Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination

Ivan Dokmanić,[a] Mile Šikić[a] and Sanja Tomić[b]*

[a]Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia, and [b]Rudjer Bošković Institute, Bijenička 54, HR-10000 Zagreb, Croatia

Correspondence e-mail: sanja.tomic@irb.hr

Metal ions are constituents of many metalloproteins, in which they have either catalytic (metalloenzymes) or structural functions. In this work, the characteristics of various metals were studied (Cu, Zn, Mg, Mn, Fe, Co, Ni, Cd and Ca in proteins with known crystal structure) as well as the specificity of their environments. The analysis was performed on two data sets: the set of protein structures in the Protein Data Bank (PDB) determined with resolution <1.5 Å and the set of nonredundant protein structures from the PDB. The former was used to determine the distances between each metal ion and its electron donors and the latter was used to assess the preferred coordination numbers and common combinations of amino-acid residues in the neighbourhood of each metal. Although the metal ions considered predominantly had a valence of two, their preferred coordination number and the type of amino-acid residues that participate in the coordination differed significantly from one metal ion to the next. This study concentrates on finding the specificities of a metal-ion environment, namely the distribution of coordination numbers and the amino-acid residue types that frequently take part in coordination. Furthermore, the correlation between the coordination number and the occurrence of certain amino-acid residues (quartets and triplets) in a metal-ion coordination sphere was analysed. The results obtained are of particular value for the identification and modelling of metal-binding sites in protein structures derived by homology modelling. Knowledge of the geometry and characteristics of the metal-binding sites in metalloproteins of known function can help to more closely determine the biological activity of proteins of unknown function and to aid in design of proteins with specific affinity for certain metals.

1. Introduction

The metalloproteome is defined as the set of proteins that have metal-binding capacity (Nelson & Cox, 2000). Metals are important for the biological activity of proteins and their removal or the replacement of one metal by another is often accompanied by a loss of or reduction in the biological activity of the protein (*e.g.* Thomson & Gray, 1992; Nelson & Cox, 2000; Bertini *et al.*, 2001; Dudev & Lim, 2003). Metal ions in proteins can act as structure promoters or can take part in enzymatic reactions. Divalent metal cations such as Zn^{2+}, Mg^{2+}, Cu^{2+} and Ca^{2+} are often associated with the catalytic or regulatory activities of proteins that constitute some of the fundamental chemical life processes (*e.g.* Nelson & Cox, 2000; Dudev & Lim, 2003; Nicholson, 1997; Cowan, 1998). For example, Mg in chlorophyll is important for photosynthesis,

Cu (together with Fe) has a role in oxygen-carrying proteins and Zn, Mg and Ca can serve as Lewis acids.

Knowledge of the type and number of amino-acid residues that coordinate a metal ion is important in order to clarify how specific they are for the presence of a certain metal and to obtain insight into their possible connection to the function of the protein.

The number of proteins of known three-dimensional structure is growing exponentially and so is the number of proteins with metal ions built in. This enables various statistical analyses of metals in proteins such as the relative abundance of different metal ions in proteins, as well as analysis of the type and frequency of amino-acid residues that participate in their coordination. All these data are valuable for the purpose of understanding how a protein selects a specific metal ion and the role of ions in protein function. Recently, a combined experimental and theoretical study of divalent metal ions in proteins by Babu *et al.* (2003) provided insight into the plausible role(s) of metal cations and catalytically important acidic residues in protein function. Also, the results of such an analysis are important for the identification of metal-binding sites in proteins for which the three-dimensional structure has been derived on the basis of the known structure of a homologue (either crystallographic or NMR).

In this study, we used the set of protein structures in the Protein Data Bank (PDB) determined with a resolution of 1.5 Å or better and a pre-calculated nonredundant PDB set (http://www.rcsb.org/pdb/statistics/clusterStatistics.do; see §2). Entries in the latter are chains extracted from clusters of chains with sequence similarity above 70%. In each cluster, the chains are ranked according to the quality factors (resolution and R value), deposition date and alphabetically. The best ranked chains are considered to be a representative data set. This consists of a total of 9999 PDB files including selected chains. When multiple copies of a metalloprotein were present in the crystal asymmetric unit only one copy was considered.

Marjorie M. Harding has performed a number of studies (Harding, 1999, 2000, 2001, 2004, 2006) on the geometry of the metal-binding sites in proteins, but on much smaller data sets and without attempting to find the specificities of a metal ion-binding site such as the distribution of coordination numbers (CN) and the combinations of amino-acid residues involved in coordination as a function of the CN. Our study is in many respects complementary to her studies and together they enable a better understanding of protein specificity and selectivity towards metal ions. These statistical analyses could further be related and completed by quantum-mechanical studies on metal ions (particularly Zn^{2+} and Mg^{2+}) and their ligands, as accomplished by Dudev and Lim (Dudev & Lim, 2001, 2002, 2004; Dudev *et al.*, 2003).

In the present study, we considered the metals Cu, Zn, Mg, Mn, Fe, Co, Ni, Cd and Ca. They predominantly have a valence of two and can be classified as trace elements in the majority of cells in living organisms. Apart from Na and K, they are the most commonly occurring metals in biological systems as well as in the Protein Data Bank (Berman *et al.*, 2000).

Recently, proteins have been considered as a tool for removing heavy metals from the environment in order to reduce pollution (Hennig, 1986; Sano *et al.*, 2006a,b). In this respect, knowledge of the preferred coordination surrounding a metal ion could help in choosing proteins that would be more effective in removing certain metal ions from the environment than those already available.

2. Method

As mentioned in §1, analysis was performed on two protein data sets. The metal-ion abundance in proteins, the distribution of their coordination numbers and the frequency of appearance of amino-acid residues and certain electron donors in the coordination sphere of metal ions were determined using the nonredundant set of chains, while the geometrical factors describing the coordination were determined using a set of protein structures with resolution ≤1.5 Å extracted from the PDB in October 2007. Redundancy is not critical for precise evaluation of the geometrical factors, but high resolution and a large number of events are necessary. In the selected set there are a total of 1400 entries in which the metal ions considered are present (the list is available on request). It should be noted that quite a few of the metal ions found in the selected PDB files are not directly coordinated to the protein atoms. This has no effect on most of the analyses we have undertaken, *e.g.* on the determinations of the mean distances between metal ions and side-chain functional groups or the determinations of the most common combinations of amino-acid residues in a metal-ion coordination sphere, but does influence analyses of the metal abundance in proteins and the distributions of metal ions with low coordination numbers.

The representative set of chains and corresponding PDB files was extracted from the PDB (April 2006) according to a pre-calculated nonredundant set of 'cluster70 chains' downloaded from the PDB site. From each cluster we chose the best ranking chain (and the corresponding PDB file) and checked whether any of the metal cations Cu^{2+}, Zn^{2+}, Mg^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cd^{2+} or Ca^{2+} was bound. For some of selected PDB files it was not possible to definitely associate the metal ion with exactly one protein (chain) and these structures were not considered in further analysis, *i.e.* the final statistics are based on a set of about 10 000 proteins comprising all protein types.

The coordination number is defined as the number of electron donors (O, N, S and Cl) found within 3 Å of the metal ion. Similarly, study of the amino-acid residue types taking part in metal coordination was performed in such a way that all amino-acid residues that have an electron donor within 3 Å of the metal ion are considered to be coordinating residues. Structures with a metal ion positioned at the symmetry element, *i.e.* when a metal ion lies at a special position, were analysed separately; we found a total of about 25 such structures (a list is available on request). For each metal ion the distribution of coordination numbers was determined as well as the abundance of the selected amino-acid residue triplets and quartets.

In cases where either the metal ion or some of its electron donors were disordered, the atom with the higher occupancy was selected. We found a total of 45 such cases.

Metal–carboxyl interactions were classified as bidentate or monodentate and distances were calculated separately for these two cases. A carboxyl group is considered to be monodentately coordinating in cases where only one of its O atoms participates in the coordination of a particular metal ion and bidentate where both atoms coordinate. Special cases in which the O atoms participate in the coordination of different metal ions (see, for example, Fig. 2 in Harding, 2006) are considered as bidentate interactions, not monodentate as in Einspahr & Bugg (1981).
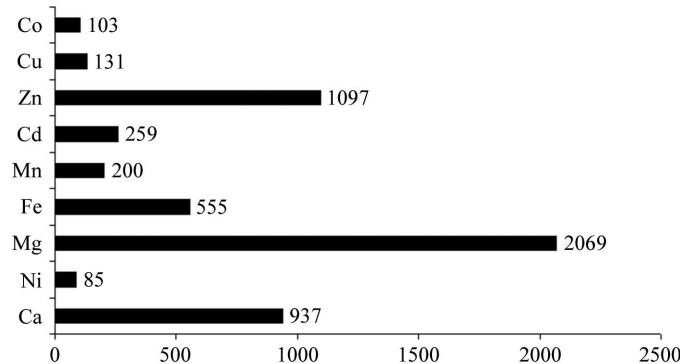
We have not discriminated water O atoms from the side-chain O atoms of Ser and Thr designated as O (hydroxyl). However, except in the case of Ca and Mg, the fraction of O atoms from these amino-acid residues in a metal coordination sphere is practically insignificant. Tyrosine may also coordinate a metal ion through an anionic O^- and we have considered this distance separately.

3. Results and discussion

In the set of nonredundant chains comprising all types of proteins the most abundant metal ion is Mg, followed by Zn, Ca, Fe, Mn, Cd, Cu, Co and Ni (see Fig. 1). Here, we should stress that artificially prepared proteins with a huge number of Fe ions were excluded from this analysis.

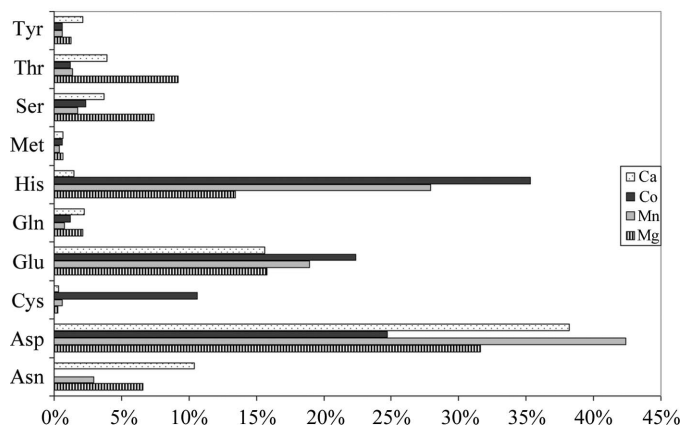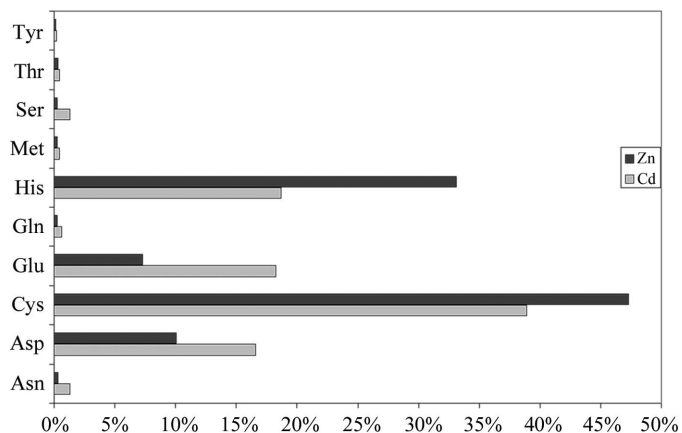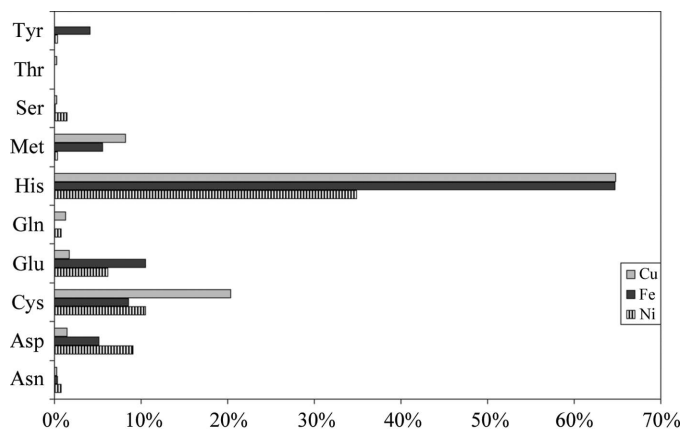### 3.1. Presence of amino-acid residues in a metal coordination sphere

The relative abundance of an amino-acid residue in the coordination spheres of the metals considered is given in Table 1. The amino-acid residues that most often coordinate the metal ions considered in this study are His, Asp, Cys and Glu, followed by Asn, Gly, Thr, Ser, Tyr, Met and Gln. With the exception of Gly, they usually participate in coordination using the side-chain electron donor. The other amino acids coordinate a metal ion solely through the carbonyl backbone (main-chain) O atom. If we consider the relative occurrence of

amino-acid residues in proteins (ROIP; Nelson & Cox, 2000), Cys and Asp exchange positions (see the last column in Table 1).

The relative presences of the most abundant amino-acid residues in the coordination spheres of each metal ion are shown in Fig. 2. The number of occurrences of each amino-

Figure 1
Metal-ion abundance in the set of nonredundant proteins from the PDB (in 2006).

Figure 2
Percentage of metal sites in which the amino-acid side-chain functional group is present as an electron-donating group. Statistics are only given for the amino-acid residues that most often participate in metal coordination (Asn, Asp, Cys, Gln, Glu, His, Met, Ser, Thr and Tyr). The metal ions are grouped according to the similarity of their coordination spheres: Cu^{2+}, Fe^{2+} and Ni^{2+} (top), Zn^{2+} and Cd^{2+} (middle) and Ca^{2+}, Mn^{2+}, Mg^{2+} and Co^{2+} (bottom).

Table 1
Occurrences of amino-acid residues in metal-ion coordination spheres (OAAMI).

The analysis was accomplished on the set of nonredundant proteins for Cu^{2+}, Zn^{2+}, Mg^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cd^{2+} and Ca^{2+}.

	OAAMI (%)	ROIP†	OAAMI/ROIP
Ala	1.1	7.80	0.14
Arg	0.5	5.10	0.10
Asn	4.2	4.30	0.99
Asp	22.3	5.30	4.21
Cys	18.8	1.90	9.92
Glu	12.3	6.30	1.94
Gln	1.1	4.20	0.27
Gly	2.7	7.20	0.38
His	25.3	2.30	11.02
Ile	0.8	5.30	0.16
Leu	0.7	9.10	0.08
Lys	0.9	5.90	0.15
Met	1.3	2.30	0.56
Phe	0.5	3.90	0.13
Pro	0.3	5.20	0.06
Ser	2.3	6.80	0.34
Thr	2.5	5.90	0.42
Trp	0.2	1.40	0.17
Tyr	1.3	3.20	0.40
Val	0.8	6.60	0.12

† Relative occurrence of amino-acid residues in proteins (Nelson & Cox, 2000).

acid residue and of water in the metal-ion coordination spheres is given in Supplementary Table 1[1].

Fig. 2 provides information about the specificity of the environment of the metal ions. Cu^{2+}, Fe^{2+} and Ni^{2+} are most often coordinated by His, followed by Cys and less often by Met, Asp and Glu. In the case of Cu^{2+}, Met is more abundant than Glu and Asp, while the opposite is the case for Fe^{2+} and Ni^{2+}. Zn^{2+} and Cd^{2+} are most often coordinated by Cys. The next most abundant amino-acid residues in the Zn and Cd coordination spheres are His, and Glu and Asp, respectively. Asp and Glu are the amino-acid residues that most frequently coordinate Ca^{2+}, Mn^{2+} and Mg^{2+} and are very often present in the coordination sphere of Co^{2+}. Mg^{2+} is frequently coordinated by the hydroxyl O atom of either Ser or Thr (see also Supplementary Table 1[1]). The presence of these two residues in the coordination sphere of other metals is low at <3%, although in the case of Ca^{2+} it is about 7%. Mg^{2+} and Ca^{2+} are also more frequently coordinated by the backbone carbonyl O atom than other metal ions; the most abundant electron donor is Gly. It is interesting that Cd and Fe, in contrast to all the other metal ions considered, are more often coordinated by Glu than by Asp.

3.2. The metal-ion coordination numbers

Distributions of metal-ion coordination numbers (as defined in §2) are given in Fig. 3. The absolute number of observed coordination numbers for each of the analysed metal ions is given in Supplementary Table 2[1]. The distribution of

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: EN5255). Services for accessing this material are described at the back of the journal.

coordination numbers and the type of coordination differ significantly among the studied metal ions. However, for the majority of metals the most frequent coordination numbers are four and six. Cu, Zn and Cd prefer coordination number four, while Mn, Co and Fe prefer CN six and Ca prefers seven. We found about the same number of Ni ions with CN four and six and similar numbers of Mg ions with CN five and six. Dudev & Lim (2004) noticed that zinc is four-coordinate in all structural sites, while five-coordinate and six-coordinate (rather than four-coordinate) structures represent energized states in zinc-containing enzymes. Apparently, most Zn ions in proteins are structure promoters.

Cases in which metal ions have a CN of three or lower should be treated with caution. It is possible that a metal ion interacts with water molecules that are not visible in the electron-density map or that there are mistakes in the crystal structures, in the sense that some amino-acid residue side chains are not in the correct positions. Most metal ions with low coordination number are positioned on protein surfaces and their presence can be explained by the crystallization procedure used. Such ions can often be found in the files for protein–DNA/RNA complexes, where they take part in stabilization of the bound polynucleotide (in the non-redundant set of protein structures, we found about 450 files with polynucleotides present). Omitting these files from the analysis resulted in a significant reduction in the number of ions with a low coordination number (for example, the number of Mg ions with CN below three fell from 280 to 184).

3.3. Correlation of amino-acid residues in the metal-ion coordination sphere

In addition to the coordination number, the specificity of a metal-ion environment is also characterized by the combination of amino-acid residues surrounding it and the type of electron-donor atom. For this purpose, we analysed how
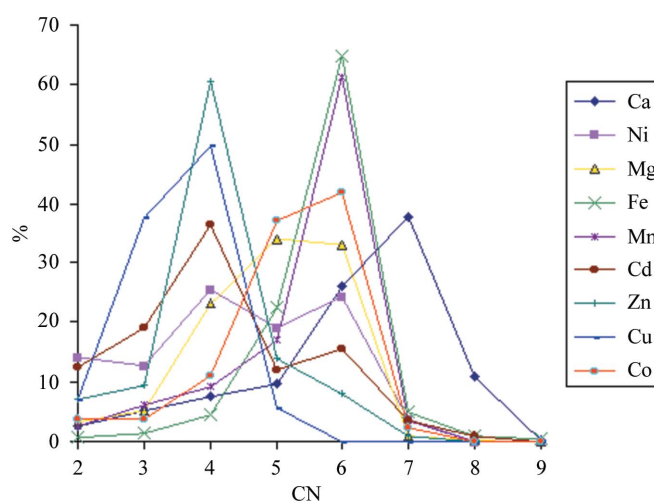
Figure 3
The relative distribution of coordination numbers (2–9) for various metal cations. The value for ions coordinated with a particular CN is divided by the total number of metal ions with coordination numbers in the range specified.

frequently certain combinations of His, Cys, Asp and Glu, *i.e.* the amino-acid residues that most often participate in co-ordination of the metal ions analysed, occur in the vicinity of a certain cation. We have not found any metal ions coordinated by five identical amino-acid residues, but a number of them are coordinated by four identical residues (Supplementary Table 3). We found 191 Zn^{2+} ions coordinated by four Cys residues. The majority of proteins with such geometry are transcription factors and DNA-binding proteins, followed by oxidoreductases, transferases and hydrolases. In most of these proteins (188) Zn has CN four. Fig. 4 shows the ratio of metal ions with CN four coordinated by selected quartets of amino-acid residues. 38 Cd^{2+} ions were coordinated by exactly four Cys residues and in the coordination sphere of two of them other amino-acid residues apart from the Cys quartet were also found. Four structures (all [NiFe] hydrogenases; PDB codes 1yqw, 1ubk, 1e3d and 1yq9) were found to contain Ni^{2+} coordinated by four Cys residues. His quartets were found in 14 proteins, most of which were oxidoreductases (nine cases). In seven of them Cu^{2+} was coordinated by four His residues, in one case in combination with water. In the coordination sphere of Fe^{2+} four His residues are combined with other amino-acid residues (for example Cys). We have not found a protein structure in which a metal ion is coordinated by only four Asp residues, but in 17 proteins we found 26 ions, 24 Ca^{2+} and two Mg^{2+}, that were coordinated by four Asp residues in combination with other amino-acid residues and/or water (Supplementary Table 3). Because of the similar nature of the side-chain electron donors of Asp and Glu, we considered these two amino-acid residues as equivalent. We found 70 metal cations, mostly Ca^{+2} (58 cases), that were coordinated by four Glu/Asp amino-acid residues. A quartet composed of an Asp/Glu combination constitutes the coordination sphere of about 12% of the calcium ions found.

Unlike Cys, Met coordinates metal ions *via* the uncharged S and we did not find any Met quartets. In general, Met rarely coordinates metal ions; it was found 59 times in the coordination sphere of six-coordinated Fe^{2+} and 32 times in four-coordinated Cu^{2+} (statistics for the nonredundant data set). We also searched for other quartets consisting of different combinations of Cys and His (see Fig. 4). We found that more than 200 ions, mostly Zn^{2+}, are coordinated by quartets made up of different combinations of these two amino-acid residues.

The most common homo triplet for metal ions with CN three was that consisting of histidines (Fig. 5). Only four Zn^{2+} ions with CN three were found to be coordinated by triplets of Cys, although this triplet commonly occurs in the coordination sphere of Zn^{2+} in combination with a heteroatom or other amino-acid residues.

3.4. The specificity of the electron donors

In accord with the results presented above, analysis of the atom types with which an amino-acid residue participates in metal-ion coordination revealed that Zn^{2+} (see Fig. 6) and Cd^{2+} are most often coordinated by the S atom from Cys, followed either by the imidazole N atom (Zn) or the hydroxyl O atom (Cd), while Cu^{2+} and Fe^{2+} are most often coordinated by the imidazole N atom. It is interesting that Mg^{2+} and especially Ca^{2+} are almost solely coordinated by oxygen. The hydroxyl O atom more frequently coordinates Mg^{2+} than the carboxyl O atom, while in the case of Ca^{2+} their frequencies are similar. Ca^{2+} is often coordinated by the main-chain carbonyl O atom of Gly. Ni^{2+} is most often coordinated by the
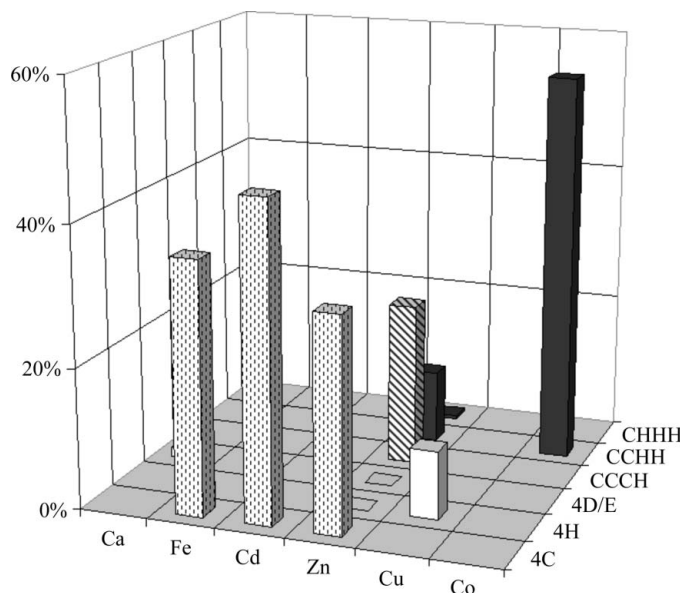
Figure 4
Percentage of metal ions with coordination number four coordinated by the specified quartets of amino-acid residues denoted by the normal single-letter codes. Quartets consisting of any Asp/Glu combination are denoted 4D/E. Analysis was accomplished on the nonredundant set of proteins.
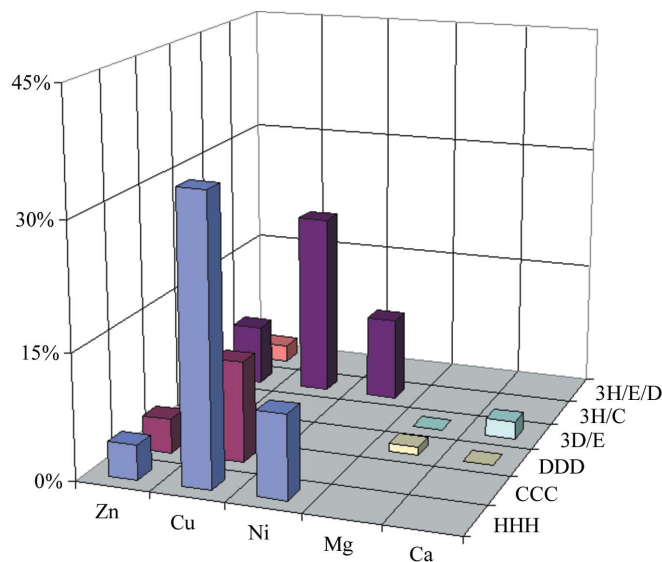
Figure 5
Percentage of metal ions with coordination number three coordinated by the specified triplets of amino-acid residues denoted by the normal single-letter codes. Triplets consisting of any Asp/Glu combination are denoted 3D/E and similarly those consisting of any His/Cys and of any His/Glu/Asp combinations are denoted 3H/C and 3H/E/D, respectively. Analysis was accomplished on the nonredundant set of proteins.

Table 2
Mean distances between metal cations and electron donors from different functional groups.

The mean distance (Å) is given in the first column, the standard deviation (SD) (Å) in the second and the number of occurrences of a functional group in the cation coordination sphere in the third column.

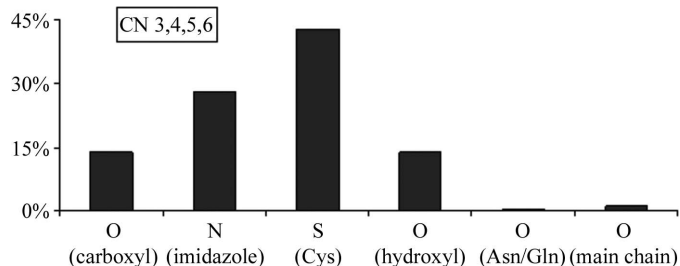|	O (carboxyl, mono)			N (imidazole)			S (Cys)			O (hydroxyl)†			O (Asn/Gln)			O (backbone carbonyl)		
	Mean	SD	No.	Mean	SD	No.	Mean	SD	No.	Mean	SD	No.	Mean	SD	No.	Mean	SD	No.
Ca	2.36	0.11	806	2.39	0.09	7	—	—	—	2.42	0.16	2036	2.35	0.10	219	2.36	0.13	1095
Ni	2.09	0.24	15	2.10	0.15	58	2.28	0.11	85	2.19	0.26	98	2.31	0.02	4	2.62	0.26	8
Mg	2.13	0.19	314	2.14	0.20	38	—	—	—	2.14	0.18	2458	2.22	0.18	31	2.38	0.28	197
Fe	2.11	0.14	72	2.08	0.11	642	2.31	0.06	178	2.19	0.19	218	2.45	0.41	2	2.74	0.28	57
Mn	2.13	0.15	132	2.20	0.10	147	2.59	—	1	2.19	0.20	243	2.31	0.27	5	2.12	0.14	70
Cd	2.32	0.15	38	2.23	0.22	132	2.51	0.05	39	2.38	0.23	449	2.21	0.02	8	2.43	0.19	59
Zn	2.11	0.20	189	2.07	0.12	752	2.33	0.07	276	2.18	0.28	501	2.39	0.34	8	2.27	0.27	57
Cu	2.17	0.04	3	2.03	0.10	279	2.19	0.07	87	2.23	0.34	85	2.19	0.13	6	2.5	0.30	20
Co	2.22 (2.19)‡	0.19 (0.13)‡	16	2.07	0.12	47	2.30	0.13	8	2.22	0.31	76	2.26	—	1	—	—	0

† Water O atoms and side-chain O atoms of Ser and Thr.   ‡ By replacing the current Asp168 rotamer in 1r6x with another of approximately the same energy, the mean distance dropped to 2.19 Å and the SD to 0.13 Å.

imidazole N atom (closely followed by the hydroxyl O atom) and Co^{2+} by the O atom from the carboxyl group of either Asp or Glu.

The presence of various atom types as a function of a metal coordination number is given in Supplementary Table 4. The abundances of certain types of electron donors and their correlation with the CN is closely related to the results on the metal environment defined by amino-acid presence, *i.e.* for Zn^{2+} and Cd^{2+} with CN four sulfur is the most frequent electron donor, while in the other cases Zn^{2+} is more often coordinated by imidazole N atoms, carboxyl O atoms and water, and Cd^{2+} by carboxyl O atoms and water (see Fig. 7).

3.5. Distances between metal ions and the amino-acid residues that coordinate them

The mean distances between the metal ions and the electron donors were determined using the set of protein structures with resolution ≤1.5 Å (see §2) from which structures with a metal ion (or any of the atoms that coordinate it) lying on a special position had been omitted. Despite a generously defined maximum coordination distance (3 Å), the mean distances between metal cations and their electron donors were much shorter than this cutoff (Table 2). The mean distances, as well as the discrepancy range determined from

the nonredundant set, were slightly larger than those determined from the structures with high resolution (data not shown). Similarly to Harding (2006), we also evaluated distances for structures in which carboxyl O atoms bidentately
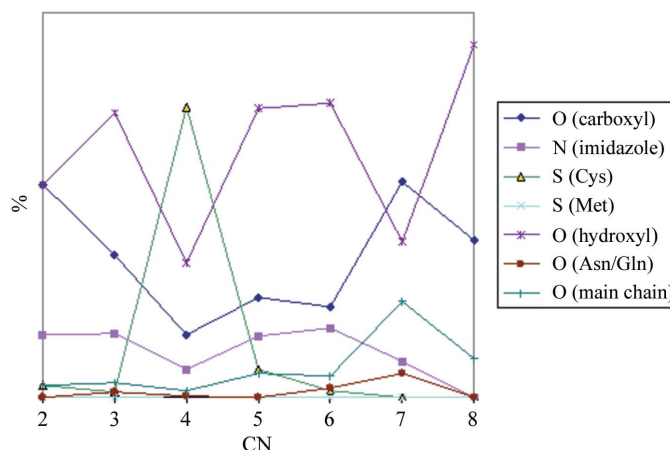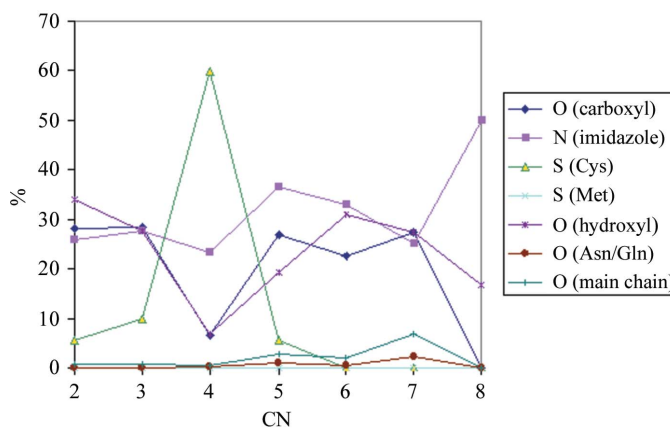
Figure 7
The normalized abundance of each functional group in the coordination sphere of Zn^{2+} (top) and Cd^{2+} (bottom) is given as a function of coordination number (CN). Analysis was accomplished on the non-redundant set of proteins. The value for the total number of electron donors used in the normalization includes several unclassified donors in addition to those specified in the legend.

Figure 6
The relative abundance of various functional groups in the coordination sphere of Zn^{2+}. O (hydroxyl) represents water O atoms and Thr and Ser side-chain O atoms. The side-chain O atoms of Asn and Gln are denoted as O (Asn/Gln).

coordinate metal ion(s) separately from those with monodentate coordination. The mean distances and standard deviation $\{SD = [\sum(d_i - \langle d\rangle)^2]^{1/2}/N\}$ are close to those determined by Harding (see Table 1 in Harding, 2006). However, our statistical sample is much larger. For example, for Zn and Mn Harding found 16 and 19 structures, respectively, with the metal monodentately coordinated to the carboxyl O atom, while we found 189 and 132 such structures, respectively. We found 16 structures with Co monodentately coordinated to carboxyl, with a mean distance of 2.22 Å. We noticed that in 1r6x this distance is quite long, 2.74 Å (Asp168 O—Co); by replacing the current rotamer of Asp168 with another one of practically the same energy, the monodentate carboxylate O—Co distance decreased to 2.19 Å. Harding did not find appropriate structures to enable her to evaluate the mean distances between Mn, Fe, Co, Cu and Zn and the main-chain carbonyl O atom, neither did she find distances between Ca and Mg and imidazole N atoms, while, with the exception of Co—O (main chain), we could evaluate these distances with reasonable precision.

In comparison to the distances to water O atoms and to Thr and Ser side-chain O atoms, the distance from a metal ion to the Tyr side-chain O atom is shorter, as expected. We found 28 structures in which Fe was coordinated to the O atom of Tyr and one structure each for Zn and Co. The calculated mean distance and standard deviation (for Fe) are 2.00 (0.08), 2.06 and 2.02 Å, respectively. There were two structures (1muw and 1xd3) in which we found Mg close to the side chain of Tyr, but in both cases the coordination distance was quite long (2.79 and 2.81 Å, respectively). In the case of 1xd3 this Mg is one of ten Mg ions, all of which were found at the protein surface and poorly coordinated by protein atoms, so we omitted 1xd3 from the present statistics. 1muw was determined with a resolution of 0.86 Å and H atoms are displayed. In this protein Mg is coordinated to the neutral Tyr.

We also determined the distribution of mean distances as a function of coordination number (see Supplementary Table 5).

For the other metals, the precision of the mean distances varies between the coordination numbers (4–6) according to the electron donor: for example, the Zn—S (Cys) distance is the most precisely determined for CN four [2.20 (0.06) Å], the Zn—N (His) distance is the most precisely determined for CN five [2.07 (0.12) Å] and the Zn—O (Asn/Gln) distance is the most precisely determined for CN six [2.27 (0.05) Å].

4. Concluding remarks

In this study, we analysed the binding-site specificity for the following metals: Cu, Zn, Mg, Mn, Fe, Co, Ni, Cd and Ca. The study, accomplished using a set of about 10 000 nonredundant protein chains from the Protein Data Bank (in 2006), revealed the frequency of amino-acid residue occurrences in the coordination spheres of these metals as well as a correlation between the number and type of electron donors in the metal-ion coordination spheres.

The quality of the force fields for protein simulation has greatly increased in the last decade; however, metal-ion parameters are still their weak point. Mean distances between metal cations and their electron donors, evaluated using a set of protein structures with resolution $\leq$1.5 Å (PDB, October 2007), should be of great help in improving these parameters.

The results obtained in this study may help in the identification and modelling of metal-binding sites in protein structures derived by homology modelling, as well as in the design of proteins with a specific affinity for a certain metal (so-called metal biosorbens). These investigations should also help biologists when replacing one metal ion in the protein with the another since they could provide a quick clue about which of the replacement ions would be the best choice with regard to the binding-site geometry and polarity.

References

Babu, C. S., Dudev, T., Casareno, R., Cowan, J. A. & Lim, C. (2003). *J. Am. Chem. Soc.* **125**, 9318–9328.

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

Bertini, I., Sigel, A. & Sigel, H. (2001). Editors. *Handbook of Metalloproteins.* New York: Marcel Dekker.

Cowan, J. A. (1998). *Chem. Rev.* **98**, 1067–1088.

Dudev, T. & Lim, C. (2001). *J. Phys. Chem. B*, **105**, 4446–4452.

Dudev, T. & Lim, C. (2002). *J. Am. Chem. Soc.* **124**, 6759–6766.

Dudev, T. & Lim, C. (2003). *Chem. Rev.* **103**, 773–788.

Dudev, T. & Lim, C. (2004). *J. Am. Chem. Soc.* **126**, 10296–10305.

Dudev, T., Lin, Y. L., Dudev, M. & Lim, C. (2003). *J. Am. Chem. Soc.* **125**, 3168–3180.

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

Harding, M. M. (1999). *Acta Cryst.* D**55**, 1432–1443.

Harding, M. M. (2000). *Acta Cryst.* D**56**, 857–867.

Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.

Harding, M. M. (2004). *Acta Cryst.* D**60**, 849–859.

Harding, M. M. (2006). *Acta Cryst.* D**62**, 678–682.

Hennig, H. F. (1986). *Environ. Health Perspect.* **65**, 175–187.

Nelson, D. L. & Cox, M. M. (2000). *Lehninger Principles of Biochemistry*, 3rd ed. New York: Worth Publishers.

Nicholson, A. W. (1997). *Ribonucleases: Structures and Functions*, edited by G. D'Alessio & J. F. Riordan. San Diego: Academic Press.

Sano, D., Myojo, K. & Omura, T. (2006a). *Water Sci. Technol.* **53**, 221–226.

Sano, D., Myojo, K. & Omura, T. (2006b). *Appl. Environ. Microbiol.* **72**, 6377–6380.

Thomson, A. J. & Gray, H. B. (1992). *Curr. Opin. Struct. Biol.* **2**, 155.