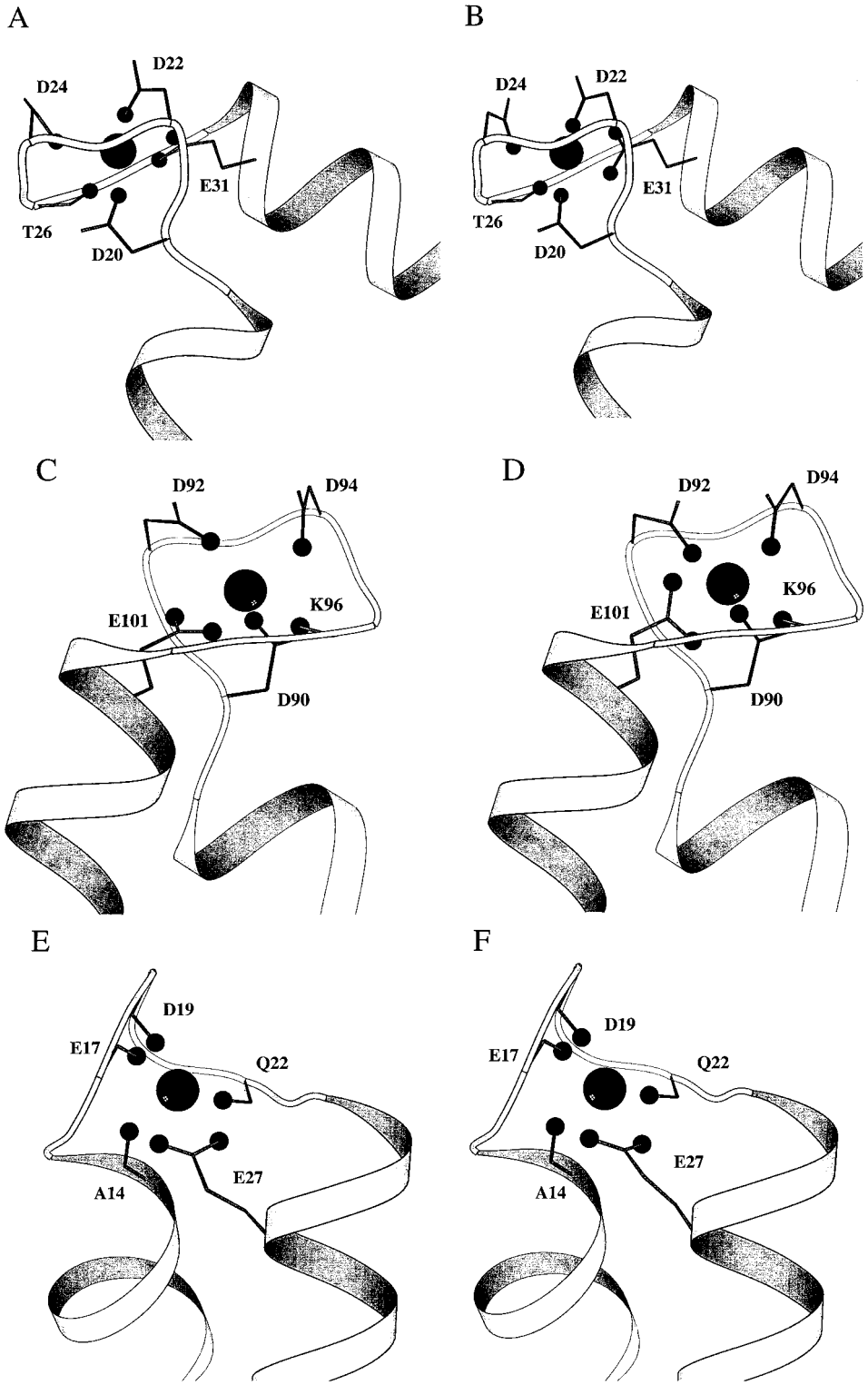



Fig. 3. The local structures of the natural (left) and reidentified (right) EF-hand calcium-binding sites are similar. (**A–B**). Site 1 of calmodulin. (**C–D**). Site 3 of parvalbumin. (**E–F**). Site 1 of calbindin$_{D9k}$. (**G–H**). Site 2 of calbindin$_{D9k.}$ (**I–J**). Site 1 of S100A12. The oxygen atoms directly linked to the calcium are shown as filled circles.

based on the bond length, angle, and side-chain clashes (see Material and Methods, eq. 1). As shown in Table III, the identified natural calcium-binding sites in calmodulin, calbindin$_{D9k}$ (site 2), S100A12, and parvalbumin have small $U(p)$ values ($\leq 20$). More strikingly, they all have the smallest $U(p)$ values among all of the constructed sites for each protein. Therefore, the $U(p)$ number can be used to rank the constructed potential sites for their deviations

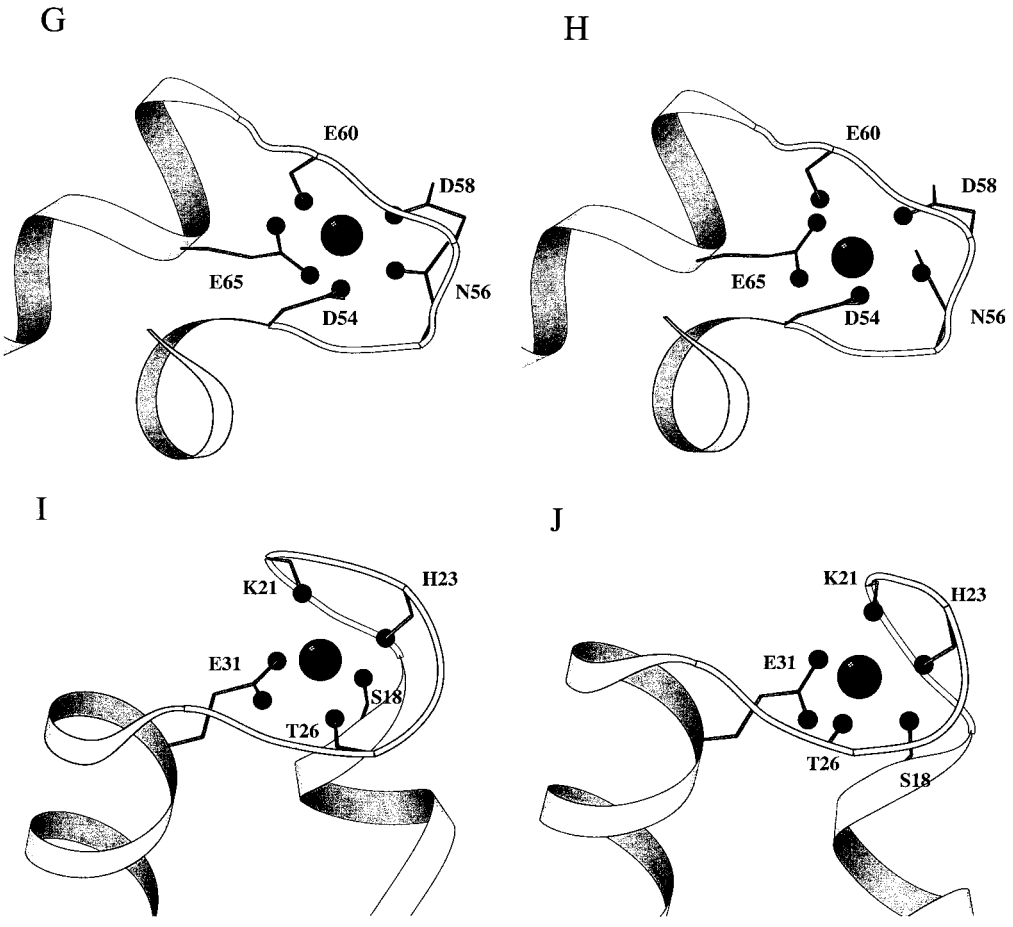



Figure 3. (Continued.)

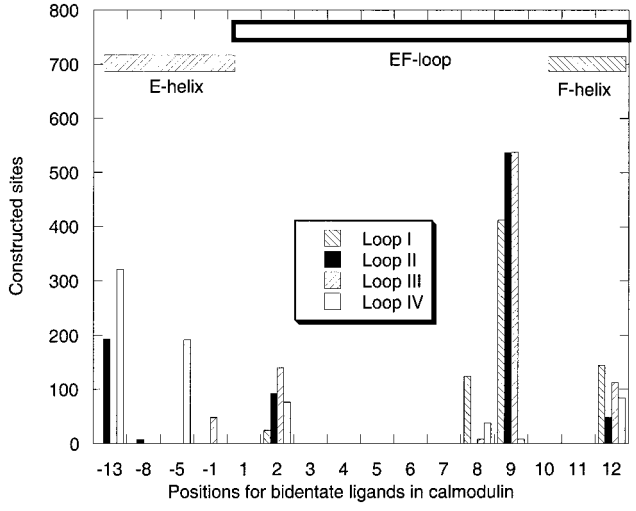



Fig. 4. Locations of residues used for the construction of the bidentate ligand in calmodulin. The sequence positions and relative orientations of the ligand residues are labeled above the amino acid sequences.

from the target site. The parameters that we use to describe the geometry of calcium-binding sites allow us to accurately identify natural calcium-binding sites. When the coordinates of the natural EF-hand sites from these proteins are minimized with the same defined parameters of the ideal pentagonal bipyramidal geometry, the differences between the natural sites in proteins can be evaluated. As shown in Table III, the difference in the pseudoenergy value between the natural classic EF-hand calcium-binding site and the identified site of parvalbumin (site 3) is noticeably greater than that of the remaining sites we examined. The reidentified natural site 3 of parvalbumin exhibits the largest deviations from those of the natural sites in both O—Ca length and O—Ca—O angle (Table II).

### *Non-EF-hand calcium-binding sites in thermitase and α-lactalbumin*

Table II shows that the local geometries of non-EF-hand calcium-binding sites in α-lactalbumin and thermitase are similar to those of EF-hand proteins although calcium-binding ligand residues are not conserved. All of the calcium-binding sites in these two proteins were identified using different residues as anchors. For example, to reidentify the calcium-binding site 2 in thermitase in the high-resolution crystal structure (1THM),[32] an anchor residue with $Ca^{2+}$ attached to the Asp was used. The other four ligands from the protein were defined using the same parameter set established for calmodulin. A total of 518 sites were constructed in the residue sequence from 50–70

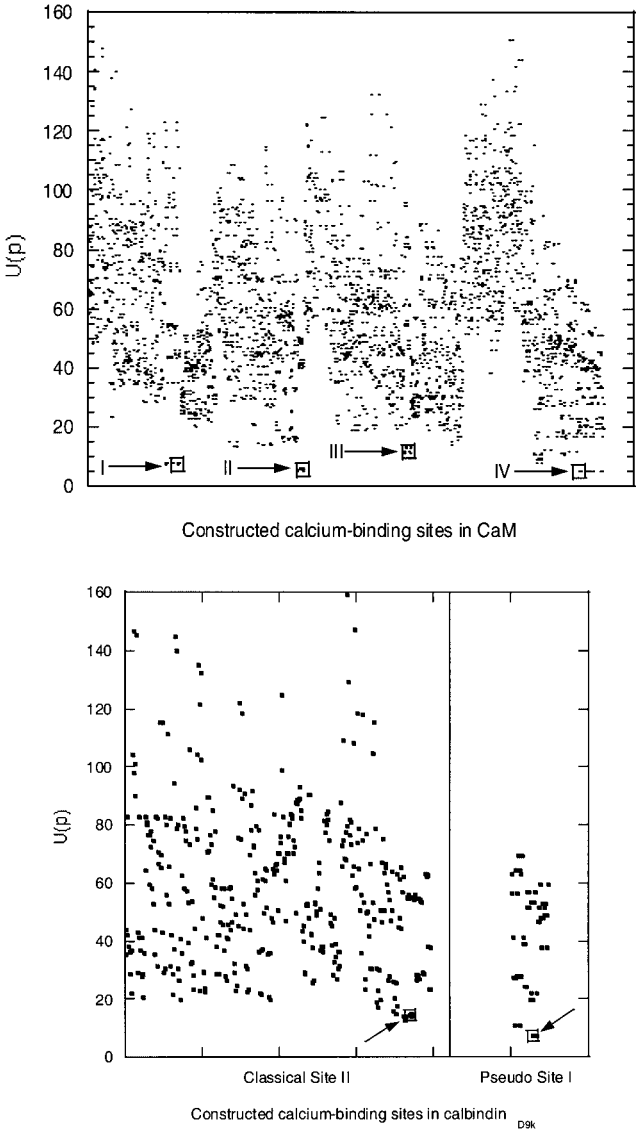



Fig. 5. Deviation of the designed sites from the ideal pentagonal bipyramidal geometry [pseudoenergy $U(p)$] for calmodulin (top) and calbindin$_{D9k}$ (bottom). The natural calcium-binding sites of these proteins are shown with an arrow.

in thermitase. The average $U(p)$ of these sites was 40.61 with a maximum value of 121.28 and a minimum value of 6.33. The reidentified natural calcium-binding site has a deviation (7.68) from the target site comparable to that of the natural calcium-binding site (7.65) (Fig. 6). The same process has been performed to identify the natural calcium-binding site in α-lactalbumin (1ALC).[31] A total of 144 sites were constructed from 75–95 in the protein sequence with an average $U(p)$ of 41.99. The reidentified natural calcium-binding site in α-lactalbumin has a $U(p)$ of 18.35, close to the lowest value of 12.90 (Fig. 6 and Table III). Therefore, the non-EF-hand calcium-binding sites can be described and identified using the same set of geometric parameters with the change of anchor residues.

**TABLE III. Pseudoenergy Value of Reidentified Sites and Natural Sites**

| Protein | Site | Natural | Reidentified |
|---|---|---|---|
| Calmodulin | 1 | 7.84 | 7.47 |
| | 2 | 5.54 | 5.72 |
| | 3 | 11.41 | 11.86 |
| | 4 | 5.44 | 5.26 |
| Parvalbumin | 3 | 5.66 | 19.83 |
| Calbindin$_{D9k}$ | 2 | 10.23 | 14.50 |
| | 1 | 4.52 | 7.02 |
| S100A12 | 2 | 19.24 | 20.04 |
| | 1 | 6.51 | 6.02 |
| Thermitase | 2 | 7.65 | 7.68 |
| α-Lactalbumin | 1 | 16.13 | 18.35 |

## Two Different Magnesium-Binding Sites in EF-Hand Proteins Can Also Be Identified

### *Common features of magnesium-binding sites*

$Mg^{2+}$ (0.65 Å) is 24% smaller than $Ca^{2+}$ (0.99 Å). The smaller size of $Mg^{2+}$ determines its preference for a coordination number of six with an octahedral geometry of its complexes. In contrast, $Ca^{2+}$ favors a coordination number of seven. The bond distances to the oxygen ligand atom typically range from 2.0–2.1 Å for $Mg^{2+}$ and 2.1–2.8 Å for $Ca^{2+}$. $Mg^{2+}$ is less comfortable than $Ca^{2+}$ in accepting large multidentate and anionic ligand groups.

Structural analysis of the available high-resolution structures of magnesium-loaded forms of calbindin$_{D9k}$ and parvalbumin were also carried out. An octahedral geometry with an average Mg—O length of 2.1 Å is observed for both proteins that differ from the pentagonal bipyramidal geometry of calcium-binding sites having an average Ca—O length of 2.4 Å. Magnesium-binding sites for both proteins exhibit several variations in ligand types in contrast to the conserved ligand types of both classic and pseudo-calcium-binding sites. As shown in Figure 7(c), magnesium binds to site 2 of calbindin$_{D9k}$ by six oxygen atoms from four residues and two water molecules. Three carboxyl groups from the side-chains of Asp54, Asn56, and Asp58 and one carbonyl group from Glu60 directly coordinate the Mg(II) metal ion. In addition, one of the Mg(II) coordinating water molecules (Wat1) is hydrogen bonded to the side-chains of Gln22, Asp58, and Asn56. The other Mg(II) coordinating water (Wat2) is hydrogen bonded by the side-chain of Glu65.[33] On the other hand, magnesium binds to site 3 of parvalbumin with all the ligands used for calcium binding except for the bidentate ligand Glu at loop position 12, which is converted to a unidentate ligand.[16,17,37]

### *Magnesium-binding site in calbindin$_{D9k}$*

The natural magnesium-binding site (site 2) of calbindin$_{D9k}$ was first described and identified (5ICB) using the common octahedral geometry.[33] A pseudoresidue Asp with magnesium attached (Asp—Mg) was used as the anchor. Four ligands at loop positions 1, 3, 5, and 7 were used to search the magnesium-binding sites in calbindin$_{D9k}$ with an O—Mg distance range of 1.0–3.0 Å, and an O—Mg—O angle range of 30–140°. As the limita-

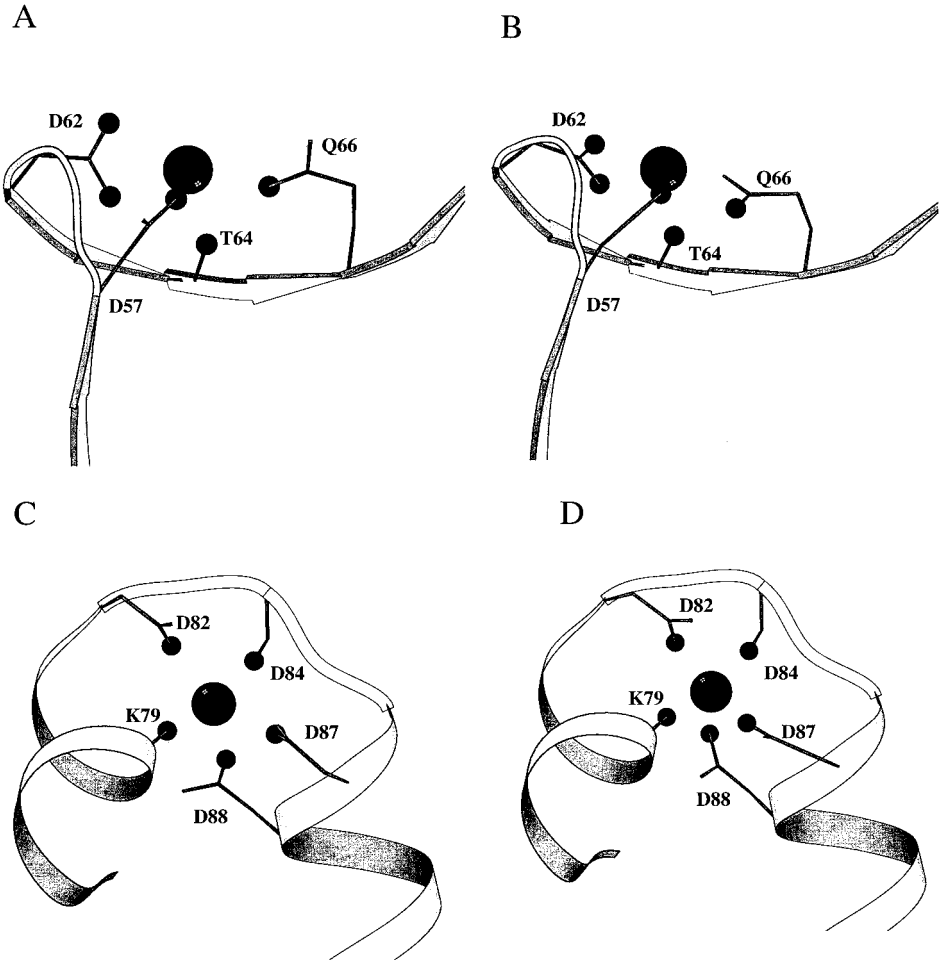



Fig. 6. Local structures of the calcium-binding site 2 of non-EF-hand proteins thermitase (**A**) and site 1 of α-lactalbumin (**C**). The corresponding reidentified sites are shown in (**B**) and (**D**).

tion is reduced with fewer ligands, the possible sites will increase. To avoid the large file size resulting from the increase of possible sites with fewer ligands used, the anchor positions were limited around the EF-hand motif 2 of the protein (residue 50–70). The octahedral geometry with an O—Mg distance of 2.1 Å was used as the target geometry for the refinement step of magnesium-binding sites. The weight coefficient constant for each deviation in eq. 1 remains the same as used for calcium-binding sites. After removing the degenerate sites by using the CLEAN program, 1171 sites were constructed. The average $U(p)$ for all these sites is 39.29 with the minimum at 11.87 and the maximum at 93.17 (Fig. 8). As shown in Figure 7 and Table III, the natural magnesium-binding site is also successfully reidentified; it also has the lowest $U(p)$ number of 14.91.

### *Magnesium-binding site in parvalbumin*

The magnesium site 3 in parvalbumin was further used to test the same common parameters of Mg—O length and O—Mg—O angle. About 2480 magnesium-binding sites with pseudoenergy from 31.87–165.87 (average of 70.22) are constructed in parvalbumin. Surprisingly, the identified natural magnesium-binding site has a pseudoenergy of 41.75, significantly greater than that of calcium.

Table IV shows the O—Mg distances and O—Mg—O angles of the natural magnesium-binding site and the identified natural site. Each O—Mg distance is close to 2.1 Å except for the side-chain oxygen of Glu101. The O—Mg distance of the side-chain of Glu101 (at loop position 12) is 2.625 Å, significantly greater than that for the rest of the ligand atoms. Such a large deviation from the target value of 2.1 Å is likely to contribute to the greater pseudoenergy of the magnesium-binding site (41.75) than that for the classic EF-hand calcium-binding sites ($< 20$). It is interesting to note that the relatively longer Mg—O distance at the C-terminal EF-loop observed here correlates with the nuclear magnetic resonance (NMR) structural studies of calmodulin by Ohki et al.[13] and Malmenda et al.[14] By monitoring chemical shift changes as a function of magnesium concentrations they observed that a magnesium ion binds to the *N*-terminal half of the EF-hand motif of calmodulin in solution. Position 12 of the EF-loop is not involved in the binding of magnesium, and position 7 only

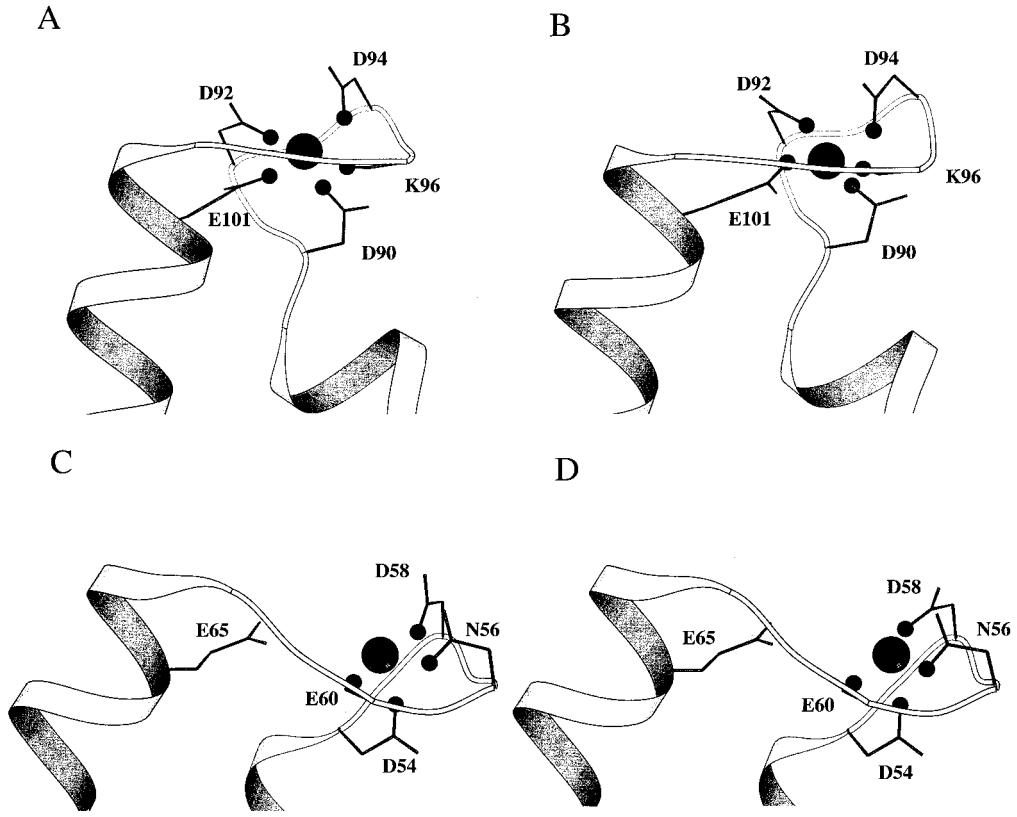



Fig. 7. Local structures of the natural (**A**) and reidentified (**B**) magnesium-binding site 3 of parvalbumin and the natural (**C**) and reidentified (**D**) magnesium-binding site 2 of calbindin$_{D9k}$.

partly coordinates to the magnesium ion. Such preferential binding of the *N*-terminal of the EF-loop by magnesium ions is likely due to different metal properties of magnesium vs. calcium and was proposed to be responsible for the lack of conformational change upon magnesium binding.[14] Magnesium rarely uses neutral oxygen donors such as carbonyls and hydroxyls as ligands, while calcium interacts with both negatively charged carboxyls and neutral carbonyls and hydroxyls.[4]

### Design of Calcium-Binding Sites in Non-Calcium-Binding Proteins

Chicken-type (c-type) lysozymes and α-lactalbumin are evolutionarily related proteins with strong similarity in their amino acid sequences and 3-D structures.[31,38] Egg-white lysozyme is not able to bind calcium because the calcium-binding ligand residues at positions 79 and 88 of α-lactalbumin were mutated at the corresponding positions in lysozyme. Our studies have shown that many calcium-binding sites can be constructed in lysozyme at the same location as α-lactalbumin with the potential calcium-binding ligand oxygen atoms from side-chains of Asp, Glu, Asn, and Gln and carbonyl main chains from any amino acids of the crystal structure 1LYZ[39] using the same parameters as for α-lactalbumin. Although several amino acids (Gln, Asp, and Asn) clustered in the same continuous loop at positions 86, 87, and 88 could serve as calcium-binding ligands, no calcium-binding sites using all-natural residues (i.e., Asp, Asn, Glu, and Gln, main-chain carbonyl) are identified in lysozyme by the program.

To further test the possibility of designing calcium-binding sites in non-calcium-binding proteins without generating artifact calcium-binding sites, proteins with different scaffolds were used. Domain 1 of CD2 is an all β-sheet protein with an IgG fold[40] while Rop has a four-helix bundle protein frame.[41] Both proteins contain many amino acids that may serve as calcium-binding ligands. For example, domain 1 of CD2 has 8 Asp, 5 Glu, 8 Asn, and 1 Gln (22% of the protein). Glu—Ca was used as the anchor residue to design calcium-binding sites in CD2 and Rop. About 7000 and 1800 calcium-binding sites can be constructed in loops and exposed positions of CD2 and Rop, respectively. $U(p)$ values are varied from 3.15–219.84 (average 47.12) for CD2 and from 6.56–172.86 (average 50.61) for Rop. As expected, no calcium-binding sites were found in either non-calcium-binding proteins using all the natural residues. All of the designed calcium-binding sites in these proteins contain at least three or more mutations to the original structure with the calcium-binding ligand residues. Our lab has shown that four designed calcium-binding sites in CD2 bind calcium strongly.[42] Our results have two important implications. First, using the established geometric descriptions, we are able to specifically identify natural calcium-binding sites with the lowest deviations. No artificial "natural" calcium-binding sites can be generated using all of the possible natural amino

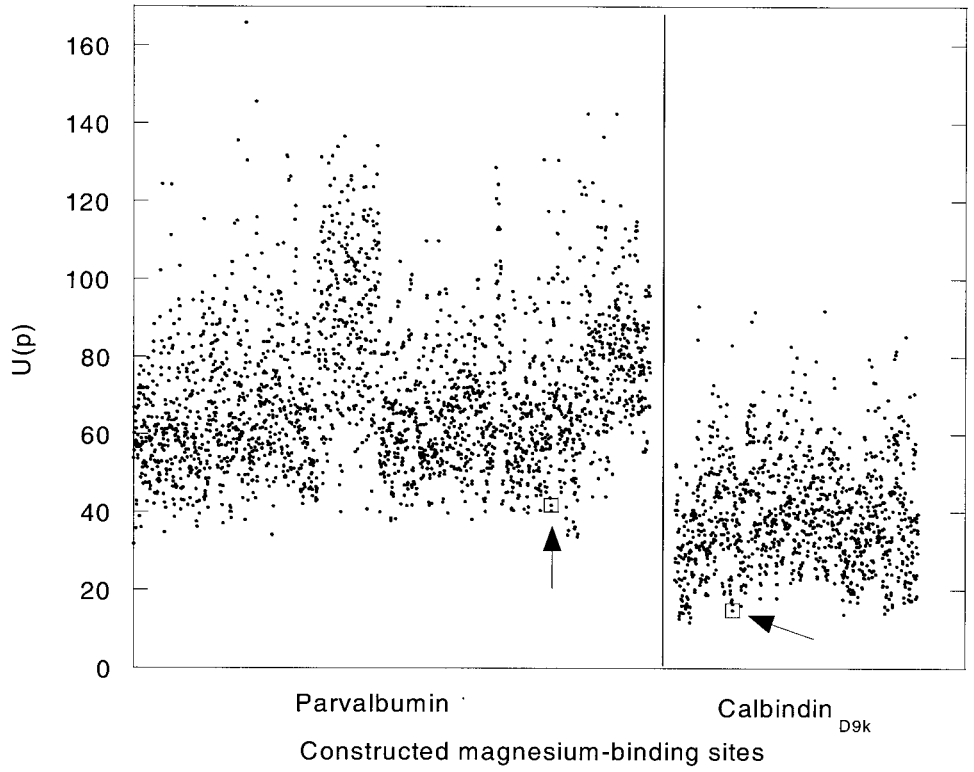



Fig. 8. Deviation of the designed sites from the ideal octahedral geometry [(pseudoenergy $U(p)$] for calbindin$_{D9k}$ and parvalbumin. The natural magnesium-binding sites of both proteins are labeled with arrows.

**TABLE IV. Mg—O Lengths and O—Mg—O Angles of the Natural and Reidentified Magnesium-Binding Sites**

| Protein | Site | O—Mg distance (Å) | | | | | O—Mg—Oδ angle (°) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| CBD | 2 | D54/Oδ | N56/Oδ | D58/Oδ | E60/O | | N56/Oδ | D58/Oδ | E60/O | |
| | Natural | 2.185 | 2.232 | 2.254 | 2.130 | | 85.0 | 82.6 | 89.5 | |
| | Identified | 2.112 | 2.137 | 2.212 | 2.232 | | 94.2 | 87.8 | 82.8 | |
| Parv | 3 | D90/Oδ | D92/Oδ | D94/Oδ | M96/O | E101/Oϵ | D92/Oδ | D94/Oδ | M96/O | E101/Oϵ |
| | Natural | 2.073 | 2.083 | 2.176 | 2.142 | 2.047 | 85.5 | 85.0 | 85.0 | 106.7 |
| | Identified | 2.166 | 1.976 | 2.180 | 2.142 | 2.625 | 75.6 | 108.0 | 87.9 | 111.1 |

acids, even when they are clustered in the protein sequence. Second, it is possible to de novo design calcium-binding sites in non-calcium-binding proteins with this established computational method. Designing calcium-binding sites in a non-calcium-binding scaffold protein contributes to our understanding of the protein design process in general and the biological role of calcium ions in particular. It allows us to define the key determinants for calcium-binding affinity and metal selectivity by establishing the required metal coordination ligands and geometry for calcium binding in a scaffold that lacks the complication of global conformational changes and the cooperative effect of multiple calcium-binding sites in proteins, which are frequently observed in natural calcium-binding proteins.[5] Because isotope-labeled metal ions can be tracked by radiological, NMR, or chemical means, our success in designing metal-binding sites into arbitrary proteins will likely lead to new ways of developing useful reagents for diagnostic tests and chemotherapy.

## CONCLUSIONS

Although there are many complications associated with calcium-binding sites in proteins because of various ligand types and irregularities in metal coordination geometries,[18,19,26] we have shown that all the natural calcium-binding sites of EF-hand (both classic and pseudo-EF-hand motifs) and non-EF-hand proteins can be described and identified using a set of geometric descriptions of the ideal pentagonal bipyramid geometry. The relative $U(p)$ values can be used to rank the success of each constructed site. The searched native-like sites in proteins have the smallest deviation from the target geometry. In addition, two different natural magnesium-binding sites in calbindin$_{D9k}$ and parvalbumin can also be identified in the metal-loaded form of proteins. None of the natural calcium- or magnesium-binding sites can be constructed in the apo-form of calmodulin. In addition, no calcium-binding sites were found in non-calcium-binding proteins

CD2 and Rop with all-natural residues. Four designed calcium-binding sites have been engineered into CD2 and they have shown a strong metal-binding ability. Our work indicates that a useful method for searching calcium-binding sites in proteins has been developed. It is possible to use the established parameters to identify calcium- and magnesium-binding sites in proteins and design de novo calcium-binding sites.

## ACKNOWLEDGMENTS

The authors thank Dr. Robert Kretsinger and Dan Adams for the critical review of this article. They thank Michelle Pu for the CLEAN computer program. They appreciate the helpful discussions from Don Harden, Kim Gernert, David Wilson, Robert Wohlhueter, Laura Morris, Yiming Ye, Anna Wilkins, Amy Carroll, Curt Coman, April Ellis, Leanne Isley, and the rest of the members in Dr. Jenny J. Yang's group. This work is supported in part by GSU start-up funds, QIF, Research Initiation, Mentoring Grant, and in part by the National Science Foundation (NSF) grant MCB-0092486 and NIH GM 62999-1 for J.J.Y.

## REFERENCES

1. Sanchez R, Pieper U, Melo F, Eswar N, Marti-Renom MA, Madhusudhan MS, Mirkovic N, Sali A. Protein structure modeling for structural genomics. Nat Struct Biol 2000;7(Suppl.):986–990.
2. Service RF. Structural genomics offers high-speed look at proteins. Science 2000;287:1954–1956.
3. Hellinga HW. The construction of metal centers in proteins by rational design. Fold Des 1998; 3:R1–8.
4. Silva JJRF, Williams RJP. The biological chemistry of the elements. Oxford, UK: Oxford University Press; 1991.
5. Linse S, Forsen S. Determinants that govern high-affinity calcium binding. Adv Second Messenger Phosphoprot Res 1995;30:89–151.
6. Falke JJ, Drake SK, Hazard AL, Peersen OB. Molecular tuning of ion binding to calcium signaling proteins. Q Rev Biophys 1994;27: 219–290.
7. Kretsinger RH. EF-hands embrace. Nat Struct Biol 1997;4:514–516.
8. Kawasaki H, Kretsinger RH. Calcium-binding proteins 1: EF-hands. Protein Profile 1995; 2:297–490.
9. Kretsinger RH, Nockolds CE. Carp muscle calcium-binding protein. II. Structure determination and general description. J Biol Chem 1973;248:3313–3326.
10. Kretsinger RH, Nockolds CE, Coffee CJ, Bradshaw RA. The structure of a calcium-binding protein from carp muscle. Cold Spring Harb Symp Quant Biol 1972;36:217–220.
11. Wnuk W, Cox JA, Stein EA. Parvalbumins and other soluble high-affinity calcium-binding proteins from muscle. In: Cheung WY, editor. Calcium & cell function, vol. 2. New York: Academic Press; 1982. pp. 243–278.
12. Berridge MJ, Bootman MD, Lipp P. Calcium—a life and death signal. Nature 1998;395:645–648.
13. Ohki S, Ikura M, Zhang M. Identification of Mg2+-binding sites and the role of Mg2+ on target recognition by calmodulin. Biochemistry 1997;36:4309–4316.
14. Malmendal A, Evenas J, Thulin E, Gippert GP, Drakenberg T, Forsen S. When size is important. Accommodation of magnesium in a calcium binding regulatory domain. J Biol Chem 1998;273: 28994–29001.
15. Malmendal A, Linse S, Evenas J, Forsen S, Drakenberg T. Battle for the EF-hands: magnesium–calcium interference in calmodulin. Biochemistry 1999;38:11844–11850.
16. Declercq JP, Tinant B, Parello J, Etienne G, Huber R. Crystal structure determination and refinement of pike 4.10 parvalbumin (minor component from *Esox lucius*). J Mol Biol 1988; 202:349–353.
17. Declercq JP, Tinant B, Parello J, Rambaud J. Ionic interactions with parvalbumins. Crystal structure determination of pike 4.10 parvalbumin in four different ionic environments. J Mol Biol 1991;220:1017–1039.
18. McPhalen CA, Strynadka NC, James MN. Calcium-binding sites in proteins: a structural perspective. Adv Protein Chem 1991;42: 77–144.
19. Glusker JP. Structural aspects of metal liganding to functional groups in proteins. Adv Protein Chem 1991;42:1–76.
20. Yang W, Lee HW, Pu M, Hellinga H, Yang JJ. Identifying and designing of calcium binding sites in proteins by computational algorithm. In: Computational studies, nanotechnology, and solution thermodynamics of polymer systems. Boston: Kluwer Academic/Plenum Publishers; 2000. pp. 127–138.
21. Brown EM, Pollak M, Hebert SC. The extracellular calcium-sensing receptor: its role in health and disease. Annu Rev Med 1998;49:15–29.
22. Pauls TL, Cox JA, Berchtold MW. The Ca2+(-)binding proteins parvalbumin and oncomodulin and their genes: new structural and functional findings. Biochim Biophys Acta 1996;1306:39–54.
23. Schafer BW, Heizmann CW. The S100 family of EF-hand calcium-binding proteins: functions and pathology. Trends Biochem Sci 1996;21:134–140.
24. Nayal M, Di Cera E. Predicting Ca(2+)-binding sites in proteins. Proc Natl Acad Sci USA 1994;91:817–821.
25. Yamashita MM, Wesson L, Eisenman G, Eisenberg D. Where metal ions bind in proteins. Proc Natl Acad Sci USA 1990;87:5648–5652.
26. Wei L, Huang ES, Altman RB. Are predicted structures good enough to preserve functional sites? Struct Fold Des 1999;7:643–650.
27. Hellinga HW, Richards FM. Construction of new ligand binding sites in proteins of known structure I. Computer-aided modeling of sites with pre-defined geometry. J Mol Biol 1991;222:763–785.
28. Moroz OV, Antson AA, Murshudov GN, Maitland NJ, Dodson GG, Wilson KS, Skibshoj I, Lukanidin EM, Bronstein IB. The three-dimensional structure of human S100A12. Acta Crystallogr Biol Crystallogr 2001;57:20–29.
29. Swain AL, Kretsinger RH, Amma EL. Restrained least squares refinement of native (calcium) and cadmium-substituted carp parvalbumin using X-ray crystallographic data at 1.6-Å resolution. J Biol Chem 1989;264:16620–16628.
30. Svensson LA, Thulin E, Forsen S. Proline *cis*-trans isomers in calbindin D9k observed by X-ray crystallography. J Mol Biol 1992;223:601–606.
31. Acharya KR, Stuart DI, Walker NPC, Lewis M, Phillips DC. Refined structure of baboon α-lactalbumin at 1.7Å resolution. J Mol Biol 1989;208:99–127.
32. Teplyakov AV, Kuranova IP, Harutyunyan EH, Vainshtein BK, Frommel C, Hohne WE, Wilson KS. Crystal structure of thermitase at 1.4 Å resolution. J Mol Biol 1990;214:261–279.
33. Andersson M, Malmendal A, Linse S, Ivarsson I, Forsen S, Svensson LA. Structural basis for the negative allostery between $Ca^{2+}$- and $Mg^{2+}$-binding in the intracellular $Ca^{2+}$-receptor calbindin $D_{9K}$. Protein Sci 1997;6:1139–1147.
34. Press WH, Flanery BP, Teuolsky SA, Vetterling WT. Numerical recipes in C. Cambridge, MA: Cambridge University Press; 1988.
35. Strynadka NC, James MN. Crystal structures of the helix–loop–helix calcium-binding proteins. Annu Rev Biochem 1989;58:951–998.
36. Linse S, Bylsma NR, Drakenberg T, Sellers P, Forsén S, Thulin E, Svensson LA, Zajtzeva I, Zajtsev V, Marek J. A calbindin $D_{9k}$ mutant with reduced calcium affinity and enhanced cooperativity. Metal ion binding, stability, and structural studies. Biochemistry 1994;33:12478–12486.
37. Allouche D, Parello J, Sanejouand YH. Ca2+/Mg2+ exchange in parvalbumin and other EF-hand proteins. A theoretical study. J Mol Biol 1999;285:857–873.
38. Nitta K, Sugai S. The evolution of lysozyme and alpha-lactalbumin. Eur J Biochem 1989;182:111–118.
39. Diamond R. Real-space refinement of the structure of hen egg-white lysozyme. J Mol Biol 1974;82:371–391.
40. Jones EY, Davis SJ, Williams AF, Harlos K, Stuart DI. Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. Nature 1992;360:232–239.
41. Banner D, Kokkinidis M, Tsernoglou D. Structure of the ColE1 Rop protein at 1.7 Å resolution. J Mol Biol 1987;196:657–675.
42. Yang W, Ye Y, Lee H, Liu Z, Hellinga H, Yang J. Rational design of calcium-binding protein. Submitted for publication.