Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The geometry of metal–ligand interactions relevant to proteins

Marjorie M. Harding

Structural Biochemistry Group, Institute of Cell and Molecular Biology, University of Edinburgh, Michael Swann Building, Edinburgh EH9 3JR, Scotland

Correspondence e-mail:
marjorie.harding@ed.ac.uk

Received 1 March 1999
Accepted 27 May 1999

Geometrical data which could be of relevance in the structure determination, structure refinement, assessment or understanding of metalloproteins have been extracted from the Cambridge Structural Database (CSD). The CSD contains crystallographic data from 'small-molecule' structures determined by X-ray or neutron diffraction to an accuracy much better than that of most current protein structure determinations. The structures selected have a crystallographic *R* factor $\leq$ 0.065 and contain Ca, Mg, Mn, Fe, Cu or Zn interacting with ligands which are analogues of the amino-acid side chains commonly found in proteins; they include carboxylate groups, alcohols, phenolates, thiolates, imidazole groups and also water molecules. For each pair, the mean metal–donor-atom distance, the sample standard deviation and the range of observed values are tabulated, using ~4500 observations in all. Where practicable, subsets with different coordination numbers and/or oxidation states are given. Also included are inter-bond angles at the ligand donor atom, the orientation of carboxylate and imidazole groups with respect to the metal–donor-atom bond and some other aspects of ligand geometry. Thus, for example, target distances and their standard deviations could be easily looked up for the validation of a metalloprotein structure or for use in restrained refinement with low-resolution data.

1. Introduction

Metalloproteins occur widely and have many important functions. In some, the metal atom or ion is a part of the active site for a catalytic process; in others, the metal appears to have a role in maintaining structure. The present study is concerned with the geometry of the interaction of the metal atom with the ligand groups around it, using the accurate information which can be obtained about this geometry from the Cambridge Structural Database (CSD; Allen & Kennard, 1993*a*,*b*). The CSD contains the results of ~190000 structures, including organic molecules and metal complexes. All have been determined by X-ray or neutron diffraction, in many cases with data in which individual atoms are well resolved ($d_{\rm min}$ = 0.8–0.9 Å or better) and sufficient observations to refine all the atom positions, or all but H, without constraints and with anisotropic displacement parameters. Interatomic distances and other aspects of molecular geometry are therefore known to a substantially greater accuracy than is usually possible with proteins. Associated with the CSD are excellent systems for search, retrieval and analysis of geometry, *QUEST* and *VISTA* (Allen & Kennard, 1993*a*,*b*). Thus, for example, the extraction of all compounds with a particular chemical connectivity, *e.g.* Zn bound to imidazole N, is straightforward and can be followed by analysis to give means and ranges for all geometrical parameters of interest.

The objective of the present study is to make conveniently available to protein crystallographers accurate information on preferred geometry for a selection of metals and ligands which occur commonly in proteins. This could be useful in the interpretation and fitting of models to electron-density maps in which individual atoms are not resolved; for example, maps at 2–3 Å resolution. The details of distances and angles could also be useful when target distances are refined in restrained refinements using medium-resolution diffraction data. Engh & Huber (1991) provided a dictionary of target distances and their standard deviations for C—C, C—O and C—N distances and angles, but the inclusion of metal–ligand geometrical information at that stage was not practicable. A third area of possible usefulness is in the validation of protein structures, a project which has been undertaken by the EU Validation Network (EU Validation Network, 1998; Dodson *et al.*, 1998). Validation of any particular structure determination should include checking of the reported geometry against accepted means and ranges for each type of bond, angle *etc*. Finally, it can be hoped that the present analysis of the preferred geometry and the range of allowed geometry will contribute to the basic understanding of the function of different metals in metalloproteins.

Orpen *et al.* (1989) provided an excellent tabulation of bond lengths in coordination complexes of transition metals, based on the CSD, and including means and standard deviations for many of the types of bonds of interest here. At the time of compilation, the CSD contained results of ∼50000 single-crystal diffraction studies, so with ∼190000 structures now available it is worthwhile to repeat relevant parts of that analysis. The present analysis also includes Ca and Mg, which are not in the tables of Orpen *et al.* (1989), and some information on angles between bonds.

Einspahr & Bugg (1981) studied carboxylate interactions with Ca ions in all relevant crystal-structure determinations published at that time. Carugo *et al.* (1993) made a detailed comparison of coordination in structures of Ca and Mg complexes with water, alcohols, ethers, carboxylates and other ligands. The study of carboxylates was extended by Carrell *et al.* (1988) to a wide variety of biologically significant metal ions. Glusker and colleagues have also examined in more detail the coordination to several of these metal ions by ligands including water and carboxylate ions; they discuss coordination numbers and details of geometry and compare the findings of crystallographic database searches with the results of *ab initio* molecular-orbital calculations (Bock *et al.*, 1994, 1995; Katz *et al.*, 1996). These analyses showed that the preferred coordination number of magnesium is six, of zinc can be four, five or six and of calcium can be six, seven or eight. There is a strong preference for magnesium and calcium to bind to O atoms as the donors, while zinc will also readily bind to nitrogen or sulfur. Chakrabarti (1990*a*,*b*) examined the interactions of metal ions with carboxylate and histidine ligands in protein structures whose coordinates were available in the Protein Data Bank (PDB; Bernstein *et al.*, 1977) and displayed distances and orientations in a manner similar to that used here (Fig. 6) and by Carrell *et al.* (1988). Over 80% of the carboxylate interactions were with Ca; the histidine interactions were more evenly spread between Fe, Zn and Ca, with a very small number involving Mg or Mn. Carrell *et al.* (1993) have also explored the directionality of imidazole coordination to metals by extracting geometrical data from the CSD and by energy calculations.

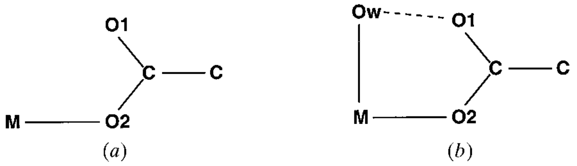
Here, the metals Ca, Mg, Mn, Fe, Cu and Zn have been included, since they are probably the metals most commonly found in metalloproteins. The ligands include imidazole and substituted imidazoles (as representatives of histidine side chains in proteins), carboxylate groups (representing aspartate and glutamate side chains), alcohols (as analogues of serine and threonine), phenolates (as analogues of tyrosine), thiolates (as analogues of cysteine) and water molecules. (Obviously it could be useful to add analogues of methionine, asparagine, glutamine and main-chain carbonyl groups in the future, as well as other metals.) Geometry within porphyrins or within Fe_4S_4-type clusters is not considered here, although the CSD could provide useful information on these. For each parameter of interest, a mean distance or angle is listed, together with its standard deviation and the range of values found in the sample. This *sample* standard deviation is a good indicator of the spread of values found. In each search of the CSD, the chemical connectivity of a 'query structure' is matched against that of a structure within the database. Obviously, it is desirable that the 'small-molecule' structures selected should be good analogues for the metal–ligand interactions in proteins, but too restrictive a query will not yield sufficient structures for a fruitful analysis.

Properties which may affect distances and angles include oxidation states and coordination numbers of metals (and, in a few cases, spin states), as well as the nature of the other ligands. Metal–ligand distances in these complexes would be expected to depend on oxidation state and spin state, if these are variable for the metal concerned; Mn and Fe in oxidation state 3 have been considered, together with oxidation state 2 for all six metals; oxidation state 1 for Cu is generally omitted, since the number of relevant complexes identifiable as Cu^I in the CSD is small. Metal–ligand distances may also vary slightly with coordination number (CN) and will certainly be affected by experimental errors. Estimating and disentangling all these is difficult, but it is important to try! Martin & Orpen (1996) have provided estimates of the possible effects of crystallographic errors and 'packing forces' on selected transition-metal complexes by examining duplicate determinations and determinations where identical chemical units occur more than once in the crystallographic asymmetric unit. Metal–ligand bond lengths of particular types generally have a wider range of variation than bond lengths within organic molecules or ligands. (Angles between bonds can, of course, be much more readily distorted from their ideal values than can bond lengths, and the magnitude of the distortions is often much larger than the experimental error in determining the angles. R.m.s. deviations of angles from ideal, such as tetrahedral, can be 3–5° and maximum deviations can be 20 or 25°.)

2. Methods and procedures

The April or October 1998 releases of the CSD have been used for all searches whose results are given here, and the programs *QUEST* and *VISTA* (Allen & Kennard, 1993*a*,*b*) were used for search and analysis. The search queries used for each ligand type are given in Figs. 1, 2, 3 and 4. All searches required $R \leq 0.065$; only non-polymeric structures with no disorder and no unresolved errors were accepted. Despite these restrictions, a few structures were found which had chemically very unreasonable bond lengths, *e.g.* a single bond C—O as short as 1.12 Å or as long as 1.77 Å; in such cases, the whole structure was excluded. The results given include the number of structures, N_E, in which the search fragment was found and the number of crystallographically independent observations of it, N_O, together with the mean, the sample standard deviation (which is a good guide to the spread of values) and the range of values found for each geometric quantity.

Searching for M—O or similar bonds in *QUEST* can be difficult for some metals, particularly Ca and Mg. An M—O bond is only found if it has been designated as such at the time that the database entry was made, normally if the $M\cdots$O distance is less than a specified threshold (radius sum plus tolerance), but sometimes it also depends on a chemical 'judgement' of the bonding situation. For Ca and Mg, the threshold distances used in compiling the CSD are rather small; consequently, the standard version of *QUEST* found rather few Ca—O or Mg—O bonds, although it could identify non-bonded Ca$\cdots$O or Mg$\cdots$O contacts up to a specified limiting distance. In order to perform other geometrical calculations and to treat all the 'bonds' and 'non-bonded contacts' together in the later part of this project, I was privileged to be allowed to use a pre-release copy of version 5.17 of the CSD program *QUEST*; this includes new facilities to recalculate the coordination environment of metal atoms at the time of search (Cambridge Crystallographic Data Centre, 1999, released April 1999). In this work, for Ca and Mg with each ligand type, a preliminary search was made for non-bonded contacts <2.8 Å (or appropriate distance); where the number found was substantial, the pre-release version (version 5.17) of *QUEST* was used and Ca$\cdots$O distances up to 2.80 Å and Mg$\cdots$O distances up to 2.45 Å were accepted as bonds; where the number of non-bonded contacts indicated was small or zero, the standard version of *QUEST* was used.



(*a*) (*b*)

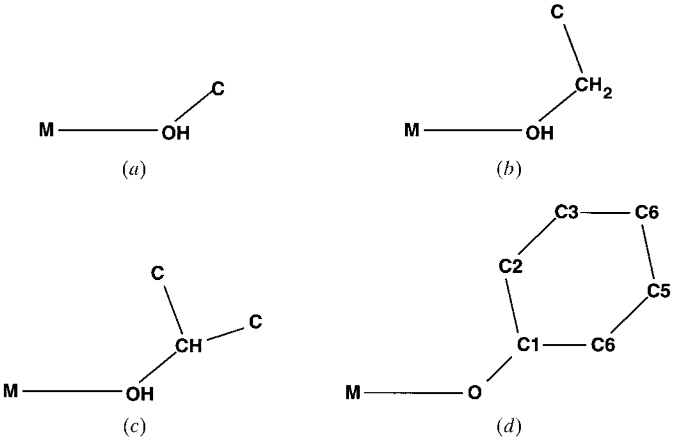
Figure 1
Search queries for (*a*) a carboxylate group coordinated to a metal and (*b*) for the 'hydrogen-bond motif' found in many carboxylates. In both, the bonds M—O2, C1—O1, C1—O2 could be of any type, C2 is tetrahedral (sp^3) carbon and no other atoms are bonded to C1, O1 or O2. $M\cdots$O1 may be a bond, but does not need to be (achieved with two separate searches). In (*b*), Ow is a water molecule and O1$\cdots$Ow must be <2.85 Å. Chelated complexes were generally identified as having O2 a 'cyclic' atom, but they were manually identified in the small number of symmetrically coordinate bidentate carboxylates.

Taking account of the oxidation states of the metal atoms in *QUEST* is not entirely straightforward. Oxidation states, where designated, are accessible from the chemical name of the compound. They are not given for all compounds and the oxidation-state number cannot be assigned or tested within the search query. Fe is particularly difficult in this respect: numerous compounds do not have oxidation states assigned and there are a number of mixed-valence compunds. For some metal–ligand combinations there are sufficient observations to treat different oxidation states and/or different metal coordination numbers as separate groups and wherever practicable this has been done; in many cases, it was not practicable. For each ligand type, the ligand geometry was examined for each metal separately. No significant differences were found and the ligand geometry is, therefore, averaged for all the metals which could be treated with *QUEST*.

3. Results and discussion

3.1. General

The numbers of examples of the different kinds of metal–ligand interaction found in the CSD are summarized in Table 1. There are dramatic variations from metal to metal. The numbers of examples in the different categories (and also in the sub-categories given later for oxidation states and coordination numbers) do give a crude indication of the relative importance or likelihood of the different interactions; for example, Ca interactions with oxygen donors are very common, but Ca complexes with nitrogen or sulfur donors are very rare. However, it should be remembered that these



(*a*) (*b*) (*c*) (*d*)

Figure 2
(*a*), (*b*) and (*c*), search queries for alcohol groups coordinated to metals, (*d*) search query for phenolate groups coordinated to metals. (*a*) and (*d*) were used for metal–oxygen distances, (*b*) and (*c*) for ligand geometry in serine and threonine-like complexes. No other atoms may be bonded to O or, in (*b*), (*c*) and (*d*), to the C atom attached to it. The M—O bond may be of any type; chelating and non-chelating ligands are distinguished by making the O 'cyclic' or 'acyclic'. In (*a*) to (*c*), all C atoms are sp^3. In (*d*), all C atoms are sp^2, all C—C bonds are of aromatic type and substitution is allowed at any or all of C2 to C6.

Table 1
The number of entries (N_E) and number of observations (N_O) found in the CSD which satisfied the search schemes in the figures indicated for different ligand types in combination with different metals (given as N_E/N_O).

Metal	Ca	Mg	Mn	Fe	Cu	Zn
$M-OH_2$	119/328	69/199	134/232	77/119	543/759†	131/246
$M-O_{carboxylate}$ (Fig. 1)‡	25/37	11/16	32/67	56/158	337/498†	75/121
$M-O_{alcohol}$ (Fig. 2)	53/145	12/25	32/45	21/30	187/293	29/40
$M-O_{phenolate}$ (Fig. 2)	4/7	0	84/168	96/226	192/290	28/43
$M-N_{imidazole}$ (Fig. 3)	1/1	0	10/16	13/24	88/139	10/20
$M-S_{thiolate}$ (Fig. 4)	0	0	8/30	28/75	4/5	14/25

† Mono and bidentate but excluding bridging carboxylates – see Table 3(*a*) for these. ‡ Cu^{II} only.

Table 2
Metal complexes with one or more coordinated water molecules.

Metal	Coordination number	Oxidation state	Number of observations	$d(M-O)$ (Å)		Outliers excluded [number, maximum d (Å)]
				Mean	Range	
Ca†	6	2	44	2.38 (5)	2.30–2.50	
	7	2	160	2.39 (4)	2.31–2.49	2, 2.72
	8	2	96	2.43 (5)	2.33–2.60	1, 2.70
Mg†	6	2	183	2.07 (3)	2.00–2.15‡	
Mn	6	2	121	2.19 (4)	2.10–2.31	
	6	3	34	2.26 (5)	2.14–2.33	1, 2.42
Fe	6	2	45	2.12 (4)	2.06–2.22	1, 2.28
	6	3	35	2.06 (6)	1.92–2.21	
Cu	4	2	20	1.97 (3)	1.91–2.04	
	5	2	326§ 92	1.97 (4)	1.90–2.10	
			226	2.30 (9)	2.14–2.52	8, 2.52
	6	2	345§ 135	1.97 (3)	1.86–2.06	Most obs are 1.93–2.02 Å
			210	2.42 (16)	2.11–2.87	Most obs are <2.70 Å
Zn	4	2	24	2.01 (3)	1.96–2.05	
	5	2	31	2.06 (9)	1.97–2.28	
	6	2	163	2.11 (6)	1.97–2.25	1, 2.35
Mg† in $Mg(OH_2)_6^{2+}$			100	2.06 (2)	2.01–2.13	
Ca† in $Ca(OH_2)_7^{2+}$			29	2.39 (4)	2.33–2.47	

† For Ca and Mg searches, the pre-release version 5.17 of *QUEST* was essential. ‡ For Mg, the CN = 6 distribution actually appeared as bimodal, 147 observations with mean $d(M-O)$ 2.06 (2) Å and 30 with 2.11 (1) Å, but no chemical explanation was apparent. § This search was performed with the pre-release version 5.17 of *QUEST*, allowing Cu—O up to 2.90 Å. The distributions for CN 5 and 6 are so obviously bimodal (*e.g.* see Fig. 5*d*) as a result of the Jahn–Teller effect that means and ranges are given separately for the shorter and the longer bonds. Also note that 24 observations in which the $Cu-OH_2$ bond is *trans* to a Cu···Cu bond have mean $d(Cu-O)$ 2.20 (10) Å, range 2.10–2.44 Å, and form a distinct 'bump' in the distribution.

numbers also reflect the interest of preparative and structural chemists in the different types of compound.

3.1.1. Error estimates. In all the crystal-structure determinations of these compounds, e.s.d.s of atom positions and then of bond lengths have been properly estimated, but e.s.d.s of atom positions are not stored in the CSD; one parameter, sigf, is stored with each structure; it is an indicator of the average e.s.d. of a C—C bond in that structure. From these sigf values it can be shown that for C—C distances in the structures used here, the e.s.d. is in the range 0.005–0.03 Å, and so it can be estimated that for $M-O$ and $M-N$ distances {assuming $\sigma[d(M-O)] \leq \sigma[d(C-C)]/(2^{1/2})$} the e.s.d. is in the range 0.004–0.02 Å. These estimates of errors should be borne in mind when examining the distributions of values found for the various distances given. When the sample standard deviation is comparable with the above e.s.d.s, it suggests that the sample is homogeneous and the bonds concerned are very similar electronically from one structure to another. Variations found in ligand geometry are mostly fairly consistent with this estimate of experimental errors, as are the variations in the Mg—O distance in the 100 independent observations in $Mg(OH_2)_6^{2+}$ ions, with mean 2.06 (2) Å. Within this group there should be no variations in electronic structure of the cation, only experimental errors and variations in the hydrogen-bonding environment of the cation. On the other hand, when a wide distribution with a large sample standard deviation is found, for example for $M-OH_2$ distances in $MX_n(OH_2)$, where X are any donors, it suggests that there are real though small variations in the $M-O$ distances, not just experimental errors in their measurement. Such differences must be a consequence of variation in the electronic structure of the other ligands around the metal atom.

A pilot study has recently been made (A. Gruneau & M. M. Harding, unpublished work) of the geometry of zinc–ligand interactions in 50 zinc proteins selected from the PDB. Proteins representative of different types were selected whose structures had been determined by X-ray or neutron diffraction to resolutions better than 2 Å; the PDB entries were converted to CSD format using the program *PROQUEST* under development at the CSD and in collaboration with Dr Marcel Verdonk; the metal–ligand geometry was then examined in the same way as with the small-molecule structures. 149 observations of histidine coordinated to zinc in proteins have a mean $d(Zn-N)$ of 2.07 (11) Å, compared with the small-molecule result (Table 6*a*) of 2.00 (2) Å. The larger e.s.d. from the protein set clearly reflects the larger errors in determining atom positions in proteins. The difference in the means, although not statistically significant, could point to some systematic error or artefact in the process of protein structure refinement.

3.1.2. Oxidation state. For Fe, the change from oxidation state 2 to 3 gives a decrease in $d(M-O)$ or $d(M-N)$ of between 0.03 and 0.07 Å for various ligand types, which is in the expected direction and not unreasonable in magnitude. For Mn, the $d(M-O)$ distances for carboxylates and phenolates show decreases of 0.18 and 0.14 Å, which are again reasonable, but with water and alcohols as ligands there are

Table 3
Metal–carboxylate complexes.

(*a*) Distances $d(M-\mathrm{O2})$, $d(M-\mathrm{O1})$ and $d(\mathrm{Mpl})$, the displacement of the metal atom from the OCO plane, using the search query in Fig. 1(*a*). For Mn to Zn, the data are for non-chelated complexes and mono- and bidentate complexes have been merged (because, especially for Cu and Zn, the distinction is not clear). For Ca, the data include chelated complexes (because it was not practicable to distinguish the non-chelated ones in the pre-release version 5.17 of *QUEST*) and mono- and bidentate complexes of Ca have been listed separately because the distinction is easy and clear. Also given are numbers of complexes found which have carboxylate bridging between two metal centres and the maximum $M\cdots M$ distance in these; they include chelating and non-chelating ligands for all metals.

Metal	Number of observations	$d(M-\mathrm{O2})$ (Å)		$d(M-\mathrm{O1})$ (Å)		$d(\mathrm{Mpl})$ (Å)		Number of bridging $\mathrm{CO^{2-}}$	Max. $M\cdots M$ (Å)
		Mean	Range	Mean	Range	Mean	Max.		
Ca† monodentate	32	2.37 (7)	2.27–2.49	4.4 (1)	3.0–4.6	0.6 (5)	1.8	15/23	6.7
bidentate	6	2.55 (8)	2.39–2.73			0.4 (2)	0.7	11/12	6.9
Mg	15	2.11 (9)	2.03–2.34	4.1 (2)	3.3–4.4	0.3	0.7	6/10	6.0
Mn	16	2.15 (7)	2.02–2.30	3.2 (7)	2.3–4.3	0.4 (4)	1.5	68/142	5.6
Fe	15	2.01 (5)	1.90–2.09	3.0 (6)	2.2–4.1	0.2 (2)	0.5	35/101	4.4
Cu	120	1.99 (10)	1.91–2.52	2.9 (4)	2.0–4.2	0.2 (2)	1.2	122/257	4.9
Cu short‡ only	109	1.96 (3)	1.91–2.03	2.9 (3)	2.0–3.4	0.2 (2)	0.6		
Zn	73	2.00 (7)	1.86–2.18	2.9 (3)	2.2–3.8	0.2 (3)	0.7	12/22	4.6

(*b*) Observed coordination numbers in complexes above. The coordination numbers 7 and 8 in Fe and Cu complexes are for clusters which include $M-M$ bonds or bonds to cyclopentadiene-type ligands.

CN	Ca	Mg	Mn	Fe	Cu	Zn
4					59	49
5			5	4	68	10
6	5	8	37	29	36	14
7	11	7	2	(4)	(1)	
8	11			(1)	(1)	
9	2					

(*c*) Carboxylate-ligand geometry in non-chelating complexes (but including bidentate carboxylate) for all metals Mn to Zn. The search query in Fig. 1(*a*) was used; the carboxylate was not allowed to be part of a larger chelating ligand. 354 observations from 218 structures were used.

	Mean	Range
$d(\mathrm{C-O1})$ (Å)	1.236 (21)	1.15–1.33
$d(\mathrm{C-O2})$ (Å)	1.263 (20)	1.18–1.33
$d(\mathrm{C-C})$ (Å)	1.513 (22)	1.45–1.65
$\angle(\mathrm{C-C-O1})$ (°)	119 (2)	111–125
$\angle(\mathrm{C-C-O2})$ (°)	116 (2)	107–123
$\angle(\mathrm{O1-C-O2})$ (°)	125 (3)	116–136
$\angle(M-\mathrm{O2-C})$ (°)	118 (14)	89–159
Selecting only the 23 carboxylate ligands with near symmetrical coordination $d(M-\mathrm{O1}) < 2.35$ Å		
$d(\mathrm{C-O1})$ (Å)	1.257 (11)	1.24–1.29
$d(\mathrm{C-O2})$ (Å)	1.252 (17)	1.19–1.27

(*d*) The hydrogen-bonded motif (search query Fig. 1*b*): dimensions and number of observations with $\mathrm{Ow}\cdots\mathrm{O1}$ closer than 2.85 Å. Coordination number is 6 in most, 5 in 15 of the Cu examples and 4 in one Cu complex.

	Mean	Range
$d(\mathrm{O-Ow})$ (Å)	2.64 (8)	2.50–2.81
$\angle(\mathrm{O2}-M-\mathrm{Ow})$ (°)	91 (4)	85–99

	Number of observations
Ca	1
Mg	1
Mn	16
Fe	6
Cu	28
Zn	12§

† For Ca searches, the pre-release version 5.17 of *QUEST* was required. ‡ 'Short' bonds to Cu, with $d(M-\mathrm{O2}) < 2.10$ Å. § Two additional Zn compounds are tetrahedral, with longer $\mathrm{O}\cdots\mathrm{Ow}$ distances of 2.96 and 2.99 Å.

increases, which are not expected. For both metals, differences in spin state (not identifiable in the CSD) or mixed-valence compounds (not usually excluded here) may be confusing the results. Ca, Mg and Zn are normally present only in oxidation state 2 so they do not present problems. $\mathrm{Cu^{I}}$ compounds have been excluded, leaving only $\mathrm{Cu^{II}}$.

3.1.3. Coordination numbers. For Zn coordinated to water molecules, $d(M-\mathrm{O})$ increases by ~0.05 Å with the increase of coordination number (CN) from 4 to 5 or from 5 to 6; for Ca the increase is ~0.04 Å from CN 6 to 7 or from 7 to 8. These changes are roughly in line with those expected for ionic compounds (see, for example, Brown, 1992). For other metals, similar changes would be expected, but there are only sufficient observations to show this clearly for Cu.

3.1.4. Jahn–Teller effect. For $\mathrm{Cu^{II}}$ with electron configuration d^9, the Jahn–Teller effect dominates the geometry of five- and six-coordinate complexes; usually, the four equatorial bonds are short and the axial bond(s) are longer. Inspection of the data for $\mathrm{Cu^{II}}$ coordinated by water (Table 2) shows that for CN 4 and for the short bonds in CN 5 and 6 (the 'short' bonds are selected as those <2.10 Å), the mean distances are very nearly the same; the change is only ~0.01 Å from CN 4 to 5 or from CN 5 to 6. The 'long' bonds in CN 5 complexes are ~0.3 Å longer than the short ones, and the two long bonds in CN 6 complexes are 0.4–0.5 Å longer than the short ones; while the short bonds of $\mathrm{Cu^{II}}$ complexes show great consistency, the long bonds are very variable (see Fig. 5*d*). $\mathrm{Mn^{III}}$, with electron configuration d^4, would be expected to show a similar though weaker effect; this is clearly seen in the 'outliers' noted for $\mathrm{Mn^{III}}$ complexes in Tables 2(*a*), 4(*a*) and 5(*a*).

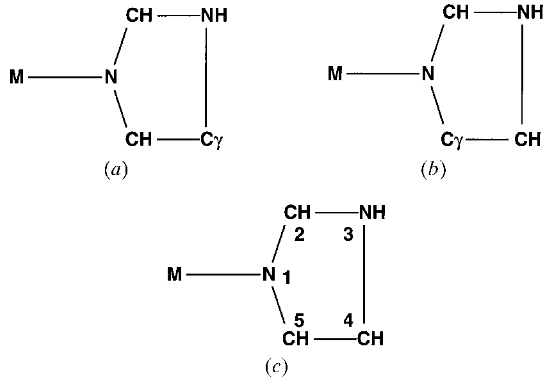
3.2. Water molecules coordinated to metal atoms

The number of observations in Table 2 is large enough for reasonable statistics and to break down most groups according to oxidation state and coordination number. In each category, there is a spread of values of $d(M-\mathrm{O})$ characterized by the sample standard deviation. According to the estimates of experimental error (above) and the observations in

$Mg(OH)_6^{2+}$, only a part of this spread should, in most cases, be attributed to experimental error in the determination of $d(M-O)$. There must also be real variations in the bond length as a result of variations in the donor groups around M and their electronic structures. In Figs. 5(*a*)–5(*c*), distributions for Ca, Mg and Zn may be compared. The spread for Zn is widest and it is much wider than for Mg or Ca; this may well be associated with its flexibility or adaptability in different chemical situations. The distribution for Cu (coordination number 6, oxidation state 2, Fig. 6*d*) shows the results of the Jahn–Teller effect: 'short' bonds with a narrow spread of values and 'long' bonds with a very wide spread.

3.3. Metal carboxylates

Carboxylate complexes are analogues for coordination of metals in proteins by the side chains of aspartate and glutamate. Carboxylate complexes in the CSD show a great deal of flexibility and variability, and carboxylate complexes with Ca show several types of geometrical behaviour not found with the other metals considered here. Carboxylates have often been categorized (*e.g.* Carrell *et al.*, 1988) as *syn*, *anti* or symmetrical bidentate [(I), (II) and (III)] and the geometry of the first two is attributed to the preference for the $O-M$ bond to lie near the direction of one of the lone pairs of the carboxylate O atom. One of the objectives of the present study was to see how precisely the structures found in the CSD conform to this geometry. Further, the carboxylate group may readily act as a bridge between two metal centres (IV), again with variations in geometry (*syn* or *anti*). A bidentate Ca carboxylate group may also be linked to one or two other Ca ions [(V) and (VI)].



(*a*) (*b*)

(*c*)

Figure 3
(*a*) and (*b*): search queries used for imidazole groups coordinated to metals appropriate for N^{ε} and N^{δ} coordination by histidine. All bonds could be of any type. Substitution was not allowed at any atom of the five-membered ring except at C^{γ}; C^{γ} was required to be an sp^2 carbon. (*c*) Search query for unsubstituted imidazole group coordinated to a metal, showing the labelling used for C and N atoms. All bonds could be of any type; no substitution was allowed. Complexes are designated as chelated if the metal atom is 'cyclic', *i.e.* if *any* of its ligands is chelating.

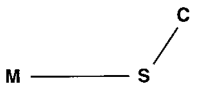
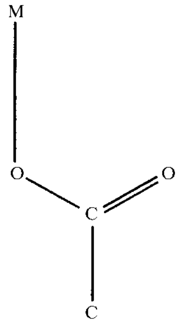
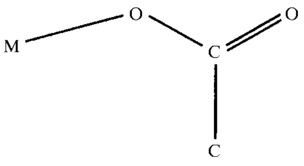


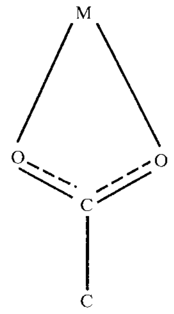
Figure 4
Search query for S of thiolate group coordinated to metal. The $M-S$ bond may be of any type, no substitution is allowed at S, carbon is sp^3. The complex is designated as chelated if S is a 'cyclic' atom.



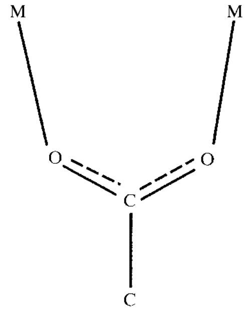
(I) *syn*-monodentate



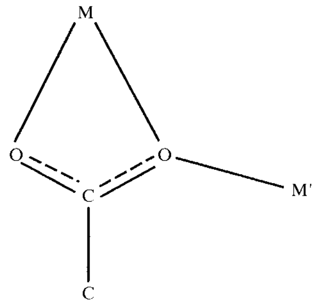
(II) *anti*-monodentate



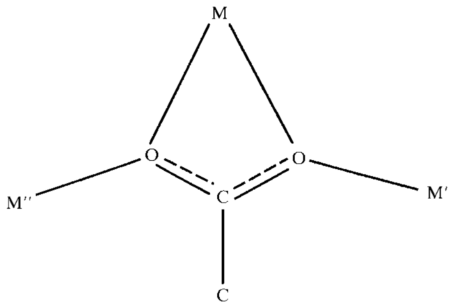
(III) symmetrical bidentate



(IV) bridging (different geometries possible)



(V) bridging



(VI) bridging (3 metal centres)

3.3.1. Carboxylate bonded to one metal centre. Fig. 6 shows the distribution of observed geometry. Similar figures were presented by Glusker and colleagues (Carrell *et al.*, 1988); the present figures include many more structures determined in the last ten years and also distinguish between

Table 4
Metal–alcohol complexes.

(a) Metal–oxygen distances and numbers of structures (N_E) and of observations (N_O), for predominant coordination numbers (CN) and oxidation states (OS). The search query in Fig. 2(a) was used. 15 structures with $d(\mathrm{O-C})$ outside the range 1.35–1.51 Å have been excluded.

	CN	OS	N_E/N_O	$d(M-\mathrm{O})$ (Å)		
				Mean	Range	
Ca†	6	2	4/10	2.35(3)	2.30–2.37	
	7	2	20/51	2.42 (4)	2.34–2.52	
	8	2	25/74	2.45 (5)	2.36–2.56	
Mg†	6	2	12/25	2.10 (4)	2.01–2.22	
Mn	6	2	12/19	2.22 (6)	2.15–2.40	
	6	3	16/22	2.27 (5)	2.18–2.34	2 outliers, max. 2.41 Å
Fe	6	2	10/15	2.14 (5)	2.07–2.21	
	6	3	11/15	2.11 (6)	2.00–2.25	
Cu	4	2	13/14	1.99 (4)	1.93–2.04	
	5	2	58/88‡	2.00 (3)	1.93–2.08	48 obs
				2.32 (10)	2.10–2.50	39 obs, 1 outlier, 2.65 Å
	6	2	53/85‡	2.00 (3)	1.96–2.08	26 obs
				2.38 (15)	2.16–2.53	53 obs, 6 outliers, 2.77 Å
Zn	5	2	14/15	2.12 (6)	2.01–2.27	
	6	2	15/24	2.16 (10)	2.08–2.46	

(b) Geometry of alcohol ligands: $d(M-\mathrm{O})$, metal–oxygen distance; $\angle(M\mathrm{OC})$, angle at oxygen.

	Non-chelating ligands	Chelating ligands
Serine-like (search query Fig. 2b)		
N_E/N_O	64/116	24/34
$d(\mathrm{O-C})$ (mean, range) (Å)	1.43 (2), 1.36–1.49	1.42 (3), 1.36–1.46
$\angle(M\mathrm{OC})$ (mean, range) (°)	112 (7), 96–136	129 (6), 115–142
Threonine-like (search query Fig. 2c)		
N_E/N_O	26/50	2/2
$d(\mathrm{O-C})$ (mean, range) (Å)	1.45 (1), 1.41–1.48	Values were 1.44 and 1.45
$\angle(M\mathrm{OC})$ (mean, range) (°)	115 (8), 91–132	Values were 122 and 133

† For Ca and Mg, the pre-release version 5.17 of *QUEST* was required. ‡ Bimodal distribution, split into ranges as in Table 2.

carboxylate groups which are part of chelating ligands and those which are not. The *anti* conformation is normally the only one possible with chelating ligands such as α-hydroxy- or α-amino-carboxylates. The very few chelating ligands for which the conformation is nearer to *syn* have very large chelate rings, *e.g.* with a hydroxyl and carboxylate group separated by a long flexible carbon chain. The metal atom is normally found near the plane of the carboxyl group, but need not be exactly in it, as shown by Fig. 6, and by the magnitude of the deviations from this plane, d(Mpl), given in Table 3(a). Non-chelating ligands all give complexes with the *syn* or symmetrical bidentate conformation or something near it; the extent of divergence from exact *syn* or exact symmetrical bidentate is evident in Fig. 6 and is quite considerable. The non-chelated complexes are slightly better analogues for aspartates and glutamates than are the chelated ones; for this reason, geometrical data are given for non-chelated complexes in Table 3 (although there are no very marked differences when chelated complexes are also included). Ligand geometry is given in Table 3(c); there was no significant difference between all ligands and non-chelating ligands.

Inspection of Fig. 6 suggests that for Ca, Mg and Mn the carboxylate interactions are less strongly directed, which is entirely consistent with the expectation of greater ionic character for these than for Fe, Cu and Zn. Even for Fe, Cu and Zn, the carboxylate-to-metal bond is not as strongly directional as the imidazole-to-metal bond (compare with the non-chelating complexes in Fig. 8). It also shows that *syn* monodentate and symmetrical bidentate carboxylates are not totally distinct patterns, especially for Cu and Zn, but that a full range of intermediates between them is possible. The monodentate carboxylates should have a single bond $M-\mathrm{O2}$ and no bond between M and O1, and the bidentate carboxylates should have two bonds, $M-\mathrm{O1}$ and $M-\mathrm{O2}$, each of order one half. In Fig. 6, *syn* monodentate carboxylates will appear with M directly 'above' the left-hand O atom, whereas symmetrical bidentate carboxylates will have M directly above the mid-point between the two O atoms. [The atom numbering is as in Fig. 1a, with $d(M-\mathrm{O2}) \leq d(M-\mathrm{O1})$ whether they are bonded or not.] For Cu, this intermediate geometry seen is easily understood on the basis of the Jahn–Teller effect; the primary bond, $M-\mathrm{O2}$, is a short bond within the square-planar base, 1.91–2.03 Å, and the secondary bond, $M-\mathrm{O1}$, is to oxygen in a near axial position and is usually 2.2–2.6 Å. Even some of those which appear to be symmetrical bidentate may be unsymmetrical and disordered (Stebler & Burgi, 1987; Simmons *et al.*, 1987). Zinc is different. There can be no Jahn–Teller effect giving rise to differences in Zn—O distances, and its 4-coordinate complexes are tetrahedral, not square planar like those of Cu. Nevertheless the (non-chelated) zinc carboxylate complexes show an almost continuous range of $M-\mathrm{O1}$ distances, from $d(M-\mathrm{O1}) \simeq d(M-\mathrm{O2}) \simeq 2.2$ Å, with both O atoms equally bonded to M, to those with $d(M-\mathrm{O1}) > 3$ Å, $d(M-\mathrm{O2}) \simeq 2.0$ Å, in which $M-\mathrm{O2}$ is clearly a single bond and $M\cdots\mathrm{O1}$ not a bond at all. Complexes can be identified which have $d(\mathrm{Zn-O1})$ in the range 2.4–2.8 Å, which must be clear examples of Zn—O1 bonds of partial order; for example, the complex shown in Fig 7. Such intermediate states, described as aniso-bidentate, have been found in zinc complexes synthesized as models for the active site of carbonic anhydrase (Cronin, 1997; Cronin & Walton, 1999). The additional partial bond modifies the geometry around Zn from tetrahedral towards distorted trigonal bipyramidal or from trigonal bipyramidal towards octahedral. Additional partial bonds are not found when the coordination number of zinc is already 6. Zinc clearly offers a flexibility in this respect, which may well be significant in relation to its catalytic activity.

Ligand geometry is given in Table 3(*c*); there was no significant difference between all ligands and non-chelating ligands, and the C—O distances are close to the length of the two equal delocalized bonds in the carboxylate anion, 1.254 (10) Å, given by Allen *et al.* (1987). An inverse correlation between $d(\mathrm{C—O1})$ and $d(\mathrm{C—O2})$ would be expected; scattergrams show that it is present, but the error levels are such that it is far from clear. However, for the small number of carboxylates which are unambigously bidentate $d(\mathrm{C—O1})$ and $d(\mathrm{C—O2})$ do appear to be very nearly equal; see Table 3(*b*).

3.3.2. Hydrogen-bonded motif. The presence of the bonding pattern shown in Fig. 1(*b*) and its effect on the carboxylate-group orientation was noticed in a number of zinc complexes and a systematic search was subsequently made for it with all metals. Table 3(*d*) shows the results. The hydrogen-bonding pattern is found mainly in the six-coordinate complexes of Mn and Zn and in the five- and six-coordinate complexes of Cu. The carboxylate is always monodentate, *i.e.* when this motif is present it does not appear to be possible for O1 to be within the first coordination sphere around the metal atom, except in the one Ca structure (CSD code CANIAC) which is seven-coordinate.

3.3.3. Carboxylate bonded to two or more metal centres. Table 3(*a*) also shows that carboxylate bridges between two

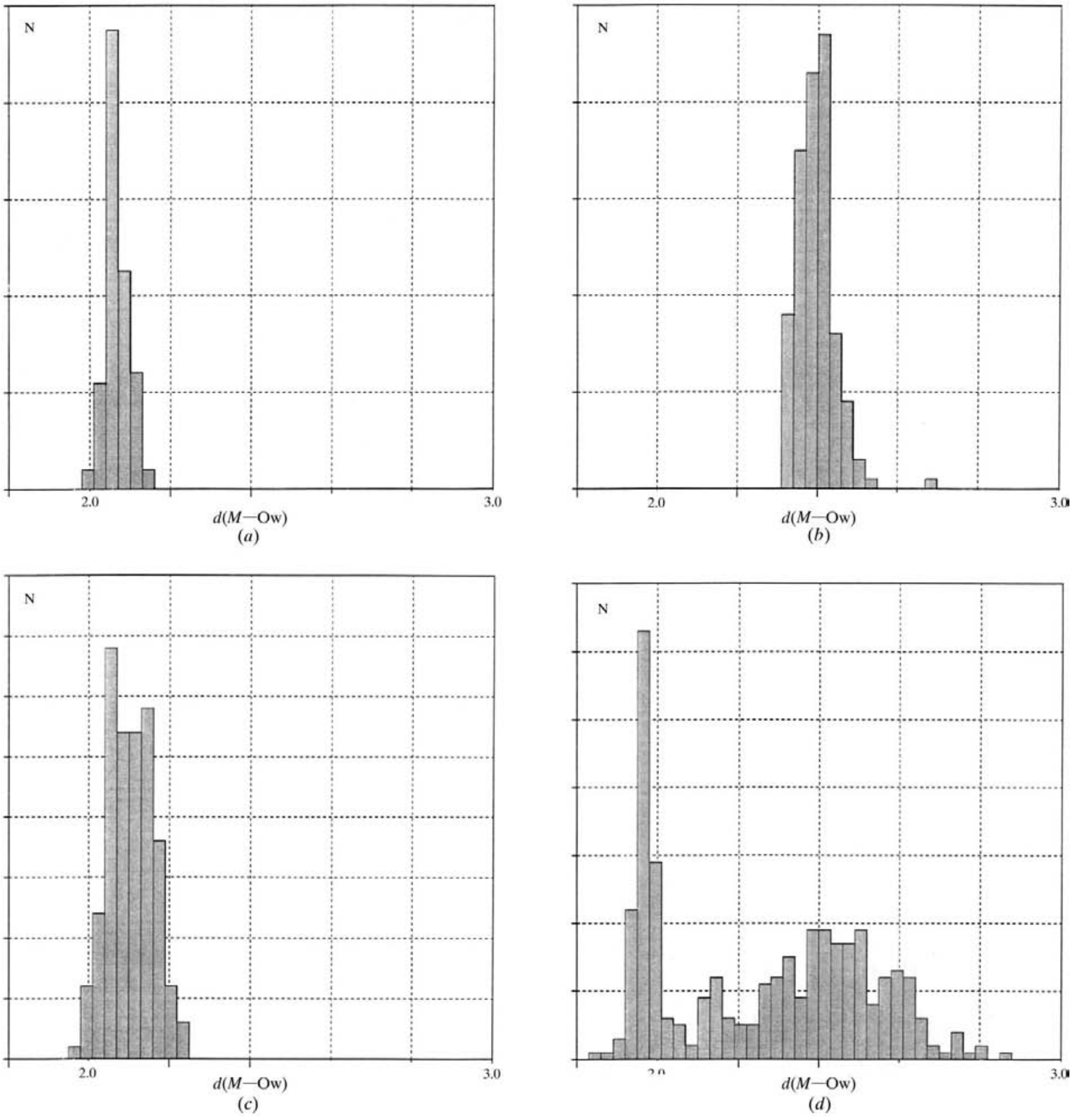


Figure 5
Histograms showing the distribution of $d(M—\mathrm{OH_2})$ distances (in Å) in complexes containing one or more water molecules in the coordination sphere of the metal. (*a*) for six-coordinate Mg complexes, (*b*) for six-coordinate Zn complexes, (*c*) for seven-coordinate Ca complexes and (*d*) for six-coordinate $\mathrm{Cu^{II}}$ complexes.

metal centres (of the same metal) [as in (IV)] are common in the CSD; the mean bonded M—O distances do not differ significantly from the mean $d(M$—O2) found in the simpler complexes (Table 3*a*). For Mn to Zn, these include many complexes in which the two metal centres are also bridged by a single atom or by one or more additional carboxylate groups, and these have short $M\cdots M$ distances, *e.g.* 2.8–3.7 Å (even shorter distances are found which can be designated as M—M bonds). For all the metals, there are examples of structures with longer $M\cdots M$ distances, and in these the only link

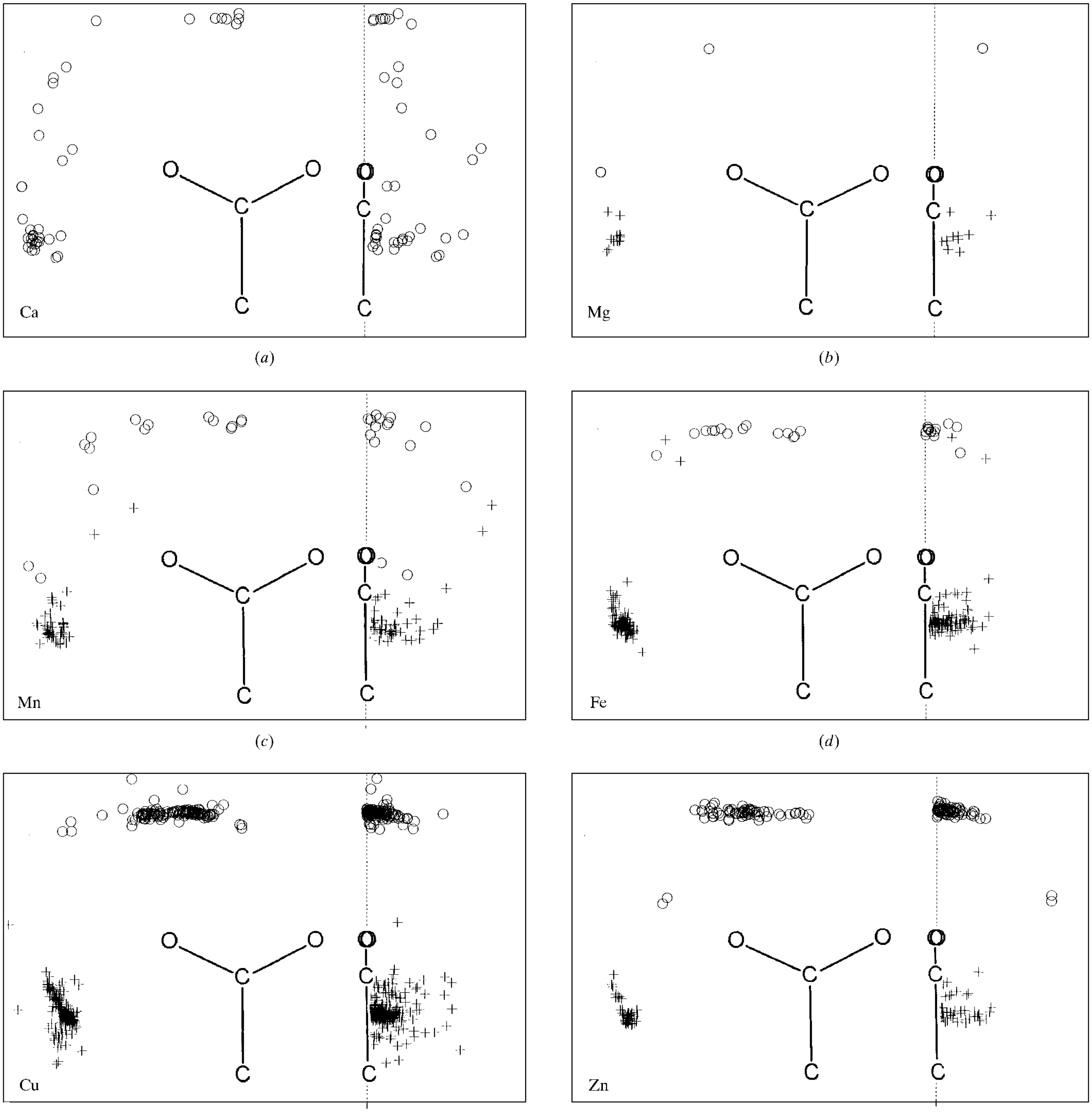


Figure 6
Scatter plots showing the distribution of metal-atom positions relative to the carboxylate group for (*a*) Ca, (*b*) Mg, (*c*) Mn, (*d*) Fe, (*e*) Cu and (*f*) Zn. For each metal, the left view is projected perpendicular to the plane of the carboxylate and the right view parallel to it; the right view thus shows displacements of the metal from the carboxylate plane. Crosses indicate chelated complexes (usually with α-hydroxy or α-amino ligands) and open circles non-chelated complexes (except for Ca, for which no distinction is made).

Table 5
Metal–phenolate complexes†.

(*a*) Metal–oxygen distances and numbers of structures (N_E) and observations (N_O) for principal coordination number (CN) and oxidation state (OS). The search query is in Fig. 2(*d*). As with the alcohol donors, the distribution of the $d(\mathrm{O-C})$ distances included a few very widely scattered values and it was judged wise to exclude these structures entirely (*e.g.* one at 1.06 Å).

	CN	OS	N_E/N_O	$d(M-\mathrm{O})$ Å		Comments
				Mean	Range	
Mn	5	3	18/47	1.88 (2)	1.85–1.91	For 39 obs plus 8 scattered in range 1.97–2.05 Å
	6	3	53/110	1.88 (2)	1.83–1.92	For 99 obs plus 10 scattered in range 1.95–2.12 Å
Fe	5	3	22/60	1.90 (2)	1.86–1.95	
	6	3	63/144	1.94 (5)	1.84–2.10	
Cu	4	2	122/192	1.89 (2)	1.86–1.93	8 outliers, 1.81–1.95 Å
	5	2	59/83	1.92 (3)	1.87–1.99	76 obs
				2.24 (7)	2.14–2.35	7 obs
	6	2	11/15	1.93 (2)	1.90–1.96	11 obs
				2.35, 2.38		
Zn	4	2	11/19	1.91 (2)	1.88–1.95	
	5	2	15/21	1.98 (4)	1.91–2.05	

(*b*) Geometry of phenolate ligands: $d(\mathrm{O-C})$, O—C distance; $\angle(M\mathrm{OC})$, angle at oxygen ; $d(\mathrm{Mpl})$, the distance of metal atom from the plane of O and the six C atoms of phenolate ion.

	Number	Mean	Range	Comments
$d(\mathrm{O-C})$ (Å)	480/900	1.32 (2)	1.26–1.39	18 outliers
$d(\mathrm{Mpl})$ (Å)	480/900	0.4 (3)	0–1.7	
$\angle(M\mathrm{OC})$ (°)				
Chelating ligands	453/843	126 (6)	105–143	Mainly 122–136 Å‡
Non-chelating ligands	27/42	134 (6)	117–150	Plus 5 outliers§

† Note: a small number of other complexes were found: Ca, four phenolate complexes found using the pre-release version 5.17 of *QUEST*, mean $d(M-\mathrm{O}) = 2.29$ (8) Å; Mn, 11 complexes, oxidation state 2, mean $d(M-\mathrm{O}) = 2.02$ (11) Å; Fe, 13 observations for oxidation state 2, mean $d(M-\mathrm{O}) = 1.97$ (6) Å and seven for oxidation state 3, CN 4, mean $d(M-\mathrm{O}) = 1.88$ (3) Å; Zn, two structures with CN 6. ‡ There are more than 600 six-membered chelate rings in which $\angle(M-\mathrm{O-C}) \simeq 129°$ and 100 five-membered chelate rings with $\angle(M-\mathrm{O-C}) \simeq 113°$. § In these five complexes, the phenol has two Bu^t *ortho* substituents; all these have $\angle(M\mathrm{OC})$ near 180°, which is probably a steric effect.

between the two metal centres is the one carboxylate group. The geometry of the carboxylate group is not significantly different from that in Table 3, except that the two C—O distances are more nearly identical [$d(\mathrm{C-O})$: mean, 1.253 (21); range, 1.18–1.30 Å] and the distribution of metal positions in relation to the carboxylate group is similar to the distributions in Fig. 6. There are a small number of structures in the CSD with carboxylate bridging two different metal centres, for example Zn and Fe. In some Ca salt structures one or both carboxylate O atoms may be bonded to more than one Ca ion [as in (V) or (VI)]; the Ca—O distances agree well with those in Table 3(*a*), with the characteristic mono- and bi-dentate distances being clearly distinguished.

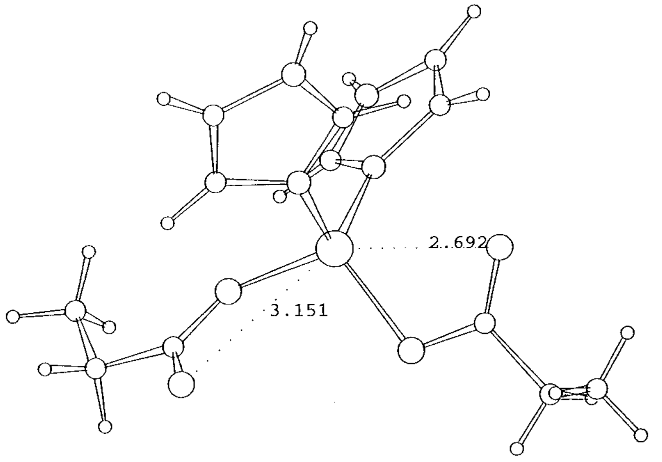


Figure 7
An example of a zinc complex in which one carboxylate group is clearly monodentate, but the other is intermediate between mono- and bidentate. The complex is bis(imidazole) bis(propionato)zinc (CSD refcode ECBIMB10). Distances shown are in Å, the largest atom is Zn, the next largest atoms are O, then N, then C and the smallest is H.

3.4. Metal complexes with other oxygen donors – alcohols, phenolates

Metal–alcohol complexes (search queries in Figs. 2*a*–2*c*) provide analogues for metal coordination to the side chains of serine and threonine. Table 4(*a*) gives the results for metal–oxygen distances and Table 4(*b*) for ligand geometry. There were not sufficient complexes found with the search queries of Figs. 2(*b*) and 2(*c*) to use these for $M-\mathrm{O}$ distances. The $d(M-\mathrm{O})$ distances follow those for $d(M-\mathrm{OH_2})$ in Table 2 very closely, with the distance larger by 0.03 (2) Å in the alcohol complexes. The $d(\mathrm{O-C})$ distance is not significantly changed from that in the free ligands (Allen *et al.*, 1987; Engh & Huber, 1991, for CH1E-OH; Engh and Huber give names to 31 different atom types in proteins, such as CH1E for tetrahedral carbon with one H, and then use the CSD to list average bond lengths and angles with e.s.d.s for various combinations).

Metal phenolates (search query in Fig. 2*d*) provide analogues for the side chain of tyrosine coordinated to metals and the results are given in Table 5. Note that in contrast to the searches for alcohol complexes where the ligand was neutral $R-\mathrm{O-H}$, here the searches were for the phenolate anion, $\mathrm{Ar-O^-}$, coordinated to a metal. The presence of a group of slightly longer Mn—O distances in the Mn(III) phenolates is attributable to the Jahn–Teller effect. The distances are all very significantly shorter than for the alcohol complexes, presumably as a result of the greater ionic character of the interaction. Also, the $\mathrm{O-C_{aryl}}$ distance, 1.32 (2) Å, is considerably shorter than in the free ligand [Allen *et al.* (1987) give 1.362 (15) Å, Engh & Huber (1991) give 1.376 (21) Å for CY2-OH]. *Ortho*-, *meta*- and *para*-substituted phenolates were included in the search query, but only ten structures were found which did not have substituents *ortho* to the phenolate O atom.

Table 6
Metal–imidazole complexes.

(a) Metal–nitrogen distances $d(M{-}\mathrm{N})$, displacement of metal atom from imidazole plane $d(\mathrm{Mpl})$ and numbers of observations (N_O) for principal coordination numbers (CN). The results of search queries in Figs. 3(a), 3(b) and 3(c) have been combined.

	N_O	$d(M{-}\mathrm{N})$ (Å)		$d(\mathrm{Mpl})$ (Å)		Number of observations		
		Mean	Range	Mean	Max.	CN = 4	CN = 5	CN = 6
Ca	1	2.44		0.17				1
Mg	0							
Mn	16	2.19 (8)	2.03–2.29	0.19 (16)	0.50		2	14
Fe	24	2.08 (9)	1.95–2.22†	0.18 (17)	0.55		4	20
$\mathrm{Cu^{II}}$‡	135	1.99 (5)	1.90–2.28	0.14 (11)	0.69	37	61	37
Zn	20	2.00 (2)	1.97–2.04	0.20 (11)	0.54	12	8	

(b) Ligand geometry: bond distances within the imidazole ring for search queries of Figs. 3(a), 3(b) and 3(c) combined; the differences are not significant. Angles $\angle(M{-}\mathrm{N}{-}\mathrm{C})$ are given for the three queries separately. $\angle(M{-}\mathrm{N}{-}\mathrm{C})$ is the angle $M{-}\mathrm{N}{-}\mathrm{C2}$ in Fig. 3(c) and the bonds are also defined there.

Bond	Mean (Å)	Range (Å)
N1–C2	1.319 (17)	1.26–1.38
C2–N3	1.337 (15)	1.30–1.41
N3–C4	1.362 (21)	1.28–1.45
C4–C5	1.352 (18)	1.26–1.46
C5–N1	1.378 (16)	1.30–1.42

Angle	Number of observations	Mean (Å)	Range (Å)
$\angle(M{-}\mathrm{N}{-}\mathrm{C})$; Fig. 3($a$)	18	126 (2)	122–130
$\angle(M{-}\mathrm{N}{-}\mathrm{C})$; Fig. 3($b$)	77	131 (6)	121–149, many are chelated
$\angle(M{-}\mathrm{N}{-}\mathrm{C})$; Fig. 3($c$)	106	126 (3)	113–132
$\angle(M{-}\mathrm{N}{-}\mathrm{C})$; all	201	128 (5)	113–149

† The distribution of $d(M{-}\mathrm{N})$ for Fe is obviously bimodal, 10 observations with $d(M{-}\mathrm{N})$ 1.95–1.99 Å and 14 observations with $d(M{-}\mathrm{N})$ 2.08–2.22 Å; both groups have oxidation state III and probably correspond to low-spin and high-spin complexes. ‡ Plus two outlying $\mathrm{Cu^{II}}$ complexes, with $d(M{-}\mathrm{N})_{\max}$ 2.59 Å and $d(\mathrm{Mpl})_{\max}$ 1.46 Å. There is also one $\mathrm{Cu^{I}}$ complex with two observations of $d(M{-}\mathrm{N})$, 1.90 and 1.91 Å, $d(\mathrm{Mpl})$ 0.09 and 0.16 Å, CN 3.

3.5. Metal–imidazole complexes

The search fragments, Figs. 3(a) and 3(b), are analogues for metal coordination at N^{ε} and N^{δ} of a histidine side chain, respectively. The metal–nitrogen distances given in Table 6(a) correspond to the query in Fig. 3(c); the separate queries in Fig. 3(a) and Fig. 3(b) did not appear to give any significantly different results and the number of observations of complexes in which a substituent atom is present at C_{γ} is small except for Cu. The scatter plots in Fig. 8 show that the metal atom lies close to the bisector of the C—N—C angle (C2—N1—C5 in Fig. 3c), and is near to, but not necessarily in, the imidazole plane; these deviations are quantified by $\angle(M{-}\mathrm{N}{-}\mathrm{C})$ in Table 6(b) and $d(\mathrm{Mpl})$ in Table 6(a). The ideal coordination direction is expected to be along this bisector and in the plane. Carrell *et al.* (1993) showed, from energy calculations and from observations in well determined structures, that 'out-of-plane' distortion is easier than 'in-plane' distortion. Here, the mean out-of-plane distortion is ~5° and the maximum is 20° (apart from the two outlying Cu complexes at 40°). For imidazoles other than those with substituents at C5, the r.m.s. in-plane distortion [taken as $\angle(M{-}\mathrm{N}{-}\mathrm{C}) - 126°$] is 2–3°, with a maximum of 13°; for those with substituents at C5, many of which are chelating, the angle $\angle(M{-}\mathrm{N}{-}\mathrm{C})$ is larger by 5°, with a maximum enlargment of 23°. This gives another indication of the magnitude of deviations from ideality which may be caused by the chemical-bond strains of chelation.

Bond lengths in the imidazole ring are tabulated in Table 6(b). Small differences were found between the values found with the queries of Figs. 3(a), 3(b) and 3(c), but none is greater than one standard deviation. The observed distances are well reproduced by Engh and Huber's tabulated values (Engh & Huber, 1991) for *neutral* histidine with the assignments in Engh and Huber terms, N1 as NR, C2 as CRH, N3 as NH1, C4 as CR1E and C5 as C5; the mean difference is $\sim\sigma/2$, the maximum difference $\sim\sigma$. (Unfortunately, no one choice of Engh and Huber names will allow all the bond lengths for the structure in Fig. 3a, including $\mathrm{C}_{\beta}{-}\mathrm{C}_{\gamma}$, to be looked up in Engh and Huber's table.)

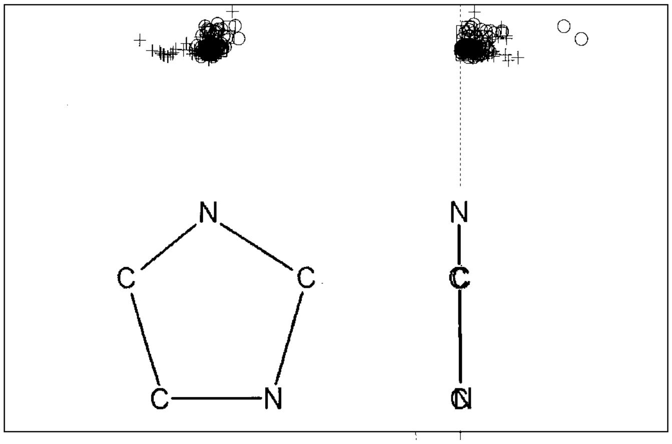


Figure 8
Scatter plots showing the distribution of metal-atom positions for Mn to Zn relative to the imidazole group, viewed perpendicular to the plane of the imidazole ring (left) and parallel to this plane (right). Structures with unsubstituted imidazole are shown as open circles (query of Fig. 3c), structures which satisfy the queries of Fig. 3(a) and 3(b) as squares and crosses, respectively. The spread of $M{-}\mathrm{N}$ distances is because of the presence of the four different metals (see Table 6 for details), but the angular characteristics of their distributions are not significantly different.

3.6. Metal–thiolate complexes

Metal–thiolate complexes are analogues for metal coordination by the cysteine side chain in anionic form. The numbers of complexes are fairly small and some were excluded because the chemical environment was very different from that in proteins. Geometrical data are given in Table 7. $d(\mathrm{S}{-}\mathrm{C})$ distances in the ligand are distributed around a mean of

Table 7
Metal–thiolate complexes.

(*a*) Metal–sulfur distances, $d(M{-}\mathrm{S})$, and numbers of structures (N_E) and observations (N_O). The search query in Fig. 4(*a*) was used. There are no thiolate complexes for Mg and none for Fe or Zn other than those shown; two structures with Mn, CN = 6, were found and five structures with $\mathrm{Cu^I}$ or $\mathrm{Cu^{II}}$, various coordination numbers.

Metal	CN	N_E/N_O	$d(M{-}\mathrm{S})$ (Å)		Oxidation state
			Mean	Range	
Mn	4	3/8	2.41 (2)	2.38–2.44	All 2
	5	5/22	2.33 (1)	2.30–2.36	All 3
Fe	4†	20/56	2.28 (2)	2.24–2.32	2, 3 or unassigned
	5	7/15	2.25 (4)	2.18–2.35	2, 3 or unassigned
	6	7/13	2.27 (11)‡	2.21–2.40‡	Mostly unassigned
Zn	4, 5	14/25	2.29 (4)	2.22–2.37	2

(*b*) Geometry of thiolate ligands, all metals.

	Mean (Å)	Range (Å)	Observations
$d(\mathrm{S{-}C})$, search query Fig. 4(*a*), excluding 4 outliers	1.83 (2)	1.79–1.88	115
$\angle(M{-}\mathrm{S{-}C})$, search query Fig. 4(*b*), excluding 1 outlier	101 (4)	91–115	70
$\lvert(M{-}\mathrm{S{-}C{-}C})\rvert$, search query Fig. 4(*b*), distribution is bimodal:			
Chelating ligands plus 2 others	36 (14)	6–72	52
Non-chelating ligands	128 (9)	75–180	18

† Includes a number of clusters related to $\mathrm{Fe_4S_4}$, some of which are designated as having additional Fe···Fe bonds. ‡ Omitting 2/5 with $d(\mathrm{S{-}C}) < 1.77$ Å.

1.83 (2) Å, close to the values of 1.817 (3) Å given by Allen *et al.* (1987) for uncoordinated $-\mathrm{S{-}C}(sp^3)$ and 1.822 (20) Å given by Engh & Huber (1991) for CH2E-S. A very large proportion of the ligands are chelating ligands, some multidentate; the ranges of angles $\angle(M{-}\mathrm{S{-}C})$ and possibly their mean will reflect this; also, torsion angles greater than ~100° are not possible with most chelating ligands.

4. Conclusions

It is practicable to tabulate means and ranges of distances for many combinations of metal and ligand using the results of accurate X-ray or neutron-diffraction structure determinations in the CSD. Sometimes, but not always, it is possible to assign means for specific coordination numbers or oxidation states. The values given here may be useful in model building or as target distances in restrained refinement, or they may be useful in assessing a metalloprotein structure which has been refined without restraints in the region of the metal atom. Substantial deviations of observed geometry from the norm would suggest either errors in coordinates or interpretation or a distortion at the metal site towards some form of active intermediate.

I am grateful to Dr Jenny Glusker for many useful discussions and for suggesting this topic in the first place; also to Dr C. K. Prout who welcomed us both into the Chemical Crystallography Laboratory, Oxford on study leave at the same time. I thank Professor M. D. Walkinshaw for laboratory space and computing facilities at Edinburgh and Dr P. Taylor for computational support there. I am grateful to Drs Frank Allen, W. D. S. Motherwell, G. P. Shields and M. Verdonk at the Cambridge Crystallographic Data Centre for help and for development versions of new programs. Finally, I thank Dr R. O. Gould, Edinburgh University for helpful comments, the EPSRC's Chemical Database Service where the earliest part of this work was performed, and Miss A. Cotton at Liverpool University who participated in it.

References

Allen, F. H. & Kennard, O. (1993*a*). *Chem. Des. Autom. News*, **8**, 1.

Allen, F. H. & Kennard, O. (1993*b*). *Chem. Des. Autom. News*, **8**, 31–37.

Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (1987). *J. Chem. Soc. Perkin Trans II*, pp. S1–S19.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bock, C. W., Katz, A. K. & Glusker, J. P. (1995). *J. Am. Chem. Soc.* **117**, 3754–3765.

Bock, C. W., Kaufman, A. & Glusker, J. P. (1994). *Inorg. Chem.* **33**, 419–427.

Brown, I. D. (1992). *Acta Cryst.* B**48**, 553–572.

Cambridge Crystallographic Data Centre (1999). *Metal Coordination in the CSD*. Cambridge Crystallographic Data Centre, England.

Carrell, A. B., Shimoni, L., Carrell, C. J., Bock, C. W., Murray-Rust, P. & Glusker, J. P. (1993). *Receptor*, **3**, 57–76.

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

Carugo, O., Djinovic, K. & Rizzi, M. (1993). *J. Chem. Soc. Dalton Trans.* pp. 2127–2135.

Chakrabarti, P. (1990*a*). *Protein Eng.* **4**, 49–56.

Chakrabarti, P. (1990*b*). *Protein Eng.* **4**, 57–63.

Cronin, L. (1997). D. Phil. Thesis, University of York, England.

Cronin, L. & Walton, P. (1999). In preparation.

Dodson, E. J., Davies, G. J., Lamzin, V. S., Murshudov, G. N. & Wilson, K. S. (1998). *Structure*, **6**, 685–690.

Einspahr, H. & Bugg, C. E. (1981). *Acta Cryst.* B**36**, 264–271.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.

EU 3-D Validation Network (1998). *J. Mol. Biol.* **276**, 417–436.

Katz, A. K., Glusker, J. P., Beebe, S. A. & Bock, C. W. (1996). *J. Am. Chem. Soc.* **24**, 5752–5763.

Martin, A. & Orpen, A. G. (1996). *J. Am. Chem. Soc.* **118**, 1464–1470.

Orpen, A. G., Brammer, L., Allen, F. H., Kennard, O., Watson, D. G. & Taylor, R. (1989). *J. Chem. Soc. Dalton Trans.* pp. S1–S82.

Simmons, C. J., Hathaway, B. J., Amornjarusiri, K., Santarsiero, B. D. & Clearfield, A. (1987). *J. Am. Chem. Soc.* **109**, 1947–1958.

Stebler, M. & Burgi, H. B. (1987). *J. Am. Chem. Soc.* **109**, 1395–1401.