Protein Engineering vol.4 no.1 pp.49–56, 1990

Interaction of metal ions with carboxylic and carboxamide groups in protein structures

P.Chakrabarti

Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90024, USA

Present address: Division of Physical Chemistry, National Chemical Laboratory, Pune 411008, India

An analysis of the geometry of metal binding by carboxylic and carboxamide groups in proteins is presented. Most of the ligands are from aspartic and glutamic acid side chains. Water molecules bound to carboxylate anions are known to interact with oxygen lone-pairs. However, metal ions are also found to approach the carboxylate group along the C–O direction. More metal ions are found to be along the *syn* than the *anti* lone-pair direction. This seems to be the result of the stability of the five-membered ring that is formed by the carboxylate anion hydrogen bonded to a ligand water molecule and the metal ion in the *syn* position. Ligand residues are usually from the helix, turn or regions with no regular secondary structure. Because of the steric interactions associated with bringing all the ligands around a metal center, a calcium ion can bind only near the ends of a helix; a metal, like zinc, with a low coordination number, can bind anywhere in the helix. Based on the analysis of the positions of water molecules in the metal coordination sphere, the sequence of the EF hand (a calcium-binding structure) is discussed.

Key words: binding geometry/carboxylic ligand/ligand water/metals/protein secondary structure

Introduction

The carboxylate anion in protein structures plays an important role in binding metal ions, especially calcium, both by direct binding (Martin, 1984), as well as by influencing electrostatic interactions away from the metal center (Linse *et al.*, 1988). Einspahr and Bugg (1981, 1984), have analyzed the geometry of calcium-binding to carboxylate anions in small model compounds. The directionality of metal–carboxylate interactions has been analyzed (Carrell *et al.*, 1988) in terms of the binding of various metal ions with the *syn* and *anti* lone-pair orbitals; metal ions have a preference for the *syn* lone pair on a carboxylate oxygen atom (Figure 1). Similarly, in four out of five complexes of carboxypeptidase A and thermolysin, the metal coordination displays *syn* stereochemistry with respect to the carboxylate group of the inhibitor (Christianson and Lipscomb, 1988). In this paper we analyze the geometry and the *syn* versus *anti* selectivity of metal ions interacting with carboxylate groups in all known protein structures. Besides the side chains of aspartic (Asp) and glutamic (Glu) acids, the C-terminus carboxyl group can also act as a ligand. Although side chains of aspargine (Asn) and glutamine (Gln) cannot exist as anions, they have similar size as those of Asp and Glu respectively, and have been included in this survey. Water molecules are ubiquitous in coordination spheres of many cations and their orientations with respect to carboxylate groups have also been analyzed. Secondary structural features of residues providing ligands to metal ions are analyzed in order to understand how metal binding sites are formed in proteins.

Materials and methods

The analysis is based on atomic coordinates from the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977). Only one protein was used to represent a family of homologous molecules, with the exception of trypsin and its zymogen, trypsinogen and phospholipase A_2, which have two crystal structures with different numbers of metal ions. We have not analyzed recently determined structures (not yet included in the PDB) of many metalloproteins, like α-lactalbumin (Stuart *et al.*, 1986),

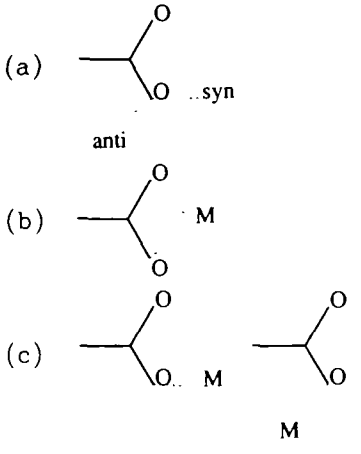

Fig. 1. (a) *Syn* and *anti* directions of lone-pair electrons on a carboxylate oxygen (b) Bidentate mode of binding of a metal ion (c) Unidentate mode of binding of a metal ion along the *syn* and *anti* directions, which have ϕ values (as explained in Figure 2) +60 and −60° respectively. Metals can also be close to the C–O direction (ϕ of 0°).

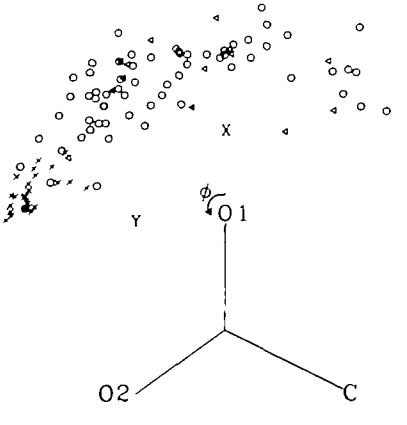

Fig. 2. Projection of metal ions on the carboxylate plane for structures given in Table I. Unidentate and bidentate carboxylate ligands are represented by circles and crosses respectively; carboxamide ligands (Asn and Gln) are represented by triangles. The diagrammatic representation of the spherical polar co-ordinates used to express metal–carboxylate (or carboxamide) interactions is also shown.

Table I. Geometric parameters for metal(M)−carboxylate (and carboxamide) interactions

PDB code[a]	Protein name	Metal[b]	Residue [b]	Structure[c]	Distance (Å)[d]		Angle (°)[d]		Angle(°)[e]	
					M−O1	M−O2	M−O1−C	M−O2−C	ϕ	θ
1BP2	phospholipase A_2, bovine	Ca	D49[f]		2.47	2.66	94 9	86 3	85 1(93.7)	82.0(82 6)
1P2P	phospholipase A_2, porcine	Ca(1)	D49[f]	H(7)	1 98	2.43	104 0	81 7	74.8(98 7)	67 8(72 0)
		Ca(2)	E92	H(3)	2.67		146 2		28 9	71 8
3CNA	concanavalin A	Ca	D10[f]	E	2 09		126.9		41.2	52 6
			N14	C	2 04		121 4		−32.6	38.1
			D19[f]	C	2 26		148.9		23.6	68.9
		Mn	E8[f]	E	2 68		99 4		73 4	34.7
			D10		2.67		95 8		−75 3	23.9
			D19		2.32		133.9		46 0	86.5
5CPA	carboxypeptidase A	Zn	E72	C[g]	2.18	2 31	95 1	87 6	84.9(92 4)	88 7(88 8)
2CPV	parvalbumin B	Ca(1)	D90[f]	C[g]	2.90		121.1		−41.8	48 2
			D92	T	2.00	2 53	103.7	92.9	76.0(87.1)	77.5(81.0)
			D94	S	2.51		122.8		45 9	43 4
			E101[f]	H(3)	2.21	2.27	93.5	103 6	86 5(76.4)	85 6(85.9)
		Ca(2)	D51	C[g]	2 00		170.0		6.2	69.3
			D53	S	2.29		152.3		15 6	84.1
			E59	E	2.69		108 3		65 6	68.4
			E62[f]	H(2)	2.44		101.3		76.5	74 5
1CSE	subtilisin	Ca	Q2	C	2.40		141.9		−32.0	68 0
			D41	T	2 46	2.60	97.7	91 4	82.2(88.6)	84.3(84.6)
			N77	S	2.39		135 1		34 4	57.3
3CLN	calmodulin	Ca(1)	D20[f]	C[g]	2.34		147.1		0 3	58.0
			D22	T	2.42		145.7		14 9	59.0
			D24[f]	S	2.61		135 3		13.3	48.5
			E31	H(3)	2.28	2.38	96.0	90.7	84.0(89.3)	83.3(83.7)
		Ca(2)	D56[f]	C[g]	2.21		153.2		12.4	65 4
			D58[f]	S	2.48		135.0		41 5	69 3
			N60	S	2.46		138.7		30.9	60 9
			E67[f]	H(3)	2.31	2 49	97.5	90.0	82 5(90 0)	89.8(89 8)
		Ca(3)	D93[f]	C[g]	2.14		158 1		14.8	74.4
			D95[f]	T	2.22		128.0		31 9	45.9
			N97	S	2.39		124.4		15.7	36.6
			E104	H(3)	2.32	2 76	106.0	83 1	73 7(97 0)	77.9(79 9)
		Ca(4)	N129	C[g]	2 17		155.4		−2 1	65.5
			D131[f]	S	2.56		117 0		51.4	48.7
			D133[f]	S	2.07		135.7		26.0	52.6
			E140[f]	H(3)	2 32	2.57	97.9	86.5	82.0(93.6)	82.3(83.1)
4TNC	troponin C	Ca(1)	D106[f]	C[g]	2.24		149 4		1 9	58.9
			N108	T	2.43		140 9		32.1	65.6
			D110[f]	S	2.45		117.5		39.1	36.9
			E117	H(3)	2 00	2.49	114.2	89 0	65 8(91.0)	89.1(89.3)
		Ca(2)	D142[f]	C[g]	2 09		150.6		−2 2	60.1
			N144	T	2.40		152.3		21.8	72.2
			D146[f]	S	2 41		127 9		46.6	62.6
			E153[f]	H(3)	2 26	2 68	108.9	90.9	71 0(89 1)	83 1(84.2)
3ICB	Ca-binding protein	Ca(1)	E27[f]	H(3)	2.27	2.51	99 8	89 2	80.2(90.8)	89.2(89.3)
		Ca(2)	D54	C[g]	2.38		150.9		−2 3	61 4
			N56	S	2 45		167.0		2.6	73 9
			D58	S	2.29		131.0		42.4	62.6
			E65	H(3)	2.29	2.61	101.7	86.4	78.3(90.8)	86.4(86.8)
1SGT	trypsin *Streptomyces griseus*	Ca	D165[f]	H(1)	2.43	2.52	92.1	88 6	87.9(91.4)	89.4(89.4)
			E230[f]	E	2.43		140.4		39.3	84.0
3EST	elastase	Ca	E70[f]	C	2.46		113.5		65.1	77.5
			N77	C	2.73		118.6		−43.4	40.7
			E80[f]	C	2.39		137.4		−40 2	73.9
1HMQ	hemerythrin	Fe(1)	E58	H(7)	2.28		123 1		56.6	81.8
			D106	T	2.06		144.5		35 5	89.3
		Fe(2)	E58[f]		2.04		124 7		49.9	60.5
			D106[f]		2.09		133.9		45.8	81 0
3PGK	phosphoglycerate kinase	Mg or Mn	D372	H(1)	2.32		115.3		58.8	55.7
1PPT	pancreatic polypeptide	Zn	N23[h]	H(9)	1.97		112.7		66.9	78.6
2PRK	proteinase K	Ca(1)	D200	C	2.48	2 70	97.3	86.3	82.4(93 8)	78.4(79 3)
		Ca(2)	D260	C	2.49	2 54	93.1	90.9	86 8(89.1)	76 7(77.0)
3PTN	trypsin	Ca	E70	C	2 39		145 4		28 5	77.6
			E80[f]	C	2.28		141 7		−38.1	85.5
4SBV	SBMV	Ca(A)	D138(A)	C	2 35		123 7		−54.6	73 1
			D141(A)	S	2 39		128 1		50 3	74 7
			N259(B)	C	2.02		140.1		39.8	88 8
			L260(B)[i]	C	2 43		154.4		−17.8	71 3
		Ca(B)	D138(B)		1 80		162.7		17.1	87.9

Table I. *Continued*

PDB code[a]	Protein name	Metal[b]	Residue [b]	Structure[c]	Distance (Å)[d]		Angle (°)[d]		Angle(°)[e]	
					M−O1	M−O2	M−O1−C	M−O2−C	ϕ	θ
			D141(B)[f]		2.83	2.84	90 8	90.1	89 1(89.9)	61.7(61.8)
			N259(C)		2 07		168.8		−1.2	78.8
			L260(C)[i]		1.95		171 1		−8.6	87.8
		Ca(C)	D138(C)		2.14		129.3		−48.6	72.9
			D141(C)[f]		2.60		104.8		74.1	69.7
			N259(A)		2.09		149.6		7.4	59.7
			L260(A)[i]		2 38		166 1		−11 5	82 2
2STV	STNV	Ca(1)	E25[f]	C	2 15		172.9		−6.8	88.0
			D194[f,h]	C	2.34		149.3		0.6	59.3
2SOD	superoxide dismutase	Zn(A)	D81(A)	E	1.91		147.5		22.4	66 1
		Zn(B)	D81(B)		2.07		131.0		45.7	69.9
		Zn(C)	D81(C)		2.00		122.8		55 1	71.2
		Zn(D)	D81(D)		1.99		126.8		52.6	79.8
2SNS	staphylococcal nuclease	Ca	D21	S	2.37		150.6		−12.8	63.0
			D40	B	2.59		165.9		−9.3	78.3
3TLN	thermolysin	Ca(1)	D185[f]	C	2.52		132.3		45.7	70.6
			E177	H(4)	2.53	2.87[j]	102.5	85.7	76.9(94.4)	71.4(73.6)
			E190	T	2.40	2.51	94.1	87.5	85 7(92.6)	72 4(73.2)
			D138	H(2)	2.46		133.5		−23.9	48.8
		Ca(2)	D185		2.43		127 9		37.6	47.3
			E177[f]		2.42		124 9		20.1	39.1
			E190[f]		2.40		134.5		−28.2	52.7
		Ca(3)	D57	E	2.23	2 71	104.6	81 5	75.4(98.5)	84.8(85.7)
			D59	S	2 34		141 7		29.8	63 6
		Ca(4)	D200	C	2.24		147.3		32 7	88.9
		Zn	E166	H(8)	2.08	2.66[j]	108.5	77 1	71.4(103.0)	81.2(83.1)
2TMV	TMV	Ca	D116[f]	H(3)	2 29		115.5		53.6	45.2
1PFK	phosphofructokinase	Mg(1)	D103(A)[f]	H(1)	1.95		147.3		32 1	83 7
		Mg(2)	E187(A)	S	2 14		135 6		44.2	84.0
2TGA	trypsinogen	Ca	E70	C	2.23		134.6		37.9	75.6
			E80[f]	C	2 23		142 3		−34.7	72.9
1TON	tonin	Zn	E148[f,h]	S	2.07	2.71[j]	101.5	73 2	75 4(109.1)	52 7(62.4)
2WRP	*Trp* repressor	Na	E60	H(4)	2 75		146.3		−33.7	89.3

[a]1BP2, Dijkstra *et al.* (1981), 1P2P, Dijkstra *et al.* (1983); 3CNA, Hardman and Ainsworth (1972); 5CPA, Rees *et al.* (1983); 2CPV, Moews and Kretsinger (1975), 1CSE, Bode *et al* (1987); 3CLN, Babu *et al.* (1988); 4TNC, Satyshur *et al.* (1988), 3ICB, Szebenyi and Moffat (1986); 1SGT, Read and James (1988); 3EST, Meyer *et al.* (1988); 1HMQ, Stenkamp *et al.* (1983); 3PGK, Watson *et al* (1982); 1PPT, Blundell *et al.* (1981), 2PRK, Betzel *et al.* (1988), 3PTN, Marquart *et al.* (1983); 4SBV, Silva and Rossmann (1985); 2STV, Jones and Liljas (1984); 2SOD, Tainer *et al* (1982), 2SNS, Cotton *et al.* (1979), 3TLN, Holmes and Matthews (1982); 2TMV, Namba *et al.* (1989); 1PFK, Shirakihara and Evans (1988); 2TGA, Marquart *et al.* (1983), 1TON, Fujinaga and James (1987); 2WRP, Lawson *et al.* (1988)

[b]1, 2 or A, B, etc. are used to distinguish different metal atoms or subunits in the same protein molecule.

[c]As defined by Kabsch and Sander (1983): B, residue in isolated β-bridge; C, non-regular structure; E, extended strand, G, 3_{10}-helix; H, α-helix, S, bend, T, H-bonded turn. For a residue in a helix, its position from the nearest end of the helix is given in parenthesis. When the same residue appears in the table more than once, the secondary structural feature is assigned to one entry only.

[d]O=OD in Asp, Asn; OE in Glu, Gln. C=CG in Asp, Asn; CD in Glu, Gln.

[e]As defined in the text. Values inside the parentheses are those involving O2 for bidentate ligands.

[f]Atom labels O1 and O2, as given in PDB, have been interchanged.

[g]The residue is in between a helix on one side and two residues, both with structure T or S, on the other side; in many PDB files, such a residue has been considered as the 1st residue of the helix. Consequently, this has been assumed to have the structure H in the text.

[h]Crystallographic (or noncrystallographic)-symmetry-related position.

[i]The carboxyl group at the C-terminus of the polypeptide chain is the ligand.

[j]Not classified as a ligand atom in the original publication.

galactose-binding protein (Vyas *et al.*, 1987), alkaline phosphatase (Sowadski *et al.*, 1985), DNase I (Suck and Oefner, 1986), the large fragment of DNA polymerase I (Ollis *et al.*, 1985), phospholipase C (Hough *et al.*, 1989) and α-amylase (Buisson *et al.*, 1987).

As in the case of small molecule structures (Einspahr and Bugg, 1981), metal−carboxylate interactions (Figure 1) can be classified as unidentate (metal interacting with only one oxygen atom) or bidentate (metal chelation by both the oxygen atoms of a carboxylate group). All carboxyl groups with either oxygen atom within 3 Å of a metal ion have been used. In many cases the numbering of the oxygen atoms had to be changed. The oxygen atom (OD1 or OD2 in Asp, OE1 or OE2 in Glu) coordinating to the metal ion, or when both the oxygens are involved, the one closer to the cation, has been named O1 and the other O2. In cases where the same carboxylate group bridges two different metal ions, each oxygen atom is considered as O1 while expressing the geometry of the metal ion it binds to.

The relative position of the metal ion was expressed in the coordinate system depicted in Figure 2. The ligand O1 atom is at the origin; the x axis is along the direction of the C−O1 bond; the carboxylate group is coincident with the xy plane, such that O2 is situated in the $-x,+y$ quadrant; the z axis is along the normal to the carboxylate plane. The geometry of metal−ligand bonding is given by the following spherical polar coordinates: the M−O1 distance; the acute angle θ between the M−O1 direction and the z axis; the angle ϕ between the x axis and the projection of the O1−M direction on the xy plane. The secondary structural features of all ligand residues were defined according to the method of Kabsch and Sander (1983).

Results and discussion

All the geometrical parameters are recorded in Table I. Various groups providing carboxylate ligands are Asp (42 cases), Glu (29), Asn (11), Gln (1) and the C-terminus (1). In both the virus structures, 4SBV and 2STV (proteins are referred to by their PDB codes as given in Table I), the C-terminal end is very close to the metal ion—in the former it is a direct ligand and in the latter the bonding is through a water molecule. In both the cases, the residue preceding the terminal residue is a ligand.

Orientation of the metal ion

The deviation of metal ions from the peptide plane, as given by the angle θ, is shown in Figure 3. Most of the θ values for unidentate cases are in the range $60-90°$, indicating that there can be some deviation of metal ions from the carboxylate plane. This deviation is also known to be dependent on the type of metal ions (Carrell *et al.*, 1988). Bidentate cases are more planar and metals do not usually deviate by $>10°$ from the plane (θ values $80-90°$) as has been observed in small molecular structures (Einspahr and Bugg, 1981).

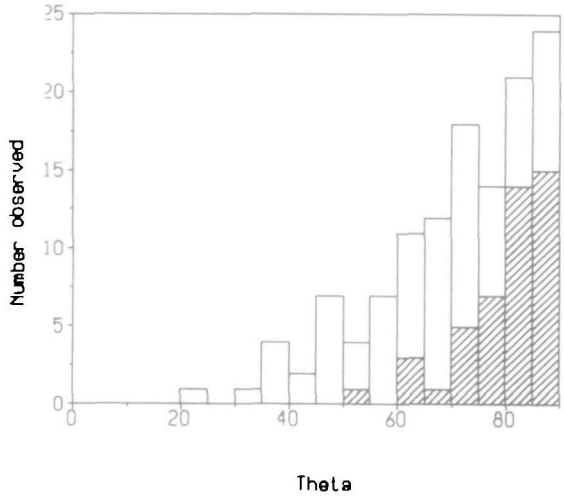

Fig. 3. Histogram of θ (deg.) showing the deviation of metal ions from the carboxylate (and carboxamide) plane. The shaded area corresponds to bidentate cases (angles involving both the oxygens have been used)

Table II. Hydrogen bonding involving the atom O2 of Asp and Glu carboxylate side chains with negative ϕ values (Table I)[a]

PDB code	Residue	Bonded to[b]	Distance (Å)
2CPV	D90	N,95	2.61
4TNC	D142	N,147	2.67
3ICB	D54	N,59	2.96
3EST	E80	N,79	3.06
3PTN[c]	E80	N,78	2.99
4SBV	D138(A)	ND2,217(A)	3.07
2STV	E25	N,63[d]	2.65
2SNS[e]	D21	OG1,22	2 08
3TLN	D138	N,139	3.04
2WRP	E60[f]	NH1,63	3.24

[a]The bridging residues (D10 in 3CNA and E190 in 3TLN) are not included; binding two metals by the same carboxylate group causes one metal to have a negative ϕ value. However, other bridging groups (E58 and D106 in 1HMQ, and E177 and D185 in 3TLN) have positive ϕ values for both the metal ions.
[b]The protein residue with the shortest hydrogen bond is given. Usually there are other hydrogen bonds
[c]2TGA is similar to this
[d]At a position related by the 3-fold rotation of the virus particle
[e]Residue D40 does not have any hydrogen-bonded neighbor It is to be noted that Ca is in the active site of the enzyme and binds to the phosphodiester oxygen of the substrate
[f]The metal ion binds both to the main-chain carbonyl as well as the side chain carboxylate anion of the residue: this causes ϕ to be negative.

The distribution of metal ions in the plane of the ligand group is given by the angle ϕ and is shown in Figure 2. The range of this angle for bidentate cases is 65.8 (in 4TNC) to 109.1° (in 1TON). The bidentate binding in small molecule structures is very symmetric (the two M−O distances nearly equal), whereby the two ϕ values are close to 90° (Carrell *et al.*, 1988). The unidentate examples display a wider distribution of ϕ values; the two limits observed are −75.3 (in 3CNA) and 76.5° (in 2CPV). However, metal ions are broadly clustered in three groups: around the *syn* lone pair ($\phi=60°$), close to the *anti* lone pair ($\phi=-60°$) and along the C−O bond ($\phi=0°$). In contrast, water molecules bound to Asp and Glu side chains usually approach along the lone-pair direction and a very few are along the C−O direction (Thanki *et al.*, 1988). In small molecule structures, calcium ions show very little tendency to be collinear with the C−O bond (Einspahr and Bugg, 1981) and similar observations have been made for many other cations (Carrell *et al.*, 1988). The linear interaction in protein structures may be caused by the need to achieve the optimum hydrogen-bonding environ-

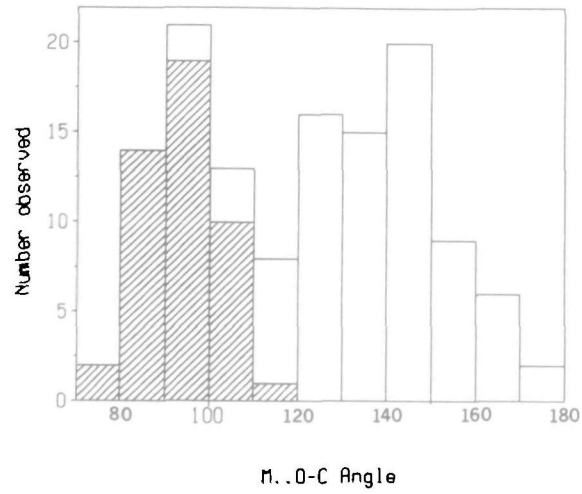

Fig. 4. Histogram showing the distribution of M−O−C angles (°) for all structures in Table I. The shaded area corresponds to bidentate cases (angles involving both the oxygens have been used).

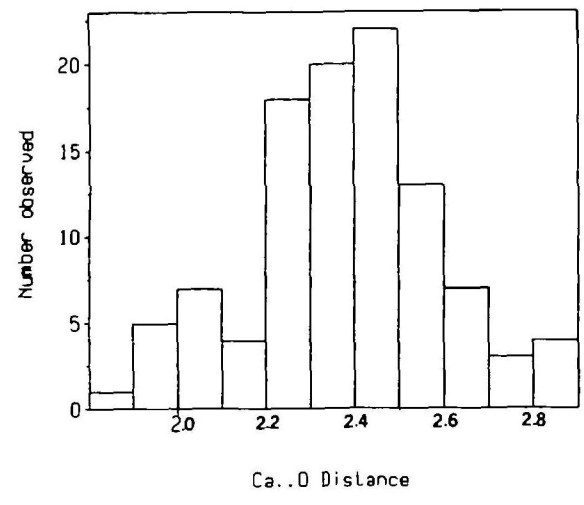

Fig. 5. Histogram showing the distribution of Ca−O distances (Å)

ment by the ligand carboxylate group, and also by the active site geometry (as in 2SNS, where the cation also binds to the phosphodiester oxygen of the substrate).

It has been proposed that the *syn* lone pair on a carboxylate oxygen atom is more basic than the one at the *anti* position (Gandour, 1981) and various calculations have been performed to assess the basicity of these two lone-pair directions (Peterson and Csizmadia, 1979; Wiberg and Laidig, 1987; Li and Houk, 1989). Chelating agents have been synthesized so that the metal can bind along the *syn* orientation (Marshall *et al.*, 1988). Although there are more metal ions along the *syn* than the *anti* lone-pair direction (Table I and Figure 2), this seems to be influenced by the position of the ligand water molecule (see later). Carboxylate groups with metal ions in the *anti* region (−ve ϕ values) usually have the O2 atom involved in hydrogen bonding (Table II). In most cases this bond is to the N atom of a residue close along the sequence and thereby imparting stability to the main-chain conformation around the metal center.

The carboxylate anion, by utilizing both its oxygens, can bind to two different metal ions. Such bridging carboxylate groups are found in 3CNA, 1HMQ, 3TLN and also in some recent structures like alkaline phosphatase (Sowadski *et al.*, 1985) and phospholipase C (Hough *et al.*, 1989). Bridging residues D19 in 3CNA, E58 and D106 in 1HMQ, and E177 and D185 in 3TLN have +ve ϕ values for both the metal ions; however, D10 in 3CNA and E190 have one +ve ϕ and one −ve ϕ value.

Bond lengths and angles

As metal ions are distributed over a wide range with respect to the carboxylate plane (Figure 2), M−O−C angles are also expected to vary over a wide range (Figure 4). For bidentate cases, the angles lie in the range 80−110° and unidentate examples display a range of 110−170°. Metal ions interact with the peptide carbonyl group close along the C=O direction; and as a result, have M−O=C angles in the range 140−170° (Chakrabarti, 1990).

The metal−oxygen distance is known to depend on the coordination number of the cation (Chakrabarti *et al.*, 1981; Einspahr and Bugg, 1984) and also on the M−O−C angle (Einspahr and Bugg, 1981). The histogram in Figure 5 shows the distribution of Ca−O distances in various metalloproteins. It is unusual in small molecule structures to have these distances <2.0 or >2.8 Å (Einspahr and Bugg, 1984). The presence of a few such bond lengths is due to the lesser accuracy of protein structures and systematic errors that might have been introduced during model building and refinement.

Secondary structural preference

A metal binding to a ligand in the middle of a helix will find the approach of any other protein ligand sterically inhibited by the helix at one side. As a result, a carboxylate side chain from the middle of a helix can bind to a metal ion either with a low coordination number or with water molecules in its coordination sphere, occupying positions not accessible to other protein ligands. This is indeed the case, as can be seen in Figure 6 by the two types of metal ions binding at two regions of the same helix in thermolysin. In nature, zinc is often four or five coordinated (Bertini *et al.*, 1985). In 3TLN a Zn ion is bound to a Glu residue in the middle of a helix; it is bound to only two other His residues and a water molecule. Similarly, a Zn bound to the middle of a helix in 1PPT is bound to only two other protein residues (one of which binds in a bidentate fashion) and a water. It is not unusual for a calcium ion to have a coordination number as high as eight (Brown, 1988), and it can bind to a helix only near its ends. As can be seen from Table I, most of the metal ions (excluding Zn) bound to a helix are within four residues from an end of the helix. The exceptions are phospholipase A_2 (where the metal binds to only three other protein ligands and two water molecules; binding in a bidentate fashion, the helix carboxylate group itself provides two ligand atoms) and hemerythrin (where all ligands are provided by adjacent helices). By the same steric consideration, Ca ion is unlikely to bind to the middle of a β-sheet. It usually binds to the end of a strand (D10 in 3CNA, E230 in 1SGT, D81 in 2SOD) or to a peripheral strand (D57 in 3TLN). Magnesium is usually coordinated by six ligands (Brown, 1988) and is expected to behave like calcium

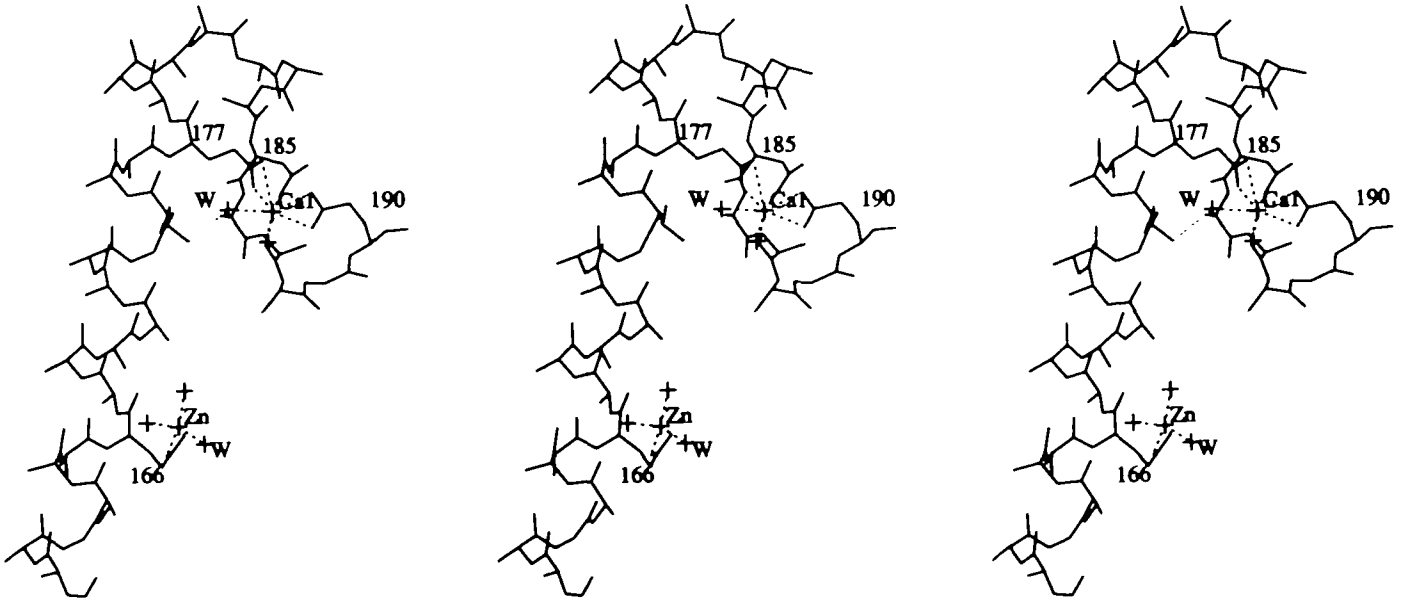

Fig. 6. Stereoview showing the binding of a calcium ion near the C-terminus of a helix and a zinc ion near the middle of the same helix (except for the ligand groups side chain atoms beyond C_β are not shown) in thermolysin. Ligand residues from the helix and the adjacent loop are indicated by their sequence numbers; positions of other protein ligands are shown as crosses. Water ligands are represented by W. The water molecule bound to the calcium ion is also hydrogen bonded (shown by thin dashed line) to T174 hydroxyl group. Metal−ligand bonds are represented by dashed lines.

in its binding characteristics with respect to a helix or a β-sheet. In the recently published structure of D-xylose isomerase (Henrick *et al.*, 1989), a Mg ion is coordinated by four carboxylate groups which are positioned at the C-terminal ends of four β-strands; Mg(1) in 1PFK is bound to a residue at the N-terminus of a helix.

Table I shows that most of the ligand residues are from helix, turn or regions with non-regular structure. In fact Ca-binding ligands in the EF hand (a Ca-binding structure resembling the E helix−loop−F helix domain of parvalbumin) (Kretsinger and Nockolds, 1973) are from two helices and intervening turns. The three ligands two residues apart along the sequence in proteins with the EF hand have structures helix(H)−turn(S or T)-turn(S). With only two such ligands, both can be from a strand (residues 8 and 10 in 3CNA), or one from a strand and the other from the adjacent turn (residues 57 and 59 in 3TLN). In three cases (2CPV, 3EST and 4SBV), ligands three residues apart bind to the same metal ion.

Position of the ligand water

Figure 6 shows that the water molecule binding the Ca ion is 'below' the side chain of the ligand glutamic acid, occupying a position that cannot possibly be taken by any other protein ligand. To hold the water molecule in this position through hydrogen bonding there is a Thr three residues preceding the Glu residue. Water seems to be an important constituent of the coordination sphere of metal ions with a bound Glu residue. The position of the water molecule with respect to the carboxylate side chain of the Glu residue usually follows one of the two types mentioned in Figure 7. In the type I mode, one oxygen of the carboxylate group binds the metal ion, whereas the other forms a hydrogen bond with the ligand water molecule. In the type II mode, the metal ion is coplanar with the carboxylate group, being bound to the carboxylate oxygens in a bidentate fashion, and the water molecule is on the top of the metal ion with reference to the plane; both the water molecules in Figure 6 show this geometry. The water molecule is quite close to the Glu residue (Table III), being hydrogen bonded to it (in type I binding) or to a residue close to it. It is interesting to note that Ca(2) in 2CPV does not have any water in its coordination sphere; instead, the hydoxyl side chain of a Ser residue acts as a ligand, simultaneously forming a hydrogen bond with a ligand Glu (Figure 8).

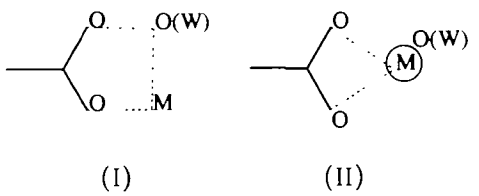

Fig. 7. Two common types of binding of a metal ion (M) involving a carboxylate anion and a water (W) molecule. (I) M and O(W) form a five-membered ring with the atoms in the carboxylate anion. Usually the ring is quite planar (The putative hydrogen in the hydrogen bond between carboxylate and water cannot be located in the X-ray analyses of proteins. If the hydrogen is considered, the structure is actually a six-membered ring.) (II) The carboxylate anion binds the cation in the bidentate mode and the M−O(W) vector is approximately perpendicular to the plane formed by the metal ion with the atoms in the carboxylate group

Type I binding, which was first observed in β-trypsin (Bode and Schwager, 1975), seems to be energetically favorable. Even in structures with a water involved in a type II binding with respect to a Glu residue, there is usually an Asp that binds the water in the type I mode (Table IV). As in the coordination sphere of Ca(2) in 2CPV, the water molecule can be substituted by a hydroxyl group from a side chain (in 3TLN) or a substrate (2TMV).

The bidentate mode of binding is likely to be very stable (Christianson and Lipscomb, 1988). However, in the presence of water (or any other ligand hydroxyl group), the type I binding seems to be equally favorable. Although not more than one carboxylate group binds in a bidentate fashion to a given metal center [exception, Ca(1) in 2CPV, Ca(1) in 3TLN], both the ligand water molecules in α-lactalbumin (Stuart *et al.*, 1986) appear to be associated with two Asp residues to make two type I bindings. This mode of binding owes its stability to the five-membered ring that the carboxylate group forms with a water and the metal ion

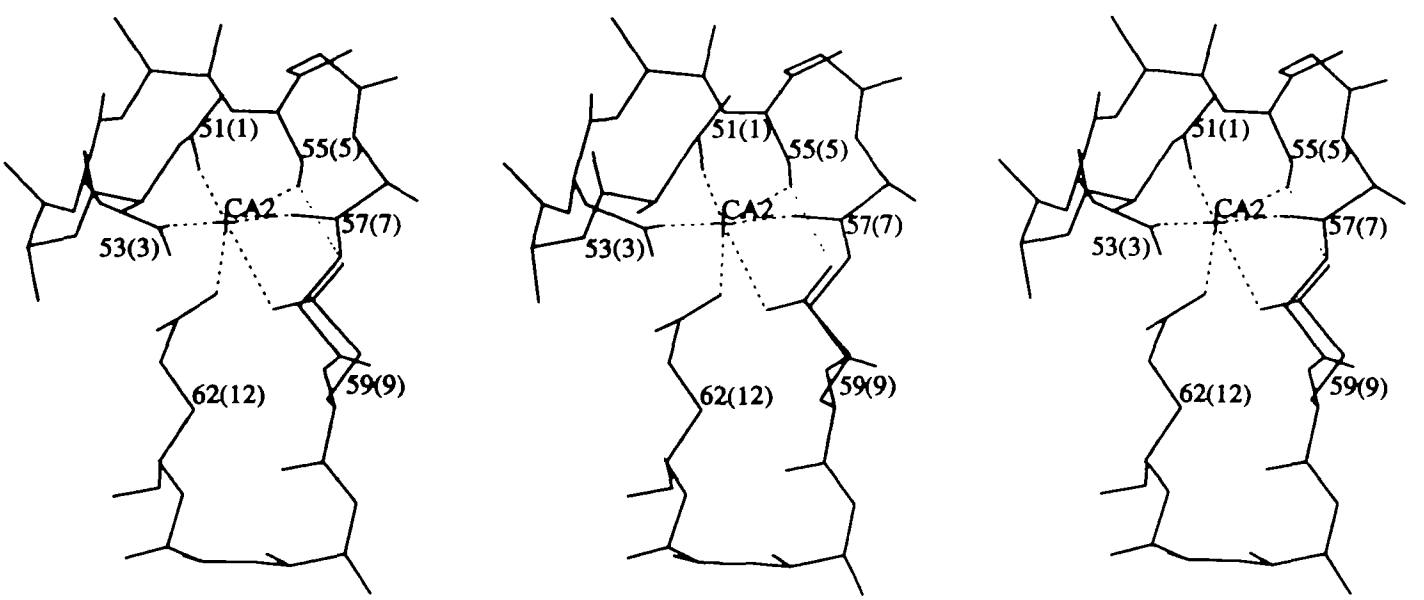

Fig. 8. Stereoview showing the coordination sphere of Ca(2) in parvalbumin B (except for the ligand groups side chain atoms beyond C_β are not shown). Ligand residues are indicated by their sequence numbers, the position of each ligand in the 12-residue loop is given in parentheses. Metal ligand bonds are represented by dashed lines. The hydrogen bond (shown by a thin dashed line) between the S55 hydroxyl group and the E59 carboxylate group creates a five-membered ring involving the metal ion, the hydroxyl oxygen and the carboxylate anion.

(Figure 7). Cyclic structures are known to be very stable and are observed when carboxylate groups dimerize (Dauber and Hagler, 1980) and have been proposed in the interaction involving a water molecule and a carboxylate anion (Cybulski and Scheiner, 1989). It is to be noted that when the water is needed in catalysis, as in carboxypeptidase A (Christianson *et al.*, 1987) and thermolysin (Matthews, 1988), it is not involved in any type I binding. In both these cases, the water molecule is in type II mode with respect to the ligand Glu, which is not known to have any other catalytic role. The type I binding involving an acidic group from one molecule and a hydroxyl group from another molecule (as in 2TMV, Table IV) can impart stability to the interaction between the two molecules. This may play an important role in protein–nucleic acid interaction.

In type I binding, the metal ion is close to the *syn* lone-pair orbital of the carboxylate anion. Cations, which are neither bound in a bidentate mode, nor are involved in the type I binding, show no significant preference for the *syn* lone pair. Assuming cations within 30° of $\phi = 60°$ to be along the *syn* direction and those within 30° of $\phi = -60°$ to be along the *anti* direction and not taking bridging carboxylates into account, in 20 cases the binding is along the *syn* direction (out of which 13 are associated with the type I binding) and in six cases the binding is along the *anti* direction. Theoretically, carboxamide can also form an isomeric type I structure by serving as a hydrogen bond donor to one lone pair on water; the other lone pair on water can bind to a metal, which in turn is bound to the oxygen of the carboxamide. However, considering distances between the water oxygen and the carboxamide oxygen in various structures, it appears that Asn and Gln with carboxamide side chains do not bind metal ions in the type I mode. Additionally, these groups cannot be involved in bindentate binding and as a result, bind metals along the C–O direction or in orientations with small ϕ values (+ve or −ve).

Table III. Position of the water molecule[a] with respect to glutamic acid in the coordination sphere of metal ions

PDB code	Metal	Glu residue	Distance (Å)[b] water–Glu	Water hydrogen-bonded to[c]	Orientation[d]
3CNA	Mn	8	2.7	OE2,8	I
5CPA	Zn	72	3.2	OE1,270	II
2CPV	Ca(1)	101	3.6	OD2,92	II
	Ca(2)	59	2.4[e]	OE2,59	I
3CLN	Ca(1)	31	3.5	OD1,22	II
	Ca(2)	67	2.8	OD1,64	II
	Ca(3)	104	2.8	OD2,95	II
	Ca(4)	140	3.5	OD1,137	II
4TNC	Ca(1)	117	3.5	OD2,114	II
	Ca(2)	153	3.3	OD1,150[f]	II
3ICB	Ca(1)	27	3.5	OE1,60	X
	Ca(2)	65	3.1	OD2,58	II
1SGT	Ca	230	2.8	OE2,230	I
3EST	Ca	70	3.1	OE2,70	I
3PTN	Ca	70	2.8	OE2,70	I
		80	2.9	OE1,77	X
2STV	Ca(1)	25	3.3	O,195[g]	X
3TLN	Ca(1)	177	3.3	OG1,174	II
	Ca(2)	190	3.3	OD1,191	X
	Zn	166	2.9	OH,157	II
1PFK	Mg(2)	187	2.8	OE2,187	I

[a]Ca(2) in 2CPV does not have any ligand water. No water has been located in the coordination sphere of Ca(2) in 1P2P.
[b]The shorter of the two distances from the ligand water to either OE1 or OE2 of Glu is given.
[c]The residue closest to the ligand Glu is given. Usually the water is also hydrogen-bonded to other protein or non-protein groups.
[d]As given in Figure 7. Cases not belonging to either of these two classes are represented by X.
[e]There is no water molecule in the coordination sphere; however, its position is occupied by the hydroxyl group of the residue S55.
[f]Bridged to the ligand water via another water molecule. However, in another troponin C structure, 5TNC (Herzberg and James, 1985), there is no bridging water.
[g]At a position related by the 3-fold rotation of the virus particle.

Table IV. Aspartic acid residues showing type I binding[a] of water molecules and metal ions

PDB code	Metal	Asp residue	Distance (Å)[b] water–Asp
1BP2[c]	Ca	49	3.22
3CNA	Ca	10	3.11
2CPV[c]	Ca(1)	92	2.29
3CLN	Ca(1)	22	3.16
	Ca(2)	58	2.65
	Ca(3)	95	2.74
	Ca(4)	131	2.74
4TNC	Ca(1)	110	2.75
	Ca(2)	146	2.64
3ICB	Ca(2)	58	2.81
2PRK[c]	Ca(1)	200	3.08
3TLN	Ca(3)	59	2.72
	Ca(4)	200	2.66[d]
2TMV	Ca	116	2.75[e]

[a]As depicted in Figure 7. Cases where the binding is not of this type are: residues 165 in 1SGT, 260 in 2PRK and 57 in 3TLN (all showing type II binding). The bridging residue 185 in 3TLN, residues 103 in 1PFK (where phosphate oxygens from two substrates bind to the same metal ion) and 194 in 2STV (where the water is close to the carboxyl group of the C-terminus) are involved in binding patterns that are different from those mentioned. The PDB file for 2SNS does not include any water molecule.
[b]The five-membered ring in type I binding is distorted in cases where the distance is >3.0 Å.
[c]Although in the type I mode the metal binding by the carboxylate group is in the unidentate fashion, it is bidentate in this case.
[d]The place of the water molecule is taken by the hydroxyl group of the residue T194. As a result, the water binding does not correspond to any pattern.
[e]The place of the water molecule is taken by the hydroxyl group of the substrate nucleotide.

Table V. Position of the aspartic acid (glutamic acid in one case) residue (marked by *) showing type I binding[a] of water molecule and Ca ion in the EF hand[b] (amino acids at positions 3, 5 and 9 in the 12-residue Ca-binding loop[c] are given)

PDB code	Metal	Position 1 in the sequence	Amino acid at position		
			3	5	9
2CPV	Ca(1)	90	Asp*	Asp	Gly
	Ca(2)	51	Asp	Ser[d]	Glu*
3CLN	Ca(1)	20	Asp*	Asp	Thr
	Ca(2)	56	Asp*	Asn	Asp
	Ca(3)	93	Asp*	Asn	Ser
	Ca(4)	129	Asp*	Asp	Asn
4TNC	Ca(1)	106	Asn	Asp*	Asp
	Ca(2)	142	Asn	Asp*	Asp
3ICB	Ca(2)	54	Asn	Asp*	Ser

[a]As described in Figure 7.
[b]The loop binding Ca(1) in 3ICB is not included, as its conformation is substantially different from that of the EF hand (Szebenyi and Moffat, 1986).
[c]An illustration is given in Figure 8.
[d]The hydroxyl group from its side chain substitutes for the water molecule and the Glu side chain at position 9 is hydrogen bonded to it.

Sequence in the EF hand

The allowed sequence in the 12-residue loop constituting the EF hand has been discussed (Tufty and Kretsinger, 1975; Herzberg and James, 1985; Szebenyi and Moffat, 1986). It has been suggested (Szebenyi and Moffat, 1986) that the 3rd and the 5th positions in the loop are expected to be Asp or Asn. As shown in Table V, an Asp residue in either of these positions participates in the type I binding. As Asn cannot be associated with this type of binding involving a water molecule, it is unlikely that it can occupy both the positions in the EF hand. A Glu residue at position 9 acts as a ligand only in the coordination sphere of Ca(2) in 2CPV (Figure 8). This is the residue which is involved in the type I binding with the hydroxyl group of the Ser residue in position 5. It appears that a Ser residue at position 5 should favor a Glu at position 9; however, it is also possible that a slight adjustment in the loop conformation will bring the Ser hydroxyl group to hydrogen bond with the ligand Asp residue at position 1. Except the residue E62 in the coordination sphere of Ca(2) in 2CPV, all ligand Glu at position 12 in the loop binds in the bidentate mode—consequently, substituting the Glu by an Gln should affect the metal binding.

Various types of structural motifs that bind metal ions have been proposed (Kretsinger and Nockolds, 1973; Stuart *et al.*, 1986; Vyas *et al.*, 1987). In this analysis we have examined the binding characteristics of individual ligand groups, instead of the whole motif. Metal–carboxylate interactions in various structures have been put on the same general basis and the important role water plays in orienting carboxylate anions in the coordination sphere of metal ions has been established. The preferred geometry of such interactions should be useful in understanding the binding affinities of various proteins for metal ions (Sekharudu and Sundaralingam, 1988; Szebenyi and Moffat, 1986). In recent years site-directed mutagenesis has been used to introduce metal-binding sites in some proteins and thereby to enhance their stability (Serpersu *et al.*, 1987; Pantoliano *et al.*, 1988; Kuroki *et al.*, 1989). Secondary structural characteristics of ligand residues, as found out in this analysis, would assist in introducing suitable ligands at those parts of the protein structure that are likely to satisfy the geometric requirements of metal binding.

Acknowledgement

I am grateful to Prof. D.Rees for discussion

References

Babu,Y.S., Bugg,C.E and Cook,W.J. (1988) *J. Mol. Biol.*, **204**, 191–204.

Bernstein,F.C., Koetzle,T F , Williams,G.J.B., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J R., Kennard,O., Shimanouchi,T. and Tasumi,M (1977) *J Mol Biol.*, **122**, 535–542.

Bertini,I., Luchinat,C. and Monnanni,R. (1985) *J Chem. Education*, **62**, 924–927.

Betzel,C., Pal,G.P. and Saenger,W (1988) *Eur. J. Biochem.*, **178**, 155–171.

Blundell,T.L., Pitts,J.E., Tickle,I.J., Wood,S.P. and Wu,C.-W. (1981) *Proc Natl. Acad. Sci. USA*, **78**, 4175–4179.

Bode,W. and Schwager,P. (1975) *FEBS Lett.*, **56**, 139–143

Bode,W., Papamokos,E. and Musil,D. (1987) *Eur J. Biochem.*, **166**, 673–692.

Brown,I.D (1988) *Acta Crystallogr., B*, **44**, 545–553.

Buisson,G., Duée,E., Haser,R. and Payan,F. (1987) *EMBO J* , **6**, 3909–3916

Carrell, C.J., Carrell,H.L., Erlebacher,J and Glusker,J.P. (1988) *J. Am. Chem. Soc.*, **110**, 8651–8656.

Chakrabarti,P. (1990) *Biochemistry*, **29**, 651–658.

Chakrabarti,P., Venkatesan,K. and Rao,C.N R (1981) *Proc. R. Soc. Lond., Ser. A*, **375**, 127–153

Christianson,D W. and Lipscomb W.N. (1988) *J Am Chem. Soc* , **110**, 5560–5565

Christianson,D.W., David,P.R and Lipscomb,W N (1987) *Proc Natl Acad Sci. USA*, **84**, 1512–1515

Cotton,F A , Hazen,E E.,Jr and Legg,M.J. (1979) *Proc. Natl. Acad. Sci USA*, **76**, 2551–2555.

Cybulski,S M and Scheiner,S. (1989) *J. Am Chem Soc* , **111**, 23–31

Dauber,P. and Hagler,A T. (1980) *Acc. Chem Res.*, **13**, 105–112

Dijkstra,B.W , Kalk,K H., Hol,W.G J and Drenth,J. (1981) *J Mol. Biol* , **147**, 97–123

Dijkstra,B.W , Renetseder,R , Kalk,K.H., Hol,W.G.J and Drenth,J. (1983) *J. Mol. Biol.*, **168**, 163–179.

Einspahr,H. and Bugg,C.E. (1981) *Acta Crystallogr., B*, **37**, 1044–1052

Einspahr,H and Bugg,C.E. (1984) In Sigel,H. (ed), *Metal Ions in Biological Systems*. Marcel Dekker, New York, Vol. 17, pp. 51–97

Fujinaga,M. and James,M N.G. (1987) *J. Mol. Biol.*, **195**, 373–396.

Gandour,R D. (1981) *Bioorg. Chem.*, **10**, 169–176.

Hardman,K.D. and Ainsworth,C.F (1972) *Biochemistry*, **11**, 4910–4919

Henrick,K , Collyer,C A and Blow,D.M. (1989) *J. Mol Biol* , **208**, 129–157

Herzberg,O. and James,M N.G. (1985) *Biochemistry*, **24**, 5298–5302

Holmes,M.A and Matthews,B.W (1982) *J. Mol Biol* , **160**, 623–639

Hough,E , Hansen,L.K., Birknes,B., Jynge,K., Hansen,S., Hordvik,A., Little,C , Dodson,E. and Derewenda,Z (1989) *Nature*, **338**, 357–360.

Jones,T A. and Liljas,L. (1984) *J. Mol. Biol* , **177**, 735–767.

Kabsch,W and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637

Kretsinger,R H. and Nockolds,C.E. (1973) *J. Biol. Chem.*, **248**, 3313–3326.

Kuroki,R., Taniyama,Y., Seko,C., Nakamura,H., Kikuchi,M. and Ikehara,M. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6903–6907

Lawson,C.L., Zhang,R., Schevitz,R.W., Otwinowski,Z., Joachimiak,A. and Sigler,P.B (1988) *Proteins*, **3**, 18–31.

Li,Y. and Houk,K.N. (1989) *J. Am. Chem. Soc.*, **111**, 4505–4507.

Linse,S., Brodin,P., Johansson,C., Thulin,E., Grundström,T. and Forsén,S. (1988) *Nature*, **335**, 651–652.

Marquart,M., Walter,J., Deisenhofer,J , Bode,W. and Huber,R. (1983) *Acta Crystallogr , B*, **39**, 480–490.

Marshall,L., Parris,K., Rebek,J.,Jr, Luis,S.V. and Burguete,M I (1988) *J. Am. Chem. Soc* , **110**, 5192–5193

Martin,R.B (1984) In Sigel, H (ed.), *Metal Ions in Biological Systems*. Marcel Dekker, New York, Vol. 17, pp 1–49.

Matthews,B W. (1988) *Acc Chem. Res.*, **21**, 333–340.

Meyer,E., Cole,G., Radhakrishnan,R. and Epp,O. (1988) *Acta Crystallogr , B*, **44**, 26–38

Moews,P.C and Kretsinger,R.H. (1975) *J. Mol. Biol.*, **91**, 201–228

Namba,K., Pattanayek,R. and Stubbs,G. (1989) *J. Mol. Biol.*, **208**, 307–325

Ollis,D.L., Brick,P., Hamlin,R , Xuong,N.G. and Steitz,T.A. (1985) *Nature*, **313**, 762–766

Pantoliano,M.W., Whitlow,M., Wood,J.F , Rollence,M.L., Finzel,B.C., Gilliland,G.L , Poulos,T.L. and Bryan,P.N (1988) *Biochemistry*, **27**, 8311–8317.

Peterson,M.R. and Csizmadia,I.G (1979) *J. Am. Chem. Soc.*, **101**, 1076–1079.

Read,R.J. and James,M.N.G (1988) *J. Mol. Biol.*, **200**, 523–551

Rees,D.C., Lewis,M. and Lipscomb,W.N. (1983) *J. Mol. Biol.*, **168**, 367–387

Satyshur,K.A., Rao, S T., Pyzalska,D., Drendel,W., Greaser,M. and Sundaralingam,M. (1988) *J. Biol. Chem* , **263**, 1628–1647

Sekharudu,Y.C and Sundaralingam,M (1988) *Protein Engng*, **2**, 139–146.

Serpersu,E.H , Shortle,D. and Mildvan,A.S. (1987) *Biochemistry*, **26**, 1289–1300.

Shirakihara,Y and Evans,P.R. (1988) *J Mol. Biol.*, **204**, 973–994

Silva,A.M. and Rossmann,M.G (1985) *Acta Crystallogr., B*, **41**, 147–157

Sowadski,J.M., Handschumacher,M.D., Murthy,H.M K., Foster,B.A. and Wyckoff,H.W. (1985) *J. Mol. Biol.*, **186**, 417–433

Stenkamp,R E., Sieker,L C. and Jensen,L H (1983) *Acta Crystallogr., B*, **39**, 697–703.

Stuart,D.I., Acharya,K.R., Walker,N.P.C , Smith,S G., Lewis,M and Phillips,D.C. (1986) *Nature*, **324**, 84–87.

Suck,D. and Oefner,C. (1986) *Nature*, **321**, 620–625

Szebenyi,D.M.E. and Moffat,K. (1986) *J. Biol. Chem.*, **261**, 8761–8777.

Tainer,J A , Getzoff,E.D., Beem,K.M., Richardson,J.S. and Richardson,D.C. (1982) *J. Mol. Biol.*, **160**, 181–217.

Thanki,N., Thornton,J.M. and Goodfellow,J M. (1988) *J. Mol. Biol.*, **202**, 637–657.

Tufty,R M. and Kretsinger,R.H. (1975) *Science*, **187**, 167–169.

Vyas,N.K., Vyas,M N and Quiocho,F.A. (1987) *Nature*, **327**, 635–638

Watson,H.C , Walker,N.P.C., Shaw,P.J., Bryant,T.N., Wendell,P L , Fothergill,L.A., Perkins,R E , Conroy,S.C., Dobson,M.J , Tuite,M F., Kingsman,A J., and Kingsman,S.M (1982) *EMBO J.*, **1**, 1635–1640.

Wiberg,K B and Laidig,K.E. (1987) *J Am. Chem Soc.*, **109**, 5935–5943.

Received on March 1, 1990, accepted on June 15, 1990