Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Analysis of copper–ligand bond lengths in X-ray structures of different types of copper sites in proteins

Luciano A. Abriata‡

Instituto de Biología Molecular y Celular de Rosario (IBR–CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, S2002LRK Rosario, Argentina

‡ Current address: Laboratory of Biomolecular Modeling, Institute of Bioengineering, School of Life Sciences, Swiss Federal Institute of Technology, EPF Lausanne, Switzerland.

Correspondence e-mail: luciano.abriata@epfl.ch

Received 27 February 2012
Accepted 8 June 2012

An updated picture of the ligand sets and copper–ligand atom bond lengths in proteins is presented which takes advantage of (i) the approximately twofold increase in the number of entries for copper-containing proteins in the PDB since the last study of this kind, especially benefiting from the recent incorporation of the structures of proteins involved in copper homeostasis, and (ii) a preliminary classification of copper sites based on their structural, electronic and functional features. This classification allowed the calculation of reliable target copper–ligand distances for several bonds that were not available in previous work and that are in good agreement with EXAFS data and the known chemistry of these sites. The analysis presented here further disclosed an artifactual dependence of the average of the reported Cu—NHis bond lengths on structure resolution, highlighting the importance of taking this into account when computing target distances even from high-resolution structures. Finally, a relationship between the two Cu—O distances in bidentate carboxylates is disclosed, similar to that reported previously for other metal ions.

1. Introduction

Copper is an essential element for life across all kingdoms, playing key roles in several redox enzymes and in long-range electron-transfer proteins (Messerschmidt *et al.*, 2001; Vila & Fernandez, 2001; Bleackley & Macgillivray, 2011; Maret, 2010; Banci *et al.*, 2010). In proteins it is found with oxidation states of +1 and +2 and is always bound to protein residues, in contrast to other metal ions which are often anchored to external cofactors (further information about protein copper sites is available in two handbooks devoted to metalloproteins; Vila & Fernandez, 2001; Messerschmidt *et al.*, 2001). Copper presents a large variety of possible coordination spheres that include the backbone and side-chain atoms of several different amino-acid types in the first ligand shell (Bertini *et al.*, 2010; Rubino & Franz, 2012). A description of the different ligand sets of copper sites and their coordination geometries is important in order to understand the chemistry and biology of this important metal ion and to improve the tables of metal–donor distances utilized in protein structure elucidation and modelling (Harding, 2006).

Although several studies have analyzed the ligand shells of the most common metal ions such as sodium, magnesium, calcium, zinc, cobalt and iron (Harding, 2001, 2002, 2006; Alberts *et al.*, 1998), studies on copper have been less detailed and have not considered the different coordination spheres and oxidation states possible for this metal ion. This was in part a consequence of the low number of structures available

Table 1
Description of the types of copper sites studied in this work, including information about function, physiologically relevant charges and spin states, consensus sequences and ligand sets, sample proteins and codes for sample structures deposited in the Protein Data Bank.

S is the spin state. In consensus sequences, capital letters indicate amino acids in one-letter codes and lower-case letters indicate the following: s, small side chain (Ala, Gly); h, hydrophobic amino acid; x, any amino acid.

Copper site	First-shell protein ligands
Type 1 mononuclear sites (electron-transfer blue proteins). Oxidized, Cu^{II}, $S = 1/2$. Reduced, Cu^{I}, $S = 0$.	$His_2CysMet$ (plastocyanin, 1plc; type 1 site of nitrite reductase, 1as7) $His_2CysGln$ (stellacyanin, 1jer) $GlyHis_2CysMet$ (azurin, 4azu) His_2Cys (type 1 site of laccase, 1hfu) Arranged in a consensus sequence of the form (G)Hx_{35-65}CssHhshM, histidines bound through ND.
Type 2 mononuclear sites (redox enzymes). Oxidized, Cu^{II}, $S = 1/2$. Reduced, Cu^{I}, $S = 0$.	Varied ligand sets of the form $N_{2-3}O_{0-2}$ + 0–2 H_2O/OH^-. His_4 (superoxide dismutase, 1eso) (Tyr)His_3(Tyr) (amine oxidase, 1av4) His_3 (Cu_H site of PHM, 1opm) His_2Met (Cu_M site of PHM, 1opm) His_3Glu (quercetin dioxygenase, 1juh) His_2Tyr_2 (galactose oxidase, 1gog) His_3(Asp) (type 2 site of nitrite reductase, 1as7)
'Homeostatic' mononuclear sites in copper pumps, chaperones, carriers and transcription factors. Bind Cu^{I} and/or Cu^{II}.	$His_{1-2}Met_{2-3}$ (CopC, 2c9q) $His_{1-2}Met_{2-3}$ (PCu_AC, DR1885, 1x9i) $HisMet_2Trp$ (CusF, 2vb2) $HisCys_2$ (Sco family, 2gqm) Cys_2 (Cox17, 3k7r) Cys_{2-n} (Atx1, Hah1, metallothionein, 1fd8, 1rju)
Type 3 binuclear sites (O_2-dependent oxidases). Oxidized, $Cu^{II}Cu^{II}$ with $S = 0$ and a low-lying excited state of $S = 1$.	Cu1, His_3; Cu2, His_3 Both ions bridged through one or two O atoms. Sample PDB entry 1aoz (ascorbate oxidase). Also found in tyrosinase, dopamine β-hydroxylase, particulate methane monooxygenase, haemocyanin and catechol oxidase.
Binuclear Cu_A sites (electron-transfer purple site in cytochrome *c* oxidase and nitrous oxide reductase). Oxidized, $Cu^{+1.5}Cu^{+1.5}$† (mixed-valence, $S = 1/2$). Reduced, $Cu^{I}Cu^{I}$, $S = 0$.	Cu1, $HisCys_2Met$; Cu2, $HisCys_2Obb$ (*T. thermophilus* COXII, 2cua). Arranged in a consensus sequence of the form $HsAx_{\sim30}$CsE/Q/WhCGxsHsxM.

† +1.5 is the formal charge of each copper ion in the oxidized resting-state form, which is a fully delocalized mixed-valence species.

for copper-containing proteins compared with other more studied metal ions, but is further complicated by the large diversity of coordination spheres and effects observed for copper (Bertini *et al.*, 2010; Rubino & Franz, 2012). This diversity produces broad distributions for Cu–ligand distances if all types of copper sites are treated together, resulting in target distances of low reliability and that may be accurate for some but not all of the possible oxidation states and site types. For example, the target distance reported by Harding for Cu—NHis bonds of 2.02 ± 0.1 Å (Harding, 2006) is reasonable for mononuclear Cu^{I} sites but is ∼0.05–0.1 Å too long for Cu^{II} sites according to EXAFS data, which is the technique of choice for the precise measurement of metal–donor distances and has a typical uncertainty of 0.02 Å.

In addition to the radically different features of each copper site, Jahn–Teller effects and distortions induced by the entatic/rack-induced state (Solomon *et al.*, 2004; Williams, 1995; Malmström, 1994) further complicate the bond-length distributions. Finally, any subtle patterns in the distributions can be blurred by the low resolution of protein structures compared with those of small complexes. It is thus clear that preliminary classification of the copper sites is expected to remove some of the effects that lead to broadening and artifacts in the distributions of Cu-atom distances in proteins, disclosing simpler patterns and more accurate Cu-atom target distances. With this aim, the work presented here takes advantage of the latest increase in the number of structures of copper proteins released by the Protein Data Bank (an almost twofold increase since the last study in 2006 and doubling approximately every six years for the last 20 years; Supplementary Figure S1[1]) and two recent publications which have reviewed and reorganized the different types of copper sites in proteins (Bertini *et al.*, 2010; Rubino & Franz, 2012). In particular, the present study benefits from the availability of several structures of proteins involved in copper homeostasis and delivery which have only recently been the subject of structural studies (Abajian & Rosenzweig, 2006; Cobine *et al.*, 2006; Abriata *et al.*, 2008; Banci *et al.*, 2005; Abajian *et al.*, 2004).

The classification of copper sites summarized by the groups of Bertini and Rubino and employed in this work was

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MH5061). Services for accessing this material are described at the back of the journal.

defined through decades of studies based on the structural, spectroscopic and functional features of the different sites.

(i) Type 1 sites are mononuclear copper sites typical of blue electron-transfer copper proteins, also found in the blue sites of multicopper oxidases (Solomon & Hadt, 2011). In these sites the copper ion is coordinated to the ND atoms of two histidines and the S atom of a cysteine residue as equatorial ligands in the base of a nearly trigonal pyramid, plus zero, one or two variable ligands in the axial position. The main characteristic of type 1 centres is the highly covalent Cu—SCys bond, which gives rise to strong absorption bands in the visible region of the absorption spectrum (conferring the distinctive blue colour) and a low $A_{\parallel}$ value in the EPR spectrum of the oxidized sites. Type 1 sites perform efficient long-range electron transfer by cycling between the +1 and +2 redox states with minimal geometric rearrangements. These features are very different to those observed for small unstrained copper complexes with similar ligands and arise from the strain imposed by the protein in what has been named the entatic or rack-induced state (Solomon *et al.*, 2004; Williams, 1995; Malmström, 1994).

(ii) Type 2 sites are mononuclear copper sites typical of single-copper redox enzymes, in which copper is coordinated to 4 N and O atoms (typically from histidines, carboxylates, backbone atoms, water molecules or tyrosine residues) in a nearly square-planar arrangement that can be completed by an axial N/O atom. Their electronic features in the oxidized state are similar to those of small Cu^{II} complexes, *i.e.* little or no external strain, large $A_{\parallel}$ values in the EPR spectrum and low absorption in the visible region. Cycling between the +1 and +2 redox states takes place during catalytic turnover with variable structural rearrangements.

(iii) 'Homeostatic' sites are mononuclear sites present in proteins involved in copper homeostasis, such as copper chaperones, copper transporters and copper-dependent transcription factors (Davis & O'Halloran, 2008; Bertini *et al.*, 2010), which bind Cu^{II} and/or Cu^{I}. They are typically unstrained sites and contain His, Cys and/or Met as ligands. Cu^{II} sites display EPR properties typical of type 2 sites but with stronger UV–visible absorption owing to Cu—SCys charge-transfer transitions when Cys is present, although the lack of strain produces lower Cu—S covalency, giving rise to yellow–orange instead of blue colours.

(iv) Type 3 sites are binuclear O-bridged copper sites of O_2-dependent oxidases with a pair of antiferromagnetically coupled Cu^{II} ions in the oxidized resting state, which is diamagnetic, as opposed to all other oxidized copper sites (Kosman, 2010). In the reduced state the oxygen bridge is broken and the copper ions are separated, no longer being a

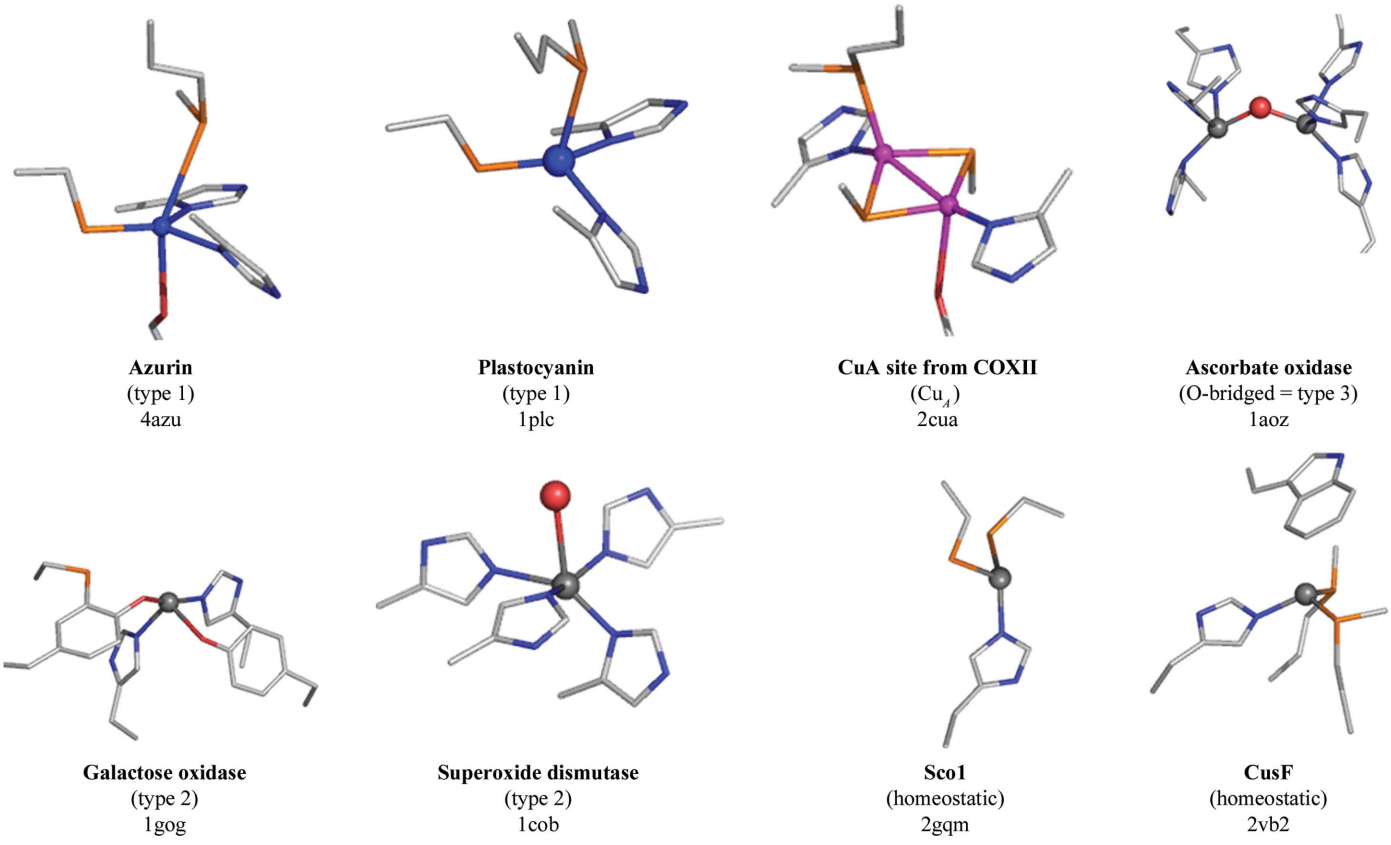

Figure 1
Examples of the types of copper sites studied in this work. Pictures were rendered using *PyMOL* v.0.97 from the indicated PDB entries, displaying copper ions as spheres (blue for type 1, purple for Cu_A and grey for the rest), copper-bound O/water atoms of type 2 and type 3 sites as red spheres and copper ligands as sticks.

Table 2
Summary of Cu-atom bond lengths (averages, with standard deviations in parentheses) for atoms found 20 or more times as first-shell ligands.

Distances with no labels are considered to be reliable for structure modelling and refinement. Distances with large standard deviations are labelled † when the variability is real as determined from data other than X-ray structures or labelled ‡ when the variability arises from unresolved broad distributions, unresolved dependence of the average of the reported distances on structure resolution and/or a low number of observations. T.B. indicates that the distribution is too broad to compute averages, never reaching zero or showing distinct populations. Values in parentheses on the second line show the number of Cu-ligand entries employed in each calculation. Values in square brackets on the third line indicate the range of observed values for bond lengths with broad distributions. As explained in the text for each case, some of these averages were computed using distributions trimmed at either 3 Å (H_2O/OH^- in type 2) or 3.4 Å (ODAsp and OEGlu) or at the population of shorter distances (Obb of type 2 and homeostatic sites). The two Cu—O distances in bidentate ODAsp and OEGlu are constrained by the relationship $d_1 + 0.6234$ Å $= 9.573/(d_2 + 0.6234$ Å). For all mononuclear sites, only structures with resolution better than 2.0 Å were analyzed (except for NDHis and NEHis, the values for which correspond to extrapolations to 0 Å in plots of average reported distance against resolution, as shown in Supplementary Fig. S3). For Cu_A and type 3 sites, all structures with resolution better than 2.5 Å were included in the analysis.

	Mononuclear Cu^{II}			Mononuclear Cu^{I}				
Ligand	Type 1	Type 2	Homeostatic	Type 1	Type 2	Homeostatic	Cu_A	Type 3
SCys	2.17 (9) (338)	—	2.23 (11) (144)	2.20 (11) (76)	—	2.26 (15) (116)	2.30 (7) (208)	—
SMet	2.83 (33)† (276) [2.25–3.57]	—	2.58 (41)† (55) [1.97–3.49]	2.81 (38)† (60) [2.12–3.34]	—	2.47 (22)† (24) [2.07–3.01]	2.5 (2)† (52) [2.11–2.99]	—
NDHis	1.96 (4) (1085)	1.92 (7) (582)	1.94 (14) (56)	2.02 (9) (142)	1.99 (9) (125)	1.98 (9)‡ (29) [1.88–2.11]	1.98 (11) (104)	1.99 (20)‡ (29) [1.85–2.45]
NEHis	—	1.95 (6) (2075)	1.96 (13) (35)	—	1.96 (12) (248)	—	—	2.01 (20)‡ (178) [1.94–2.55]
Obb	T.B.	2.41 (27)† (84) [1.88–2.97]	2.51 (17)† (23) [1.96–2.75]	T.B.	—	—	2.43 (17)† (52) [2.07–2.81]	—
Nbb	—	2.26 (25)† (117)	—	—	—	—	—	—
O (H_2O, OH^-, Obridge)	—	2.35 (34)† (573) [1.48–3.00]	—	—	—	—	—	2.1 (3)‡ (76) [1.78–2.68]
ODAsp and OEGlu	—	2.63 (62)‡ (273) [1.53-3.49]	—	—	—	—	—	—
Others	NEGln, 2.23 (12)‡ (21) [2.12–2.63]	OTyr, 2.38 (33)‡ (16) [1.77–2.94]	—	—	—	—	—	—

binuclear site. In all cases, each copper ion is coordinated to the NE or ND atoms of three histidine residues.

(v) Cu_A sites are binuclear electron-transfer centres that rely on two copper ions bridged through two SCys atoms, each ligand shell being completed strictly by one histidine and one axial methionine or backbone O atom. The oxidized resting state is a mixed-valence pair in which the formal charge of each copper ion is +1.5 because the only unpaired electron is fully delocalized, as shown by the seven-line hyperfine structure in the $g_{||}$ region of the EPR spectrum (Kroneck *et al.*, 1988, 1990) and as further supported by several types of spectroscopies and quantum-mechanical calculations (Gorelsky *et al.*, 2006; Abriata *et al.*, 2009; Xie *et al.*, 2008; Gamelin *et al.*, 1998).

The key features of the studied sites are summarized in Table 1 and include the possible oxidation and spin state(s) of the copper ion(s) and the corresponding native consensus sequences and ligand sets. The table also lists examples of proteins with sites of each type. Sample three-dimensional structures of selected copper sites are shown in Fig. 1. In addition to the recent specific reviews cited above for each type of site, further descriptions and examples of the different types of copper sites are available in handbooks and reviews (Messerschmidt *et al.*, 2001; Bertini *et al.*, 2010; Vila & Fernandez, 2001; Rubino & Franz, 2012). Other site types such as the tetranuclear CuZ were not analyzed owing to the low number of structures available.

2. Methods

The *Ligand Expo* server of the Protein Data Bank (http://ligand-expo.rcsb.org/index.html; Berman *et al.*, 2002; Bernstein *et al.*, 1977) was searched for proteins containing CU1, CU, CUA and C2O chemical components, which correspond to mononuclear reduced Cu^{I} (including type 1, type 2 or homeostatic), mononuclear oxidized Cu^{II} (including type 1, type 2 or homeostatic), Cu_A binuclear (only oxidized structures are available) and oxygen-bridged type 3 binuclear centres, respectively. Other copper-containing chemical components from *Ligand Expo* were not studied owing to their lower number of entries in the PDB. X-ray structures with resolution better than 2.5 Å were initially downloaded, resulting in 119 PDB files for proteins with CU1 sites, 555 files for CU sites, 26 files for Cu_A sites and 16 files for C2O sites. After analyzing the distribution of resolutions (Supplementary Fig. S2), only structures with resolution better than 2.0 Å were analyzed in the case of mononuclear sites, whereas in the cases of the CUA and C2O chemical components all structures

were included in the analysis. Note that most files produce more than one copper-site entry because of the number of similar proteins in the unit cell and/or because the protein itself has more than one site. Any possible sources of error arising from photoreduction were not handled.

Classification of the copper sites according to the different types described above was performed as follows. Chemical components CUA and C2O correspond exactly to Cu_A and type 3 sites, respectively, in their oxidized states, so no special treatment was needed for classification. On the other hand, the chemical components 'CU' and 'CU1' correspond to all mononuclear sites in the oxidized (CU) and reduced CU1 states, respectively, regardless of their classification as either type 1, type 2 or homeostatic. Filters based on the known ligand sets (built from the description above, Table 1 and the work of Bertini and coworkers) were applied to the mononuclear sites, returning classifications into type 1, type 2 or 'homeostatic' sites, each in reduced or oxidized states. For each group of copper sites, bond-length data were then computed from the PDB files using Python BioPDB scripts.

First-shell ligands were searched within a sphere of 3 Å around the Cu atoms to build histograms of copper–donor distances. When the histograms did not fall off to zero at 3 Å the search was extended to 4 Å to probe weaker and second-shell ligands. As described in §3, the reported target distances for Cu—NDHis and Cu—NEHis correspond to extrapolations to a resolution of 0 Å in plots of averaged bond length *versus* binned structure resolution. All other distances are averages for all structures (of resolution better than 2 Å for CU and CU1 and better than 2.5 Å for CUA and C2O). No bond lengths are reported for ligand atoms for which distributions did not fall off to zero at 4 Å or when the number of observed bonds was less than 20.

Ligand sets are only reviewed and used in this work to achieve a classification of copper sites based on the analysis of structures of wild-type proteins prior to bond-length analyses, guided by the recent articles by Bertini and Rubino (Bertini *et al.*, 2010; Rubino & Franz, 2012).

3. Results and discussion

3.1. Classification of copper sites prior to structural analysis

The Protein Data Bank organizes copper ions into so-called 'chemical components', which may or may not correspond exactly to the different types of sites as classified according to spectroscopic and functional data. The chemical component 'CU' gathers all mononuclear copper ions in the oxidized state (*i.e.* Cu^{II} ions in type 1, type 2 and homeostatic sites). In the same way, 'CU1' gathers all mononuclear copper ions in the reduced state (Cu^{I} ions in type 1, type 2 and homeostatic sites), 'CUA' corresponds to the two copper ions in oxidized Cu_A sites and 'C2O' corresponds to the two Cu^{II} ions in oxidized type 3 sites. Thus, while the CUA and C2O chemical components already correspond exactly to the structural and functional definitions of Cu_A and type 3 sites, respectively, it was necessary to classify each CU and CU1 copper ion into its corresponding type. This was performed on the basis of their fingerprint structural features as described in §1 and in Table 1. The classification process produced eight classes of copper sites, for which the detected ligand atoms are distributed as shown in Fig. 2.

3.2. Analysis of bond lengths and calculation of site-specific target distances

The availability of several Cu-atom bond lengths for different types of copper sites in both oxidation states

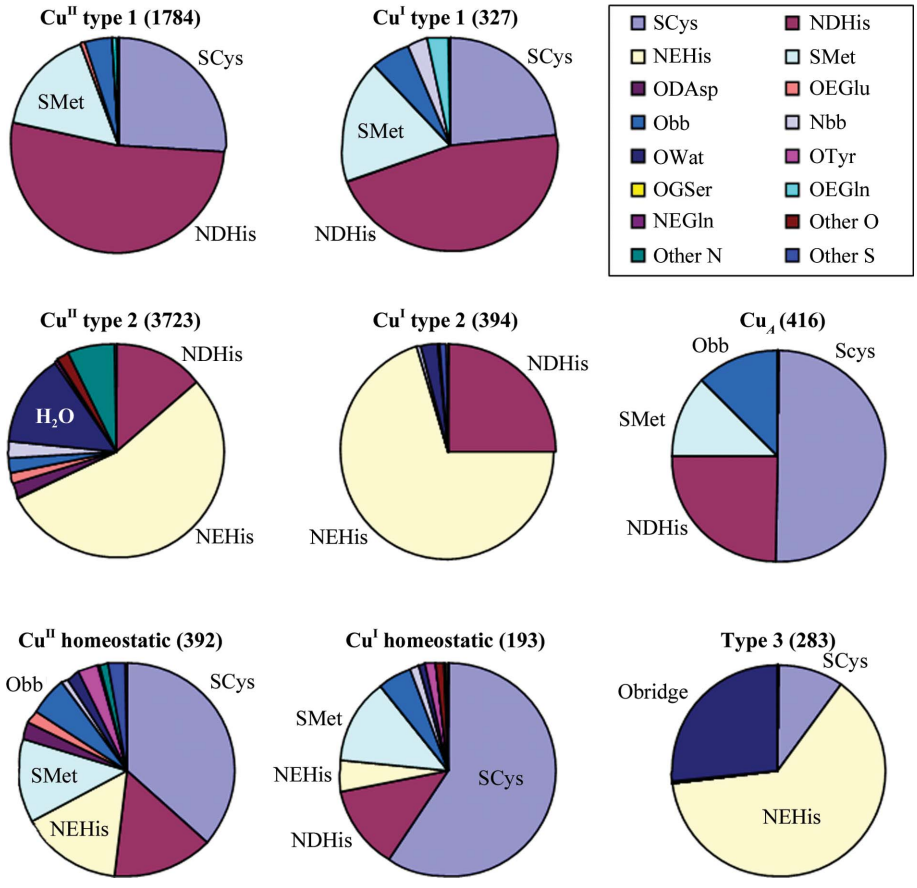

Figure 2
Pie charts showing the fraction of copper–ligand bonds retrieved for each kind of copper site. Within each group, the fractions compare well with ligand sets expected from consensus sequences as determined in previous studies and summarized in Table 1. The numbers in parentheses indicate the total numbers of retrieved bond lengths.

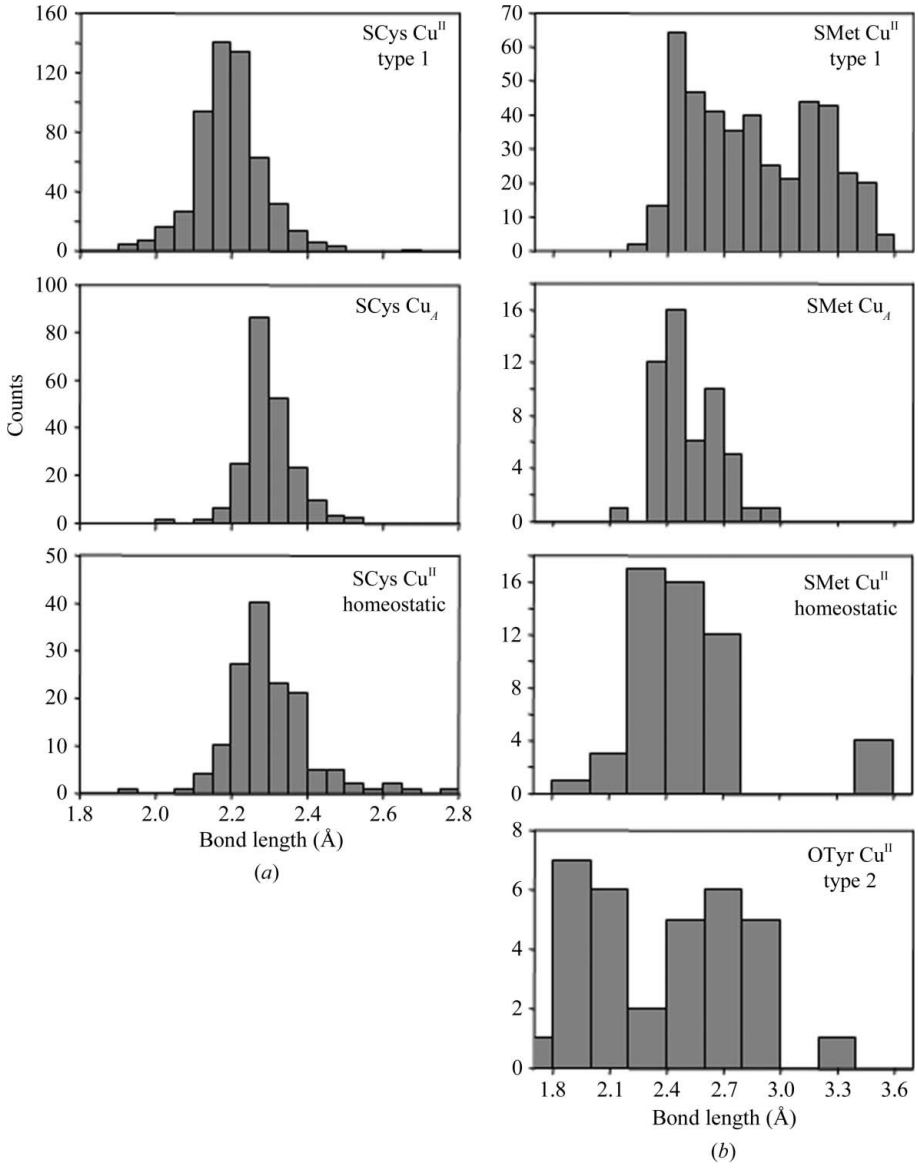

Figure 3
Distributions of (*a*) Cu—SCys and (*b*) Cu—SMet and Cu—OTyr bond lengths (discussed in the text).

provides a good starting point for the calculation of new target distances taking into account (i) oxidation state, (ii) the existence or absence of strain induced by the protein (which is of significant magnitude in type 1 and Cu$_A$ sites according to the low reorganization energies of copper in these sites) and (iii) part of the Jahn–Teller effects (since axial and equatorial ligands are usually provided by different residues, at least within site types). Cu-atom bond lengths were calculated for each class of copper site and analyzed to derive class-specific target distances, which are given in Table 2. Descriptions of the less-represented atom ligands and comments on the similarities and differences in bond lengths between and within site types are given in the Supplementary Material.

3.2.1. Cu-atom bond lengths for SCys and SMet. The average values obtained for all Cu—SCys and Cu—SMet bond lengths are in good agreement with the values derived from EXAFS data for specific proteins (examples are given in Supplementary Table S1). The standard deviations determined for Cu—SCys bonds are small enough to include the reported averages as reliable target distances. The differences observed between target distances for Cu—SCys bonds in the different types of sites (discussed in the Supplementary Material) highlight the importance of preliminary classification. Standard deviations for Cu—SMet bonds are much larger than those for Cu—SCys bonds and reflect their broader distributions (compare Figs. 3*a* and 3*b*). However, this variation arises from true structural differences as shown by the EXAFS data (Supplementary Table S1) and is actually important for fine-tuning the properties of copper sites (Garner *et al.*, 2006; Marshall *et al.*, 2009; Ledesma *et al.*, 2007). Thus, no reliable target distances can (or should) be provided for Cu—SMet bonds from statistics only.

3.2.2. Cu-atom bond lengths for NDHis and NEHis. Contrary to what was observed for Cu—SCys and Cu—SMet bonds, the averages for all Cu—NDHis and Cu—NEHis bond lengths are not consistent with the EXAFS data. Instead, the crystallographic averages are systematically larger by ∼0.1 Å. Similar differences were observed for Cu—NHis and other metal-atom bond lengths in previous works, in which this artifact was attributed to the intrinsically low resolution of protein structures (Cheung *et al.*, 2000; Tamames *et al.*, 2007; Zheng *et al.*, 2008). In order to better assess the impact of structure resolution on the computed bond lengths for histidine ND and NE atoms, they were recomputed on four subsets of structures binned by resolution and the averages were plotted against the mid-resolution of each subset (Supplementary Fig. S3*a*). The average Cu—ND and Cu—NEHis bond lengths for proteins solved at a resolution better than 1 Å still gave systematically larger values, now by ∼0.05 Å.[2] The plots further revealed that the averages of the reported distances are longer at worse resolutions, with a linear trend described by an average slope of 0.067 for the best represented mononuclear sites. Interestingly, extrapolation of these plots to a resolution of 0 Å seems to remove the artifact

[2] Similarly, Harding's target distance for Cu—NHis computed from structures of resolution better than 1.25 Å turns out to be overestimated by ∼0.05–0.1 Å for oxidized mononuclear copper.

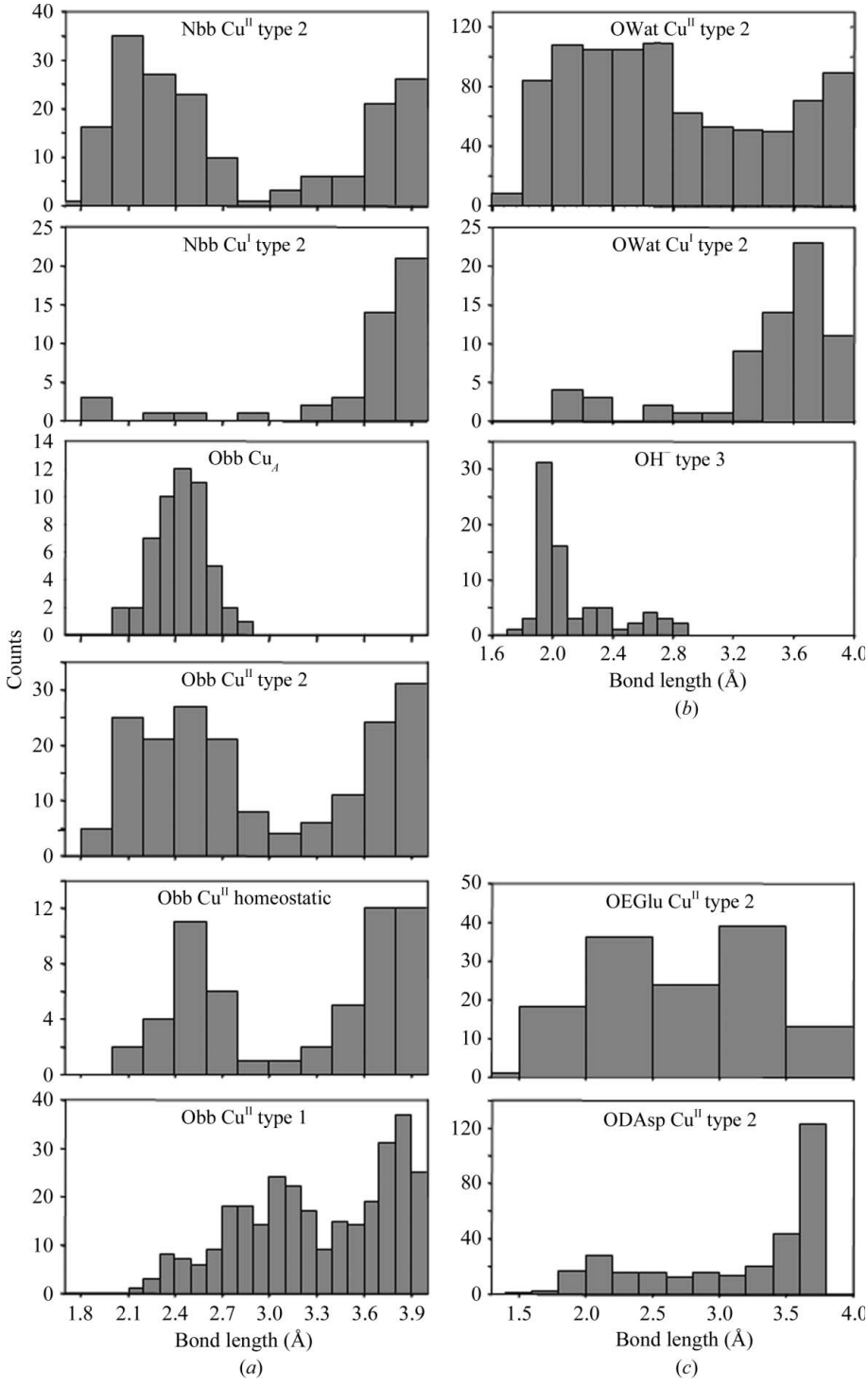

Figure 4
Distributions of (*a*) Cu—Nbb and Cu—Obb, (*b*) Cu—Owater/OH$^-$/bridgingO and (*c*) Cu—Ocarboxylate bond lengths (discussed in the text).

introduced by the low resolution, since the intercepts agree very well with EXAFS-derived bond lengths for all types of copper sites (with the exception of type 3 sites, probably owing to the low number of entries and the low resolution of the structures that contain these sites). The target distances reported in Table 2 for NDHis and NEHis in all but Cu^{I} homeostatic sites and type 3 sites correspond to these extrapolations. For the two exceptions, extrapolations gave inconsistent results owing to the low number of observed values, so they were estimated by correcting their averages with the average slope of the plots for mononuclear sites (and hence are indicative only). In brief, it appears that including structures of good to modest resolution, at least up to 2.5 Å, and considering their resolutions in the analysis produces better values for target distances than the analysis of only high-resolution structures.

Looking back to the Cu—SCys and Cu—SMet data, no correlations were evident when the average of the reported bond lengths was plotted against binned resolution (Supplementary Fig. S3*c*). Consistently, as stated above, their overall averages already match the EXAFS values well. Thus, it appears that the artifact introduced by the limited resolution is negligible for Cu—SCys and Cu—SMet bond lengths, probably because they are longer than Cu—NHis bonds.

3.2.3. Cu-atom bond lengths for backbone O and backbone N. In Cu$_A$ sites, Cu—Obb bond lengths are distributed similarly to Cu—SMet bond lengths, in agreement with the symmetric nature and stiffness of the centre, whereas the distribution is broader in the less rigid mononuclear sites (Figs. 3*b* and 4*a*). In particular, the distributions of Cu—Obb bond lengths in type 1, homeostatic and type 2 sites do not fall off to zero at any distance and appear to show two populations probably corresponding to first-shell and second-shell ligands (Fig. 4*a*). The values given in Table 2 for Cu—Obb distances correspond to averages for the populations of shorter bond lengths in type 2 and homeostatic sites, while the populations cannot be resolved for type 1 sites and hence no value is reported. In any case, the reported target distances are not reliable for modelling because of the true dispersion in values, similar to the situation for SMet.

The number of collected Cu—Nbb (deprotonated peptide N) bond lengths is only large for oxidized type 2 sites. Similar to the case for Obb atoms, there are two broad peaks arising

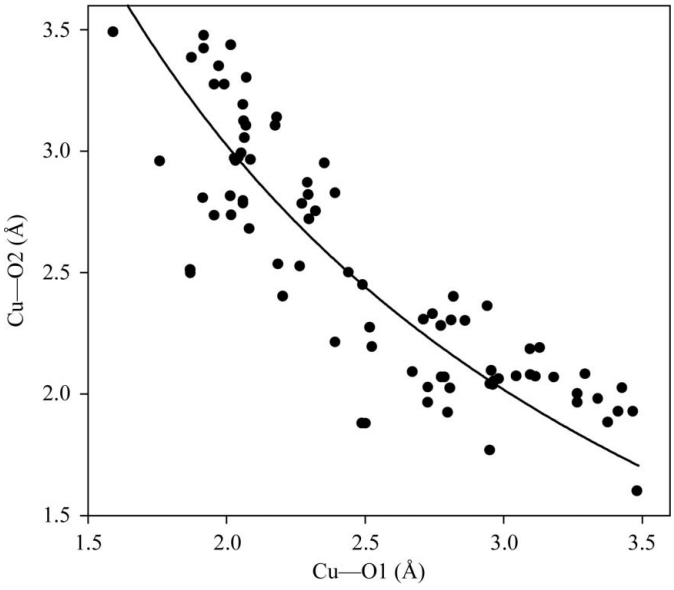

Figure 5
The relationship between the two Cu—O distances in bidentate coordination of Cu^{II} by carboxylate groups. The axes can be permuted, so each distance is shown twice and the plot is symmetrical about the diagonal line.

from first-shell and second-shell Nbb atoms (Fig. 4*a*). Again, the average reported in Table 2 corresponds to the first peak and is not reliable for modelling.[3]

3.2.4. Cu-atom bond lengths for water/hydroxide/bridging O atoms. O atoms of water and hydroxide ligands are very common in oxidized type 2 and type 3 sites. Although they should correspond to different bond lengths, they are not distinguishable in X-ray structures hence the given averages are indicative only. This fact and the expected true variability explain the broad distributions observed for type 2 sites (Fig. 4*b*). Also as expected, the distribution of bond lengths for the bridging O atom of type 3 sites is sharper and shorter on average than that for H_2O/OH^- in type 2 sites.

3.2.5. Cu-atom distances for carboxylic O atoms. Carboxylic O atoms from Glu and Asp are only abundant in structures of oxidized type 2 sites. Their distance distributions are broad (Fig. 4*c*), leading to large standard deviations that preclude the calculation of precise target distances (the average value reported in Table 2 is only indicative and was computed by trimming the distances at 3.4 Å). However, as shown previously for Zn^{II} and Mg^{II}, broad distributions in carboxylates arise from bidentate chelation in such a way that the two metal—O lengths become inversely correlated (Harding, 2006). In the data set of type 2 Cu^{II} sites analyzed here, a similar correlation was observed between the Cu—Ocarboxylate lengths (Fig. 5), which are constrained by the relationship $d_1 + 0.6234\ \text{Å} = 9.573/(d_2 + 0.6234\ \text{Å})$.

[3] The limited amount of EXAFS data indicates that part of the variation might be genuine.

4. Conclusions

An updated study of copper–ligand bond lengths is presented for the most common types of copper sites in proteins (Table 2). Many of the reported values stand as reliable target distances for modelling and refinement, with standard deviations that are smaller than those computed previously, and are supported both by EXAFS data and the known chemistry of copper. For some ligands, genuine variability in bond lengths precludes the assignment and prediction of accurate target distances; hence, they should be handled with caution in modelling and refinement protocols.

This work further highlights the importance of (i) performing a preliminary classification of the copper sites before choosing or computing target distances, (ii) including not only very high resolution structures in the computation of target distances but also structures of moderate resolution and (iii) binning the data by resolution in these computations in order to detect and resolve artifacts that cause the obtained target distances to deviate from their true values even at the best resolutions achievable in protein crystallography.

CONICET, EMBO and the Marie Curie Actions are acknowledged for postdoctoral fellowships. Three anonymous reviewers are acknowledged for their very useful suggestions and comments. María Eugenia Zaballa is acknowledged for helpful comments and for proofreading the manuscript.

References

Abajian, C. & Rosenzweig, A. C. (2006). *J. Biol. Inorg. Chem.* **11**, 459–466.
Abajian, C., Yatsunyk, L. A., Ramirez, B. E. & Rosenzweig, A. C. (2004). *J. Biol. Chem.* **279**, 53584–53592.
Abriata, L. A., Banci, L., Bertini, I., Ciofi-Baffoni, S., Gkazonis, P., Spyroulias, G. A., Vila, A. J. & Wang, S. (2008). *Nature Chem. Biol.* **4**, 599–601.
Abriata, L. A., Ledesma, G. N., Pierattelli, R. & Vila, A. J. (2009). *J. Am. Chem. Soc.* **131**, 1939–1946.
Alberts, I. L., Nadassy, K. & Wodak, S. J. (1998). *Protein Sci.* **7**, 1700–1716.
Banci, L., Bertini, I., Ciofi-Baffoni, S., Katsari, E., Katsaros, N., Kubicek, K. & Mangani, S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 3994–3999.
Banci, L., Bertini, I., McGreevy, K. S. & Rosato, A. (2010). *Nat. Prod. Rep.* **27**, 695–710.
Berman, H. *et al.* (2002). *Acta Cryst.* D**58**, 899–907.
Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bertini, I., Cavallaro, G. & McGreevy, K. S. (2010). *Coord. Chem. Rev.* **254**, 506–524.
Bleackley, M. R. & Macgillivray, R. T. (2011). *Biometals*, **24**, 785–809.
Cheung, K.-C., Strange, R. W. & Hasnain, S. S. (2000). *Acta Cryst.* D**56**, 697–704.
Cobine, P. A., Pierrel, F., Leary, S. C., Sasarman, F., Horng, Y. C., Shoubridge, E. A. & Winge, D. R. (2006). *J. Biol. Chem.* **281**, 12270–12276.
Davis, A. V. & O'Halloran, T. V. (2008). *Nature Chem. Biol.* **4**, 148–151.
Gamelin, D. R., Randall, D. W., Hay, M. T., Houser, R. P., Mulder, T. C., Canters, G. W., De Vries, S., Tolman, W. B. & Solomon, E. I. (1998). *J. Am. Chem. Soc.* **120**, 5246–5263.

Garner, D. K., Vaughan, M. D., Hwang, H. J., Savelieff, M. G., Berry, S. M., Honek, J. F. & Lu, Y. (2006). *J. Am. Chem. Soc.* **128**, 15608–15617.
Gorelsky, S. I., Xie, X., Chen, Y., Fee, J. A. & Solomon, E. I. (2006). *J. Am. Chem. Soc.* **128**, 16452–16453.
Gray, H. B., Malmstrom, B. G. & Williams, R. J. (2000). *J. Biol. Inorg. Chem.* **5**, 551–559.
Gray, H. B. & Winkler, J. R. (1996). *Annu. Rev. Biochem.* **65**, 537–561.
Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.
Harding, M. M. (2002). *Acta Cryst.* D**58**, 872–874.
Harding, M. M. (2006). *Acta Cryst.* D**62**, 678–682.
Kosman, D. J. (2010). *J. Biol. Inorg. Chem.* **15**, 15–28.
Kroneck, P. M., Antholine, W. E., Kastrau, D. H., Buse, G., Steffens, G. C. & Zumft, W. G. (1990). *FEBS Lett.* **268**, 274–276.
Kroneck, P. M., Antholine, W. A., Riester, J. & Zumft, W. G. (1988). *FEBS Lett.* **242**, 70–74.
Ledesma, G. N., Murgida, D. H., Ly, H. K., Wackerbarth, H., Ulstrup, J., Costa-Filho, A. J. & Vila, A. J. (2007). *J. Am. Chem. Soc.* **129**, 11884–11885.
Malmström, B. G. (1994). *Eur. J. Biochem.* **223**, 711–718.
Maret, W. (2010). *Metallomics*, **2**, 117–125.
Marshall, N. M., Garner, D. K., Wilson, T. D., Gao, Y.-G., Robinson, H., Nilges, M. J. & Lu, Y. (2009). *Nature (London)*, **462**, 113–116.
Messerschmidt, A., Huber, R., Poulos, T. & Wieghardt, K. (2001). *Handbook of Metalloproteins.* Chichester: Wiley.
Paraskevopoulos, K., Sundararajan, M., Surendran, R., Hough, M. A., Eady, R. R., Hillier, I. H. & Hasnain, S. S. (2006). *Dalton Trans.*, pp. 3067–3076.
Rubino, J. T. & Franz, K. J. (2012). *J. Biol. Inorg. Chem.* **107**, 129–143.
Solomon, E. I. & Hadt, R. G. (2011). *Coord. Chem. Rev.* **255**, 774–789.
Solomon, E. I., Szilagyi, R. K., DeBeer George, S. & Basumallick, L. (2004). *Chem. Rev.* **104**, 419–458.
Tamames, B., Sousa, S. F., Tamames, J., Fernandes, P. A. & Ramos, M. J. (2007). *Proteins*, **69**, 466–475.
Vila, A. J. & Fernandez, C. O. (2001). *Handbook of Metalloproteins*, edited by I. Bertini, A. Sigel & H. Sigel. New York: Marcel Dekker.
Williams, R. J. (1995). *Eur. J. Biochem.* **234**, 363–381.
Xie, X., Gorelsky, S. I., Sarangi, R., Garner, D. K., Hwang, H. J., Hodgson, K. O., Hedman, B., Lu, Y. & Solomon, E. I. (2008). *J. Am. Chem. Soc.* **130**, 5194–5205.
Xue, Y., Davis, A. V., Balakrishnan, G., Stasser, J. P., Staehlin, B. M., Focia, P., Spiro, T. G., Penner-Hahn, J. E. & O'Halloran, T. V. (2008). *Nature Chem. Biol.* **4**, 107–109.
Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L. & Minor, W. (2008). *J. Inorg. Biochem.* **102**, 1765–1776.