Journal of
Synchrotron Radiation

ISSN 0909-0495

Received 31 August 2009
Accepted 18 November 2009

Synergic approach to XAFS analysis for the identification of most probable binding motifs for mononuclear zinc sites in metalloproteins

Lisa Giachini,[a]* Giulia Veronesi,[b,c] Francesco Francia,[d] Giovanni Venturoli[d,e] and Federico Boscherini[b,c]

[a]Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia, [b]Dipartimento di Fisica, Università di Bologna, viale C. Berti Pichat 6/2, 40127 Bologna, Italy, [c]OGG-INFM-CNR, c/o ESRF, BP 220, F-38043 Grenoble, France, [d]Laboratorio di Biochimica e Biofisica, Dipartimento di Biologia, Università di Bologna, Via Irnerio 42, 40126 Bologna, Italy, and [e]CNISM, c/o Dipartimento di Fisica, Università di Bologna, viale C. Berti Pichat 6/2, 40127 Bologna, Italy. E-mail: lisa.giachini@synchrotron.org.au

In the present work a data analysis approach, based on XAFS data, is proposed for the identification of most probable binding motifs of unknown mononuclear zinc sites in metalloproteins. This approach combines multiple-scattering EXAFS analysis performed within the rigid-body refinement scheme, non-muffin-tin *ab initio* XANES simulations, average structural information on amino acids and metal binding clusters provided by the Protein Data Bank, and Debye–Waller factor calculations based on density functional theory. The efficiency of the method is tested by using three reference zinc proteins for which the local structure around the metal is already known from protein crystallography. To show the applicability of the present analysis to structures not deposited in the Protein Data Bank, the XAFS spectra of six mononuclear zinc binding sites present in diverse membrane proteins, for which we have previously proposed the coordinating amino acids by applying a similar approach, is also reported. By comparing the Zn *K*-edge XAFS features exhibited by these proteins with those pertaining to the reference structures, key spectral characteristics, related to specific binding motifs, are observed. These case studies exemplify the combined data analysis proposed and further support its validity.

Keywords: metalloproteins; XAFS; zinc binding sites; binding motifs; cytochrome *c* oxidase; thermolysin; superoxide dismutase.

1. Introduction

Metalloproteins have been receiving great interest in the scientific community since they constitute a significant proportion of all known genomes. Particular attention is devoted to the metal sites, which are often responsible for protein function. Detailed structural data for the metal sites in a metalloprotein are essential in order to fully understand the structure–function relationship that makes possible the performance of life-sustaining processes. X-ray absorption fine structure (XAFS) is an ideal tool for selectively probing the local structure of a metal ion in a protein since it can be applied to non-crystalline samples and can provide a high accuracy of the determined interatomic distances (Hasnain & Hodgson, 1999; Hasnain & Strange, 2003; Hasnain, 2004; Strange *et al.*, 2005). Significant advances in XAFS analysis have taken place in the last 20 years which have allowed the determination of reliable structural information for the metal site. Specifically we recall the use of constrained (or rigid-body; RB) and restrained refinement (Binsted *et al.*, 1992) and the combination of crystallographic and EXAFS information to extract three-dimensional information from the XAFS spectrum (Cheung *et al.*, 2000). Despite the progresses made in the precise and accurate refinement of the local structure, the identification of unknown metal binding sites, by means of XAFS data, is still a challenge when pre-existing structural information is not available. In fact, XAFS does not provide an absolute determination of the structure as X-ray diffraction does. A selection of suitable structural starting models is required, which implies that *a priori* structural information should be known. Moreover, even in cases in which such information might exist, the presence of multiple solutions [owing to the large number of structural parameters and Debye–Waller (DW) factors in the fitting model] can make the

selection of the correct local structure extremely difficult (Dimakis & Bunker, 2004, 2006).

In this paper we will show that a great improvement in this field, for mononuclear zinc binding sites, can be made by combining multiple-scattering EXAFS analysis performed within the RB refinement scheme, non-muffin-tin *ab initio* XANES simulations (Joly, 2001), structural information on amino acids and metal binding clusters provided by the Protein Data Bank (Harding, 2004), and DW factor calculations based on density functional theory (DFT) (Dimakis & Bunker, 2004). This combined approach has been recently applied (or partially applied) by the authors to investigate the Zn^{2+} binding sites in charge translocating membrane protein complexes (Giachini, Francia, Veronesi *et al.*, 2007; Francia *et al.*, 2007; Giachini, Francia, Boscherini *et al.*, 2007; Veronesi *et al.*, 2010). In all these cases the binding motif for the zinc ion was unknown. Limited preliminary structural information was available only in certain cases. By applying this combined data analysis procedure we identified the most probable binding motifs. However, since no complete structural information about the binding site existed for those proteins independent of that derived from our analysis, we could not have direct proof of the validity of our approach.

In this work further efforts have been made in order to organize this analysis approach in a coherent manner, as a general method for identifying unknown binding motifs for mononuclear zinc sites. In order to test the validity of the method, experimental data for three reference zinc proteins, for which the local structure around the metal is already known from protein crystallography, have been recorded and analysed on the basis of such an approach. As test proteins we have selected proteins which exhibit three binding motifs among the most common for mononuclear zinc metalloproteins, in the Protein Data Bank (PDB). This allows a direct comparison between the binding motif(s) selected using the present method and those solved by protein crystallography.

2. Experimental

2.1. Sample preparation

Bovine heart cytochrome *c* oxidase (COX), containing 10 nmoles of heme $a+a_3$ mg^{-1} protein, was purified as described by Errede *et al.* (1978). Thermolysin (TLS) and CuZn superoxide dismutase from bovine eritrocytes (SOD) were purchased from Calbiochem and Sigma-Aldrich, respectively. TLS was re-crystallized as described by Matsubara (1970). Measurements were performed on polyvinyl alcohol (PVA) protein films, prepared by adding 350 µl of a 10% solution of PVA (Fluka) to 1 ml of 80 µ*M* COX or to 1 ml of TLS suspension (15 mg ml^{-1} protein) or SOD suspension (3 mg ml^{-1} protein), respectively. After mixing, the protein-PVA solutions were layered into 3 cm × 3 cm × 0.3 cm Teflon holders and dried under nitrogen flow until PVA films were formed.

2.2. Data collection

Zn *K*-edge XAFS measurements were performed at the BM8 GILDA beamline of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, using a Si(311) double-crystal monochromator employing dynamical sagittal focusing (Pascarelli *et al.*, 1996). The photon flux was of the order of 10^{10} photons s^{-1} and the spot size was approximately 1 mm × 1 mm. Data were collected at room temperature using a 13-element hyper-pure Ge detector equipped with fast digital electronics with a peaking time equal to 1 µs (Ciatto *et al.*, 2004). The analyzed spectrum for COX was obtained from a scan with an integration time of 15 s per point with a maximum number of counts per channel of 6×10^3. The analyzed spectrum for TLS was obtained from the average of two scans for a total integration time of 30 s per point. The maximum number of counts per channel was 3×10^4. For SOD the analysed spectrum was obtained from the average of three scans for a total integration time of 45 s per point. The maximum number of counts per channel was 6×10^3.

3. Conceptual scheme of the data analysis procedure

When applying our analysis procedure we make use of a database of target models that we have built for mononuclear zinc metalloproteins. For target models we mean a number of binding motifs selected according to the PDB in which the local structure is fixed on the basis of average structural parameters based partly on the Cambridge Structural Database (CSD) and partly on the PDB. A file containing the atomic coordinates necessary for XAFS calculations is associated with each target model and added to the database. Similarly, for each target model we associate and add to the database text files containing the fitting parameters and their associated mathematical relationships necessary to apply the RB refinement and to use the DW factor parameterization, proposed by Dimakis & Bunker (2004) according to DFT (see §3.1 for details). A systematic inspection of the XAFS features of the simulated spectra of the target sites led to establish (i) the relationship between the white line intensity and the coordination number, (ii) the possibility of distinguishing between different first neighbours through a first-shell analysis and (iii) the relationship between the number of histidine (His) residues and their contribution in the region R = 3–4 Å in the Fourier transform. These quantitative criteria are extremely useful in order to decrease the number of possible starting models, which are generated and described in §3.3.

With the use of this database, one is then able to identify the most probable metal binding site from the raw data following a simple and straightforward procedure. As illustrated in Fig. 1, the first step is a preliminary XAFS analysis of the raw data to select a limited number of starting models among those present in the database. Each selected model has its corresponding files in the database: the text file of the atomic coordinates can be imported into a code able to calculate the theoretical amplitude and phase-shift functions [we have used *FEFF8* (Ankudinov *et al.*, 1998)]; the text file for RB refine-

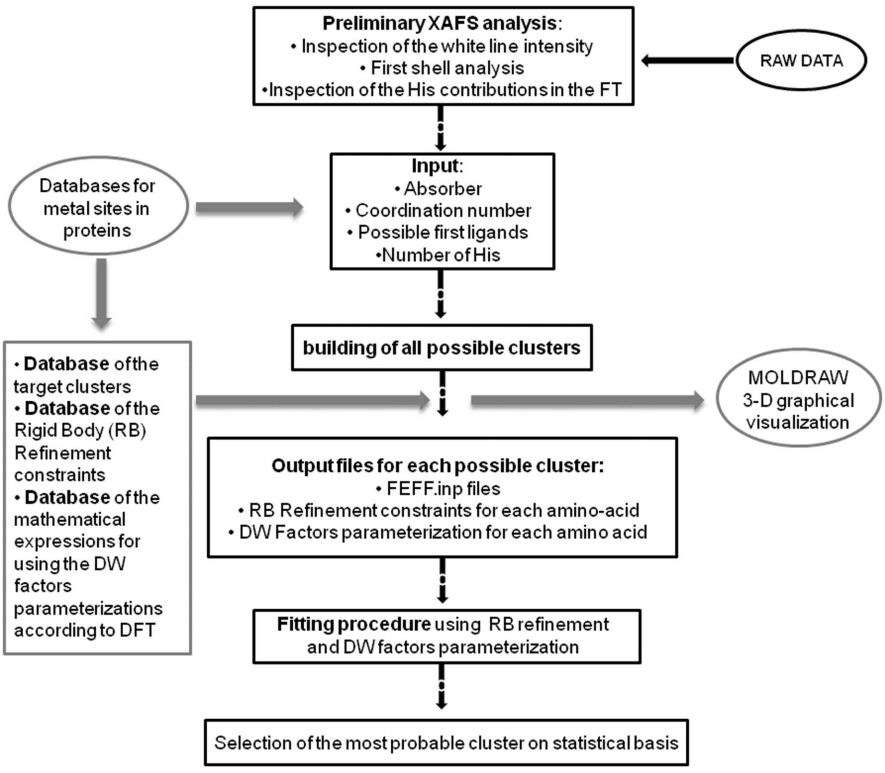

Figure 1
Block diagram of the data analysis procedure. Among a database of target models a limited number of sites are selected according to a preliminary XAFS analysis. The *ab initio* simulations of the putative clusters are fitted to the experimental data using RB refinement and parameterizing the DW factors according to DFT estimates. The identification of the binding site is done on the basis of a statistical analysis. See text for details.

ment parameters and DW factors parameterization can be imported into a code able to perform a least-squares refinement. We have used *FEFFIT*, as implemented in the *Artemis* package (Ravel & Newville, 2005) (see §3.4). The fit is then performed directly in k-space with a k-weight of 3, and the identification of the binding site is done on the basis of a statistical analysis by evaluating the reduced χ^2 (see §3.5).

3.1. Databases

To build the target clusters for a selected metal (in this case zinc) we need the following information: a list of all possible metal–ligand patterns observed in the PDB; the internal structural parameters for all amino acids; a set of target distances between the metal and the coordinating amino acids (see Table 2); and target geometries of the amino acids around the absorber. To obtain the list of all possible metal–ligand patterns observed in the PDB we use the statistical analysis tools of the Metalloprotein Database and Browser (MDB; http://metallo.scripps.edu/) (Castagnetto *et al.*, 2002) which contains quantitative information on all the metal-containing sites available from structures in the PDB distribution. Once the list of possible patterns has been defined, the coordinates of the target clusters are obtained by fixing the internal structural parameters of the amino acids to the values reported by Engh & Huber (1991), whereas the target distance from the amino acid to the metal and the geometry of the cluster are fixed according to the University of Edinburgh website which provides information about the geometry and constitution of metal coordination groups in metalloproteins (http://tanna.bch.ed.ac.uk/) (Harding, 1999, 2000, 2002, 2004, 2006; Hsin *et al.*, 2008). Clusters are built using *MOLDRAW* (Ugliengo *et al.*, 1993), which allows importing and exporting files in *FEFF* (Ankudinov *et al.*, 1998) format. Therefore, from the database of target clusters, a database of *FEFF* files, one for each target cluster, can be easily generated. Following these indications it is possible to easily build databases for all the metals of interest for metalloproteins. Up to now we have built the databases for mononuclear zinc metalloproteins.

In the case of zinc metalloproteins, tetrahedral clusters are by far the commonest Zn sites. However, as indicated at the University of Edinburgh website, other less common coordination numbers are 3, 5 and 6. The ideal stereochemistry for a coordination number of 6 is octahedral, for 5 it is trigonal bipyramidal or square pyramidal, and for 4 it is tetrahedral. The possible ligand patterns with more than 15 PDB entries for each coordination number, according to the MDB (Castagnetto *et al.*, 2002), are listed in Table 1. The target distances of the amino acids and of the water molecule from zinc according to the University of Edinburgh website are listed in Table 2. For each ligand pattern a target cluster, a *FEFF* file and a text file with constraints for RB refinement and DW factors have been built.

3.2. Preliminary XAFS analysis

3.2.1. Coordination number. The correlation between the white-line (WL) intensity and the coordination number is very well documented for metalloproteins for a wide variety of absorbing atoms (including zinc) (Feiters *et al.*, 2003; Mijovilovich & Meyer-Klaucke, 2003; Peariso *et al.*, 2003; Dau *et al.*, 2005; Banci *et al.*, 2005). In particular, it has been observed that the WL intensity increases with the coordination number. The existence of such a correlation can be understood qualitatively in the framework of molecular orbital theory. In fact, the density of final states owing to unoccupied molecular orbitals is expected to increase, for similar ligands with similar bond lengths, with the numbers of neighbours; since the absorption cross section is directly proportional to the density of final states, the observed correlation is reasonable.

Table 1
Ligand patterns for zinc metalloproteins with more than 15 PDB entries according to the MDB.

For the coordination number 6, all the ligand patterns have less than 15 PDB entries in the MDB.

Coordination number	Binding motifs	Counts in MDB
3	His His His	39
	Asp His His	19
	Cys Cys His	17
	His H_2O H_2O	17
4	Cys Cys Cys Cys	484
	Asp His His His	187
	Cys Cys Cys His	183
	Cys Cys His His	157
	His His His H_2O	74
	Cys Cys His H_2O	56
	Glu His His H_2O	27
	Cys Cys Cys H_2O	22
	Asp Asp His Ser	21
	Asp His H_2O H_2O	20
	Asp Asp His His	16
5	Glu Glu His His H_2O	35
	Asp Asp His His His	19

We therefore expected to find such correlation by performing theoretical simulations of the X-ray absorption near-edge structure (XANES) of the target clusters for mononuclear zinc binding sites. In fact, we did observe this correlation for all the XANES simulations that we have performed, which include all the binding motifs listed in Table 1 and additional less common binding sites containing a mixture of sulfur and nitrogen/oxygen atoms in the first coordination shell (see Figs. 2 and 3 of the supplementary material[1]). By comparing the WL intensity for clusters with different coordination numbers, a quantitative criterion can be established, as described in the following. Normalizing to unity the XANES spectrum sufficiently far from the edge (specifically at 80 eV after the edge), the WL intensity is <1.5 for a coordination number of 3 and 4 and is >1.6 for coordination numbers of 5 and 6. We would like to stress that such a criterion is confirmed not only by all the theoretical simulations that we have performed but also by all the experimental data shown in this work, and compares well with other previously published XAFS data of mononuclear zinc binding sites (Feiters *et al.*, 2003). Therefore, in the specific case of mononuclear zinc metalloproteins, by using this simple criterion it is possible to rule out some coordination numbers for the unknown zinc binding sites simply by measuring the WL intensity of the raw data. In Fig. 2 we show the XANES simulations for a set of Zn clusters containing the same ligands (His, carboxylic acids and water molecules), but characterized by different coordination numbers and coordinating geometry. The difference in the WL intensity between the coordination numbers 3, 4 and the coordination numbers 5, 6 is evident: in particular, in the cases reported in the figure, we obtain a value of 1.5 for $N = 3$ or 4, 2.0 for $N = 5$ and 2.1 for $N = 6$.

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: HI5605). Services for accessing these data are described at the back of the journal.

Table 2
Metal–ligand target distances for Zn according to the University of Edinburgh website.

For serine expect that the Zn—O distance is between that for water and carboxylate.

$Zn-O_{H_2O}$	$Zn-O_{Asp/Glu}$	$Zn-N_{His}$	$Zn-S_{Cys}$
2.09 Å	1.99 Å	2.03 Å	2.31 Å

XANES simulations have been performed making use of the non-muffin-tin finite difference method, as implemented in the *FDMNES* program (Joly, 2001). A much better agreement with experimental data has been obtained by using this approach, instead of a muffin-tin-based approach (Joly, 2003). All simulated spectra were convoluted with a Lorentzian function and then normalized at their value corresponding to the energy of 80 eV after the edge, in order to be compared.

3.2.2. First neighbours. Amino acids can bind metals with N, O or S. Since N/O and S have different scattering amplitudes, first-shell analysis can discriminate between the presence of N/O or S atoms. However, because Zn–N and Zn–S XAFS oscillations are almost out of phase, the relative number of the different ligands cannot be reliably determined, as already observed by Clark-Baldwin *et al.* (1998). The difficulty is greater when the data have a poor signal-to-noise ratio and thus the possible fitting range is short. Therefore we recommend that during the preliminary analysis a quick first-shell analysis is used only to rule out, eventually, the presence of S or N/O (and thus the related amino acids). Knowledge of the relative number of the different ligands would, of course, further decrease the possible starting models but it will also introduce to some extent the risk of ruling out the good model at the initial stage of the procedure.

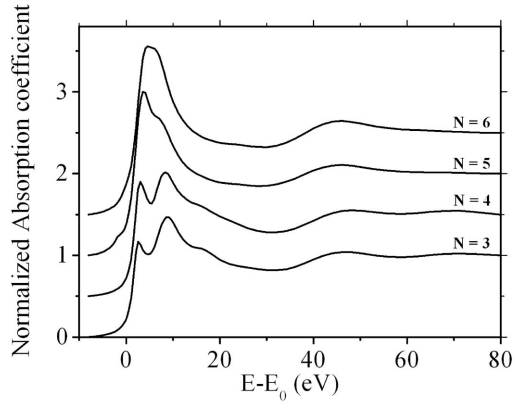

Figure 2
XANES simulations for a set of Zn clusters containing the same ligands (His, carboxylic acids and water molecules), but characterized by different coordination numbers and coordinating geometry. In particular, the simulated clusters are formed by 3 His for $N = 3$; 3 His 1 Asp for $N = 4$; 2 His 1 H_2O 1 Glu-bidentate for $N = 5$; 2 His 2 H_2O 1 Asp-bidentate for $N = 6$. The numerical values obtained for the WL intensity are 1.5 for $N = 3$ or 4, 2.0 for $N = 5$ and 2.1 for $N = 6$.

3.2.3. Number of histidines. It is well known that His residue, owing to its ring structure, generates multiple scattering contributions of significant amplitude that contribute at high distances (>3 Å) in the Fourier transform (FT) (Co *et al.*, 1981; Strange *et al.*, 1987). On the contrary, contributions in this region coming from other amino acids are very weak. Therefore the FT of the experimental spectrum contains important information about the presence of His residues present in the binding cluster from which the XAFS signal is generated. For mononuclear zinc binding sites, we have investigated such a relationship in a systematic way by performing many theoretical simulations based on the target clusters contained in the database and by comparing these results with experimental data. The DW factors of the theoretical simulations were kept fixed at the values based on DFT calculations. The range used for the FT was 2–12 Å^{-1}. From these simulations and experimental data it appears that, for mononuclear zinc binding sites, the region in the FT which contains, almost exclusively, contributions coming from His residues is included between 3 and 4 Å (see Fig. 3). Moreover we observe a systematic increase of the quantity

$$I = \int_3^4 \left|\mathrm{FT}\left[k^3\chi(k)\right]\right| \mathrm{d}k \qquad (1)$$

with the number of His present in the cluster. In particular, in the case of $N = 3, 4$ (at a temperature of 300 K) we have found that the quantity I can be related to the number of His in the following way: if only one His residue is present, $0.4 \leq I < 0.8$; for two His, $0.8 \leq I < 1.2$; for three His, $1.2 \leq I < 1.6$. In the case of $N = 5, 6$ (at a temperature of 300 K) we have found that the quantity I can be related to the number of His in the following way: if one is present, $0.5 \leq I < 1.0$; for two His, $1.0 \leq I < 1.5$; for three His, $1.5 \leq I < 2.0$.

In Table 3 we report the values for the integral I obtained from the XAFS data collected for a number of tetrahedral mononuclear zinc sites. These experimental values compare well with those obtained from theoretical calculations and thus satisfy the quantitative criterion stated above. In Fig. 3 we show, as an example, the FT of the experimental data collected at room temperature for TLS (for which it is known from protein crystallography that the zinc site contains two His) and for SOD (for which it is known from protein crystallography that the zinc site contains three His). Together with these experimental data we show in Fig. 3 as a shadowed area the sum of all His contributions, localized in the region 3–4 Å in *R*-space, for the theoretical calculations based on the target clusters of the database corresponding to the binding motif His His Glu H_2O and His His His Asp in the case of TLS and SOD, respectively. We note only minor differences (in the region 3–3.5 Å) between the experimental data and the theoretical calculations of the contributions coming from His residues. The origin of these slight differences is most likely due to the presence of low-amplitude contributions from other residues (*i.e.* carboxylic acids) or to subtle structural discrepancies between the real cluster and the target cluster of the database.

Table 3
Experimental values for the integral I in selected tetrahedral mononuclear zinc binding sites.

Protein	Number of His	I
TLS (this work)	2	1.1
SOD (this work)	3	1.5
Reaction centre (Giachini *et al.*, 2005)	2	0.9
Complex I (Giachini, Francia, Boscherini *et al.*, 2007)	2	0.8
Avian cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007)	2	0.9
Bovine cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007)	2	0.9

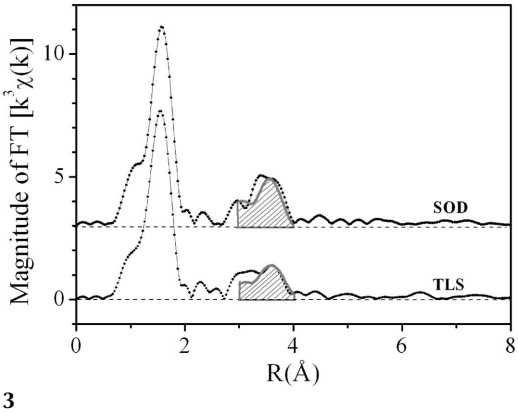

Figure 3
Amplitude of the Fourier transform of the experimental data collected at room temperature for TLS (2 His) and SOD (3 His). Together with these experimental data we show as shadowed areas the sum of all the contributions, coming from the His residues in the theoretical calculations, and localized in the region 3–4 Å in *R*-space. The theoretical calculations were based on the target clusters of the database corresponding to the binding motif His His Glu H_2O and His His His Asp, for TLS and SOD, respectively. The numerical values obtained for the integral I are reported in Table 3.

3.3. Generation of possible clusters

On the basis of the information provided by preliminary XAFS analysis (*i.e.* coordination number, possible ligands and number of His), we select those models in the list provided by the MDB that exhibit these characteristics. A possible limitation of this approach is that binding motifs which have never been observed by protein crystallography might not be identified. However, we would like to point out that this is not a conceptual limitation of the algorithm and that, in principle, it is possible to take into consideration also models originating from all permutations of the possible amino acids in groups compatible with the selected coordination number(s), and not included in the MDB list. For the new binding motifs the target model can be easily built in *MOLDRAW*, as described previously.

Once the possible starting binding motifs have been identified, the corresponding files (containing atomic coordinates, RB refinement constraints and parameters for the DW factors expressions) can be picked up from the database. Recent improvements in *Artemis* features (Ravel & Newville, 2005) allow to directly import not only *FEFF* theoretical calculations but also text files containing the parameters of the fitting models (set, guessed and defined) (see specifically the *Artemis*

manual, at http://cars9.uchicago.edu/~ravel/software/doc/Artemis/artemis.pdf).

3.4. EXAFS simulation and fitting procedure

Theoretical amplitude and phase shift functions are calculated using the *ab initio* code *FEFF8.2* by selecting the partially non-local form for the exchange potentials and including *FEFF*'s automated self-consistent potential calculations (Ankudinov *et al.*, 1998). The value of $S_0{}^2$ is calculated by *FEFF8.2* from atomic overlap integrals of each different cluster taken into account and is kept fixed during the analysis. This value changes only slightly among the binding sites included in the database, varying in the range 0.93–0.96. The data are analyzed using the *FEFFIT* program (Newville *et al.*, 1995) using as minimization algorithm a modified Levenberg–Marquardt method. The fits are performed directly in k space, with a k weight of 3, minimizing the R-factor, defined as (Stern *et al.*, 1995)

$$R = \sum_{i=1,N} \left(k_i^3 \tilde{\chi}_{i_{\rm data}} - k_i^3 \tilde{\chi}_{i_{\rm fit}}\right)^2 / \sum_{i=1,N} \left(k_i^3 \tilde{\chi}_{i_{\rm data}}\right)^2. \quad (2)$$

All of the multiple scattering signals constituted by up to five scattering processes involving atoms belonging to the same residue and with an effective length ≤ 5 Å are taken into account. The free parameters used in the fitting procedure are (i) a common shift in the energy origin E_0; (ii) the variation of the first-ligand distance, and (iii) an angular parameter for each amino acid bound to the metal. This angular parameter accounts for rotation of the residue around an axis through the atom bound to the metal and perpendicular to the plane formed by the metal, the first and the second neighbour (see Fig. 6). DW factors are parameterized, according to Dimakis & Bunker (2004), with analytical expressions of the form

$$\sigma_i^2(\Delta R, T) = \sigma_i^2(R_{0i}, T) + A_i(T)\Delta R_i + B_i(T)\Delta R_i^2, \quad (3)$$

where T is the absolute temperature, R_{0i} is the equilibrium distance for the atom i found out by DFT calculations, ΔR_i is the variation between the equilibrium and XAFS calculated distance for atom i, and $\sigma i^2(R_{0i}, T)$, $A_i(T)$, $B_i(T)$ are third-order polynomial functions derived from calculation of the DFT phonon normal modes calculations. Specifically, this means that the previous expressions are inserted into *Artemis* as mathematical relations so that the DW factors are adjusted during the fitting procedure according to the XAFS calculated distances.

The reliability of the fitting procedure is assured by a high determinacy of the system, which is described as the ratio between the number $N_{\rm ind}$ of independent points in the XAFS data set and the number p of fitted parameters included in the model ($N_{\rm ind}/p$) (Levina *et al.*, 2005). The number of independent points has been calculated using (Stern *et al.*, 1995)

$$N_{\rm ind} = \frac{2(k_{\rm max} - k_{\rm min})(R_{\rm max} - R_{\rm min})}{\pi} + 2, \quad (4)$$

where $k_{\rm max}$, $k_{\rm min}$ and $R_{\rm max}$, $R_{\rm min}$ define the intervals in the reciprocal and real space in which the analysis was performed.

3.5. Selection of the most probable cluster

The identification of the binding site is done on a statistical basis. Confidence analysis is performed on the basis of the reduced χ^2, *i.e.* χ^2_ν, defined as (Stern *et al.*, 1995)

$$\chi^2_\nu = \frac{1}{\nu}\frac{N_{\rm ind}}{N} \sum_{i=1,N} \left(\frac{\tilde{\chi}_{i_{\rm data}} - \tilde{\chi}_{i_{\rm fit}}}{\sigma}\right)^2, \quad (5)$$

where $\nu = N_{\rm ind} - p$ is the number of the degrees of freedom in the fit. For each data set a single value for the variance σ of $\tilde{\chi}_{i_{\rm data}}$ is calculated using Poisson statistics, by calculating the square root of the total number of counts. Even in the presence of 'good' fits, *i.e.* at relatively low values of the R-factor, the values calculated for χ^2_ν are usually much larger than 1. This situation is commonly encountered in XAFS analysis and attributed to small inadequacies of the model and/or to systematic experimental errors [see Kelly *et al.* (2001) and references therein]. In view of this, the standard fluctuation in χ^2_ν [which is equal to $(2/\nu)^{1/2}$] is used to rescale to $(2/\nu)\chi^2_\nu$ (Stern *et al.*, 1995; Kelly *et al.*, 2001). The comparison between two different fits of the same data set (corresponding to two different clusters, a and b) is performed by means of χ^2_ν according to the following criterion (Kelly *et al.*, 2001): fit to cluster b is considered significantly better than fit to cluster a when

$$\left[\chi^2_\nu(a) - \chi^2_\nu(b)\right] \geq \left(2\left\{\frac{\left[\chi^2_\nu(a)\right]^2}{\nu(a)} + \frac{\left[\chi^2_\nu(b)\right]^2}{\nu(b)}\right\}\right)^{1/2}, \quad (6)$$

which corresponds to the confidence level of σ.

4. Testing the method

To test the efficacy of our method we have selected three zinc metalloproteins containing a characteristic zinc binding site known from protein crystallography. In particular we have selected (i) the Zn structural site of bovine heart COX; (ii) the Zn catalytic site of TLS and (iii) the Zn catalytic site of SOD.

COX is the terminal component of the respiratory chain: it catalyses the oxidation of cyt *c* reduced by the cyt bc_1 complex, reducing O_2 to H_2O and pumping four protons across the mitochondrial membrane. There are a wide number of crystallographic structures available for bovine heart COX in different states (fully oxidized, fully reduced, azide-bound and carbon-monoxide-bound) with resolution up to 1.8 Å (see, for example, Muramoto *et al.*, 2007; Shinzawa-Itoh *et al.*, 2007; Aoyama *et al.*, 2009). All the available crystallographic structures show the existence of an endogenous Zn^{2+} bound to subunit V_b. The local structure around the zinc ion, reported in all the crystallographic structures, is formed by four cysteines. This binding motif is the most common in nature.

The XAFS spectrum for purified bovine heart COX exhibits a WL intensity of 1.3 (see Fig. 4). According to the quantitative criterion stated in §3.2.1, this indicates a coordination number of 3 or 4. First-shell analysis clearly indicates the presence of S atoms whereas no nitrogen and/or oxygen atoms have been detected. We thus have this preliminary informa-

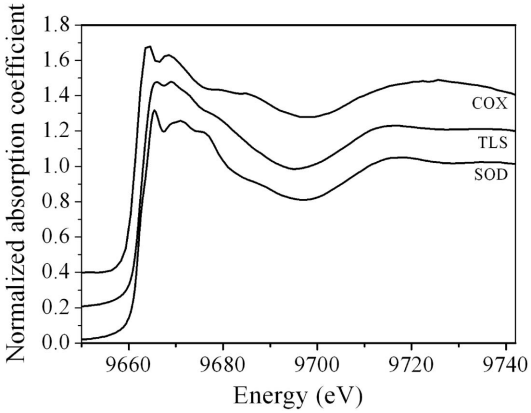

Figure 4
Experimental spectrum for COX, TLS and SOD.

tion: absorber = Zn; coordination number = 3, 4; coordinating atoms whose presence is sure = S; no His. When considering the most probable binding motifs selected in Table 1, the only possibility is the cluster formed by four cysteines. When including less common or occasionally found binding motifs (more than five PDB entries in the MDB), the list includes additionally a binding motif formed by three cysteines. We therefore proceed to apply the next step of our method, which is the comparison of the two reduced χ^2 obtained from the fitting procedure on the basis of the criterion stated in §3.5 [see equation (6)]. The fitting procedure is performed within the RB refinement scheme (as described in §3.4), moving each amino acid rigidly around the absorber with two degrees of freedom (specifically the $Zn-S_\gamma$ distance and the $Zn-S_\gamma-C_\beta$ angle, as shown in Fig. 6). As starting models we use the target clusters corresponding to the selected binding motifs. When the same amino acid occurs more than once, as is the case here, it is treated in a non-degenerate mode. This is done in order to avoid the case that a possible spread in first-shell distances may interfere in an incorrect way in the evaluation of the best-fitting cluster. DW factors are parameterized on the basis of DFT calculations, as described in §3.4 [see equation (3)]. The amplitude reduction factor $S_0{}^2$ is kept fixed at the value calculated by *FEFF8.2*, as stated previously. For the models taken into consideration for COX, TLS and SOD, such a value is 0.95. The fitting range used was 2.3–12.3 Å^{-1} as indicated by the dashed line in Fig. 5 where we show the EXAFS of the best-fitting cluster (continuous line) compared with the experimental data (open circles). The *R*-factor and reduced χ^2 that we obtained for the two binding motifs are reported in Table 4. According to the criterion stated in §3.5, the model Cys Cys Cys Cys provides a fit which is significantly better than that of the model Cys Cys Cys within the confidence interval of 1σ. The structural parameters obtained are reported in Table 5. The value provided by the fitting procedure for the common shift in the energy origin is −2 (2) eV (the starting value for the energy origin was set to 9661 eV, as determined by the first inflection point of the raw data).

We notice a certain spread in the first-shell distances ($Zn-S_\gamma$). To check if this spread is consistent with the crystallographic structure, we have compared our results with those

Table 4
Values obtained for the $R_{\rm factor}$, reduced χ^2 and its standard fluctuation for the two models selected for COX.

The 1σ-uncertainty in χ^2_ν is given in brackets. According to the criterion stated in §3.5, the model Cys Cys Cys Cys is statistically better than the model Cys Cys Cys within the confidence interval of 1σ.

Binding motif	$R_{\rm factor}$ (%)	χ^2_ν [$(2/\nu)^{1/2}\chi^2_\nu$]
Cys Cys Cys	6	87 [31]
Cys Cys Cys Cys	4	50 [19]

Table 5
Structural parameters provided by the fitting procedure for the zinc binding site of COX.

The 1σ error on the least significant figure is given in parentheses.

Ligand	$Zn-S_\gamma$ (cys) (Å)	α (°)
Cys1	2.30 (9)	104 (5)
Cys2	2.31 (8)	103 (10)
Cys3	2.32 (9)	109 (10)
Cys4	2.41 (2)	97 (10)

of the PDB file 2zxw which refers to the X-ray structure of the fully oxidized as-isolated bovine heart COX (Aoyama *et al.*, 2009) determined at 2.5 Å resolution. As mentioned above, several X-ray structural analyses have been reported for bovine heart COX. However, when the values of the $Zn-S_\gamma$ distances around Zn^{2+} are compared for the single structures, large discrepancies (up to 0.3 Å) are visible. This seems to depend not only on differences in resolution but especially on differences of the state of the protein (fully oxidized *versus* fully reduced, for example) (Aoyama *et al.*, 2009). Therefore, to make the comparison more meaningful we have chosen the case where the experimental conditions were more similar to ours, *i.e.* the X-ray structure of the fully oxidized as-isolated bovine heart COX, obtained by limiting the X-ray dose for each shot and by using many (∼400) single crystals. As shown in Table 6, the distances that result from our analysis compare well with those of the crystallographic analysis. In both cases

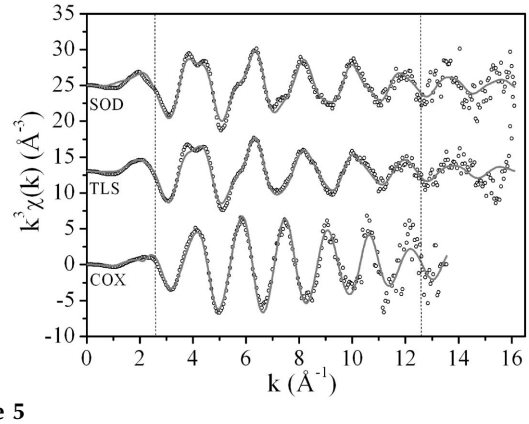

Figure 5
Experimental k^3-weighted EXAFS function (open circles) for COX, TLS and SOD. The continuous line is the fit obtained for a cluster formed by Cys Cys Cys Cys, by His His His Asp and by His His Glu H_2O in the case of COX, SOD and TLS, respectively. Dashed lines indicate the fitting range 2.3–12.3 Å^{-1}.

Table 6
First-shell distances as determined in the present work compared with those pertaining to the PDB file 2zxw (Aoyama *et al.*, 2009).

Ligand	Zn—S_γ (Å) target cluster	Zn—S_γ (Å) XAFS (this work)	Ligand	Zn—S_γ (Å) XRD PDB 2zxw
Cys1	2.31	2.30	Cys82	2.30
Cys2	2.31	2.31	Cys60	2.31
Cys3	2.31	2.32	Cys62	2.34
Cys4	2.31	2.41	Cys85	2.38
Average	2.31	2.33	Average	2.33

we observe a comparable spread in the first-shell distances, and the same average value of the four Zn—S_γ distances (2.33 Å). This is also the value obtained by performing a first-shell fit by using a single Zn—S_γ path with a degeneracy of four, that we have previously published (Francia *et al.*, 2007). To test the reliability and determinacy of our fitting procedure we have monitored the dependence of the structural results on starting values. When performing the fit with different starting values for the Zn—S_γ distances (*i.e.* target distances), we observe no changes in the structural parameters obtained from the fitting procedure.

TLS is a thermostable neutral metalloproteinase enzyme produced by the gram-positive bacterium *Bacillus thermoproteolyticus*. It contains a zinc site which catalyzes the hydrolysis of peptide bonds involving hydrophobic amino acids. Crystallographic analyses have allowed the identification of the amino acids that bind to the catalytic zinc ion. Specifically, these studies show that the zinc cluster is formed by two His, one glutamic acid (Glu) and one water molecule (H_2O) (Matthews *et al.*, 1972).

In the XAFS spectrum collected for TLS the WL intensity is again 1.3 (see Fig. 4), thus indicating a coordination number of 3, 4. First-shell analysis excludes the presence of S atoms. The coordinating atoms are N/O. The FT indicates the presence of two His according to the quantitative criterion illustrated above (§3.2.3), since we obtain a value of $I = 1.1$. We thus have this preliminary information: absorber = Zn; coordination number = 3, 4; possible coordinating atoms = N (His = 2), O. These criteria allow the selection of the following binding motifs, among the most common ones present in the MDB (listed in Table 1): His His Asp, His His Glu H_2O, His His Asp Asp. When considering less common or occasionally found binding motifs (more than five PDB entries in the MDB), the list includes additionally His His Glu, His His Asp H_2O, His His H_2O H_2O. From an XAFS point of view the carboxylic acids [*i.e.* aspartic acid (Asp) and Glu] are indistinguishable owing to the close similarity of their chains; in fact the only structural difference between the two residues is the presence of one more carbon atom in the chain of Glu, which is normally at a distance larger than 4.5 Å. Therefore the list reduces to the four patterns His His Asp/Glu, His His Glu/Asp H_2O, His His Asp Asp, His His H_2O H_2O. The fitting procedure has been performed as described previously for COX, and detailed in §3.4. The fitting range used was 2.3–12.3 $Å^{-1}$, as indicated by the dashed line in Fig. 4. The free parameters were the two individual Zn—$N_{\varepsilon 2}$ distances and the two angles Zn—$N_{\varepsilon 2}$—$C_{\varepsilon 1}$ of the two His residues, the Zn—$O_{\varepsilon 1/\delta 1}$ distance and the Zn—$O_{\varepsilon 1/\delta 1}$—$C_{\delta/\gamma}$ angle of the Glu/Asp residue, and the Zn—O distance of the water molecule (see Fig. 6).

Table 7
Values obtained for the $R_{\rm factor}$, reduced χ^2 and its standard fluctuation for the four models selected for TLS.

The 1σ-uncertainty in χ^2_ν is given in brackets. According to the criterion stated in §3.5, the model His His Glu/Asp H_2O is statistically better than the model His His Asp/Glu within the confidence interval of 2σ; the model His His Glu/Asp H_2O is significantly better than the model His His H_2O H_2O and the model His His Asp Asp within the confidence interval of 1σ.

Binding motif	$R_{\rm factor}$ (%)	χ^2_ν [$(2/\nu)^{1/2}\chi^2_\nu$]
His His Asp/Glu	10	119 [35]
His His Asp Asp	7	76 [23]
His His H_2O H_2O	5	69 [21]
His His Glu/Asp H_2O	4	35 (11)

The *R*-factor and reduced χ^2 obtained for the four binding motifs are reported in Table 7. When evaluating the reduced χ^2 for fits with different binding motifs we observe that the binding motif His His Glu/Asp H_2O is significantly better than the others. Specifically, the confidence interval of the statistical analysis is 2σ when this model is compared with that with coordination number of three (His His Asp/Glu), whereas the confidence interval decreases to 1σ when the same model is compared with the two binding motifs His His Asp Asp, His His H_2O H_2O, which have, indeed, a very similar local structure around the zinc ion. In Fig. 5 we show the EXAFS of the best-fitting cluster (continuous line) compared with the experimental data (open circles). The structural parameters provided by the fitting procedure are reported in Table 8. The best-fitting value for the common shift in the energy origin is −1 (1) eV (the starting value for the energy origin was set to 9662 eV, as determined by the first inflection point of the raw data).

SOD catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide. Crystallographic data have shown the

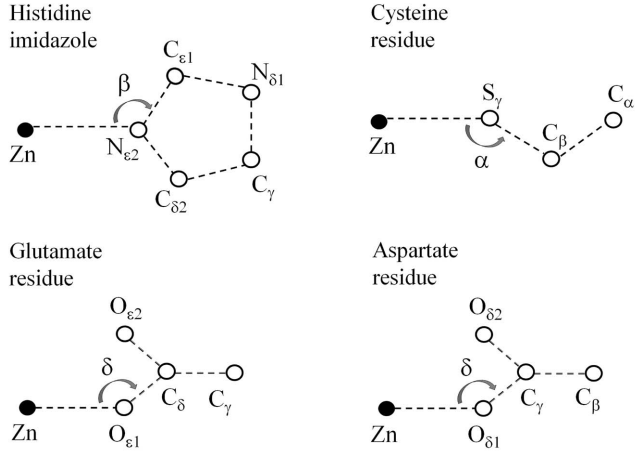

Figure 6
Sketches of the reference structural units used in the refinement.

Table 8
Structural parameters provided by the fitting procedure for the zinc binding site of TLS.

The 1σ error on the least significant figure is given in parentheses.

Ligand	Zn—Res (Å)	β, δ (°)
His1	1.97 (4)	128 (10)
His2	1.99 (5)	121 (10)
Glu†	1.95 (4)	124 (3)
H_2O‡	2.04 (2)	–

† The value found for Zn—$O_{\varepsilon 1}$ and δ (Zn—$O_{\varepsilon 1}$—C_δ) results in a Zn—$O_{\varepsilon 2}$ distance of 2.93 Å. ‡ The parameterization of the DW factors on the basis of DFT calculations is, at present, not available for water molecules. Consequently, this parameter was free to vary in the first stage of the fitting procedure. It converged to a value of 0.004 Å^2. In the final fit, from which the reduced χ^2 is calculated, this value was kept fixed at such a value.

existence of a catalytic zinc site formed by three His and one Asp (Trainer *et al.*, 1983; Hough & Hasnain, 1999). The XAFS spectrum collected for SOD exhibits a WL intensity of 1.3, thus indicating a coordination number of 3, 4. As for the case of TLS, first-shell analysis excludes the presence of S atoms. The coordinating atoms are N/O. The FT indicates the presence of three His according to the quantitative criterion illustrated above, since we obtain a value of $I = 1.5$. Therefore in this case we have the following preliminary information: absorber = Zn; coordination number = 3, 4; possible coordinating atoms = N (His = 3), O. These criteria lead to the following clusters, among those present in Table 1: His His His, His His His Asp, His His His H_2O. When enlarging the statistics to less common or occasionally found binding motifs (more than five PDB entries in the MDB), no new binding motifs are encountered. The fitting procedure for the selected starting models listed above was performed as described previously. The free parameters were the three individual Zn—N_ε distances and the three angles Zn—N_{e2}—$C_{\varepsilon 1}$ of the three His residues, the Zn—$O_{\delta 1}$ distance and the Zn—$O_{\delta 1}$—C_γ angle of the Asp residue. It is worthwhile noting that it is known from protein crystallography that the three histidines bind zinc in N_δ conformation in SOD. However, for simplicity, His residues are always built in N_ε configuration in the target clusters of our database. Therefore when considering such residues in the different binding motifs we always use the N_ε model. The extra carbon atom at approximately 3.5 Å which is present in the N_δ conformation can indeed contribute to change the reduced χ^2 slightly. However, these changes are normally smaller than those produced by two different binding motifs. We recall, moreover, that it is known from protein crystallography that one of the zinc-coordinating His (His61) is also linked to a Cu site. This can affect the DW factors. For the two aspects mentioned above, the case of SOD can be considered as an interesting case to prove the reliability of the method. The statistical results for the three models are shown in Table 9. The *R*-factor and reduced χ^2 obtained from the model His His His Asp are appreciably lower than those generated from the other two models, indicating that this binding motif is indeed the one which reproduces better the experimental spectra. When comparing the His His His Asp binding motif with that characterized by a coordination

Table 9
Values obtained for the $R_{\rm factor}$, reduced χ^2 and its standard fluctuation for the three models selected for SOD.

The 1σ-uncertainty in χ^2_ν is given in brackets. The model His His His Asp is significantly better than the model His His His H_2O within a confidence interval of 1σ. The model His His His Asp is significantly better than the model His His His within a confidence interval 2σ.

Binding motifs	$R_{\rm factor}$ (%)	χ^2_ν [$(2/\nu)^{1/2}\chi^2_\nu$]
His His His	10	138 [43]
His His His H_2O	7	97 [31]
His His His Asp	5	60 [18]

Table 10
Structural parameters provided by the fitting procedure for the zinc binding site of SOD.

The 1σ error on the least significant figure is given in brackets.

Ligand	Zn—Res (Å)	β, δ (°)
His1	1.98 (5)	120 (10)
His2	2.01 (5)	122 (10)
His3	2.01 (4)	126 (10)
Asp†	2.00 (5)	116 (9)

† The value found for Zn—$O_{\delta 1}$ and δ (Zn—$O_{\delta 1}$—C_γ) results in a Zn—$O_{\delta 2}$ distance of 2.74 Å.

number of three (His His His), we can state that the former is significantly better than the latter within a confidence interval of 2σ. When comparing the His His His Asp binding motif with the other (His His His H_2O), the confidence interval decreases to 1σ. The fact that, despite the approximations used in the target model, the binding motif His His His Asp still remains the best fitting cluster with *R*-factor and reduced χ^2 appreciably lower than those produced by the other models is a good indication of the reliability of our analysis procedure.

The EXAFS of the best-fitting cluster is shown in Fig. 5 compared with the experimental data. Table 10 reports the structural parameters provided by the fitting procedure. The value yielded by the fitting procedure for the common shift in the energy origin is -2 (1) eV (the starting value for the energy origin was set to 9662 eV, as determined by the first inflection point of the raw data).

In all the three reference structures that we have examined the most probable binding motifs, selected by applying our data analysis procedure, is the same as those reported from protein crystallography. This confirms that the present analysis approach allows the identification of binding motifs by means of XAFS data for mononuclear zinc binding sites and therefore that it can be applied to identify the binding motif when this is unknown.

We have recently applied the method to characterize different Zn^{2+} binding sites, the local structure of which was unknown, in charge translocating membrane protein complexes. Specifically, we have investigated the inhibitory zinc binding site of bacterial, avian and bovine cytochrome (cyt) bc_1 (Giachini, Francia, Veronesi *et al.*, 2007), of bovine-heart COX (Francia *et al.*, 2007) and of *Escherichia coli* transhydrogenase (TH) (Veronesi *et al.*, 2010). Moreover, in the bovine NADH-Q oxidoreductase (complex I) we have

detected the presence of an endogenous zinc binding site and we have characterized its local structure by XAFS (Giachini, Francia, Boscherini *et al.*, 2007). In all these cases the binding motif for the zinc ion was unknown. Limited preliminary structural information was available only in the case of the avian cyt bc_1 complex (Berry *et al.*, 2000). By applying this analysis procedure we were able to identify with significant confidence the most probable binding motifs. In some cases, new or occasionally found binding motifs were also taken into consideration, when the most common binding motifs did not reproduce satisfactorily the XAFS spectrum or when the limited complementary structural information available supported such rare or new binding sites. In this respect we recall that, at the present stage, the binding motifs present in the database are those included in the MDB. However, as described in §3.3, it is possible to build additional binding motifs, compatible with the preliminary information, rather easily by following the instructions given in §3.1.

The XAFS features of the six zinc binding sites mentioned above are shown in Figs. 7 and 8. For complex I we have found a tetrahedral cluster formed by two His and two cysteines; for transhydrogenase we have obtained a zinc binding site formed by two His, one cysteine and one carboxylic acid in a tetrahedral geometry; in COX an inhibitory tetrahedral zinc binding site has been identified formed by three His and one carboxylic acid; in the bacterial cyt bc_1 we have proposed an octahedral cluster formed by one His, two carboxylic acids, one glutamine or asparagine and two water molecules; in the avian and in the bovine cyt bc_1 we have characterized a tetrahedral cluster composed of two His and one carboxylic acid. As a fourth ligand, in this case, we proposed a lysine residue, even though this amino acid appears to be involved very rarely in zinc coordination. When comparing these data with those measured in this work for the three reference structures, some strong correlations between spectral characteristics and XAFS features are evident.

The octahedral site identified in the bacterial bc_1 exhibited a great increase in the WL intensity as compared with the tetrahedral sites (see Fig. 7). This increase is in line with the theoretical calculations performed by the finite difference method, described in §3.2.1. The value measured for the WL intensity is 1.87 and thus fulfils the quantitative criterion stated above.

In the case of the avian and bovine cyt bc_1, the proposed binding motif is similar to that of TLS, which contains two His and one aspartic acid. The value for the integral I calculated for these proteins (see Table 3) fulfils the quantitative criterion for the determination of the number of His residues stated in §3.2.3. It is worthwhile noting that when we have performed

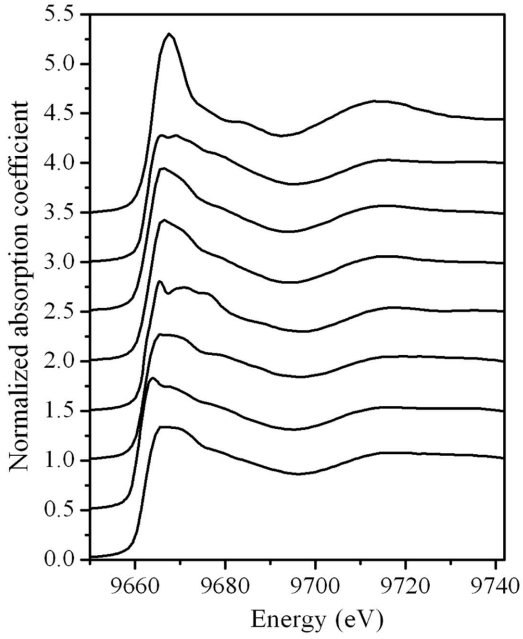

Figure 7
Experimental spectrum for all the mononuclear zinc binding sites in which we have identified the most probable binding motifs. From top to bottom: inhibitory zinc site of bacterial cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007); catalytic site of TLS (this work); inhibitory zinc site of avian cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007); inhibitory zinc site of bovine cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007); catalytic zinc site of SOD (this work); structural and inhibitory zinc site of COX, from a sample containing two zinc ions per protein (see Francia *et al.*, 2007); inhibitory zinc site of TH (Veronesi *et al.*, 2010); structural zinc site of complex I (Giachini *et al.*, 2007).

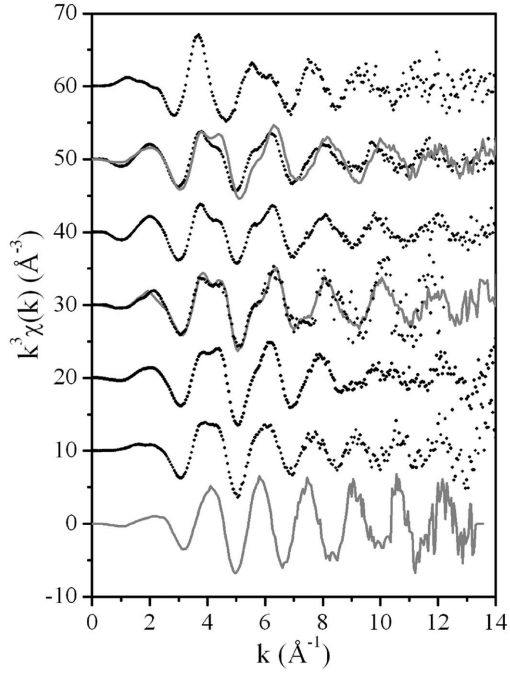

Figure 8
Experimental k^3-weighted EXAFS functions for all the mononuclear zinc binding sites in which we have identified the most probable binding motifs. Reference structures are reported with a continuous line (in grey). Zinc binding sites not deposited in the PDB are shown with filled circles (in black). From top to bottom: inhibitory zinc site of bacterial cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007); catalytic site of TLS (this work) superimposed on the inhibitory zinc site of avian cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007); inhibitory zinc site of bovine cyt bc_1 (Giachini, Francia, Veronesi *et al.*, 2007); catalytic zinc site of SOD (this work) superimposed on the inhibitory zinc site of COX (Francia *et al.*, 2007); inhibitory zinc site of TH (Veronesi *et al.*, 2010); structural zinc site of complex I (Giachini *et al.*, 2007), structural zinc site of COX (Francia *et al.*, 2007, and this work).

the analysis for the cyt bc_1 complexes we still did not use this criterion for the determination of the number of His. Therefore we have fit in those cases different models containing either two or three His. The models with two His indeed provided a better fit than the models with three His. This is a further indication of the validity of the criterion described in §3.2.3.

The binding motif that we have proposed for the inhibitory site of COX is the same exhibited by SOD. In fact the two XAFS spectra almost overlap (see Fig. 8). The data are rather noisy since they have been extracted by subtracting the spectrum of the endogenous site of COX from the spectrum acquired in a sample containing two zinc ions per protein (see Francia *et al.*, 2007). Therefore it is not possible in this case to obtain a trustworthy value for the integral I since the data do not allow a clean Fourier transformation in the interval 2–12 Å, which is the interval that we have chosen for our calculations.

For complex I and TH we proposed two sites containing a mixture of cysteines, His and carboxylic acids. The difference in the scattering properties of the different first neighbours results in spectra with features in between those of the endogenous site of COX (for which the binding motif is Cys Cys Cys Cys) and the binding motifs of TLS or SOD containing a mixture of His and carboxylic acids.

We believe that the direct validation obtained for the three reference structures, together with these observations, strongly support the reliability of the data analysis approach presented in this work.

5. Conclusions

In the present paper we have presented a data analysis procedure of XAFS data to identify most probable binding motifs for mononuclear zinc sites in metalloproteins. Zn *K*-edge XAFS measurements were performed on three selected zinc metalloproteins, each containing a different characteristic zinc binding site known from X-ray protein crystallography. The most probable cluster selected by applying our data analysis procedure to the three data sets is the same as that reported in the corresponding crystallographic structures deposited in the protein data bank. Moreover, when comparing the data for the reference models with those collected for six unknown zinc binding sites that we have previously investigated, the correspondence between the XAFS features and the binding motifs is evident. This indicates that it is possible to identify with a reasonable confidence zinc binding motifs for mononuclear sites by means exclusively of XAFS data using the analysis approach described here.

This work was supported by MIUR of Italy. Measurements at ESRF were performed within the public user program. We are grateful to the staff of the GILDA beamline of ESRF for excellent support. L. Giachini and G. Veronesi thank EMBO for supporting their participation in the school 'BioXAS Practical Course on Metalloproteins and Organism Tissue' (EMBL, Hamburg, Germany, 14–19 June 2005 and 10–15 July 2007).

References

Ankudinov, A. L., Ravel, B., Rehr, J. J. & Conradson, S. D. (1998). *Phys. Rev. B*, **58**, 7565–7576.
Aoyama, H., Muramoto, K., Shinzawa-Itoh, K., Hirata, K., Yamashita, E., Tsukihara, T., Ogura, T. & Yoshikawa, S. (2009). *Proc. Natl. Acad. Sci. USA*, **106**, 2165–2169.
Banci, L., Bertini, I. & Mangani, S. (2005). *J. Synchrotron Rad.* **12**, 94–97.
Berry, E. A., Zhang, Z., Bellamy, H. D. & Huang, L. (2000). *Biochim. Biophys. Acta*, **1459**, 440–448.
Binsted, N., Strange, R. W. & Hasnain, S. S. (1992). *Biochemistry*, **31**, 12117–12125.
Castagnetto, J. M., Hennessy, S. W., Roberts, V. A., Getzoff, E. D., Tainer, J. A. & Pique, M. E. (2002). *Nucl. Acids Res.* **30**, 379–382.
Cheung, K.-C., Strange, R. W. & Hasnain, S. S. (2000). *Acta Cryst.* D**56**, 697–704.
Ciatto, G., d'Acapito, F., Boscherini, F. & Mobilio, S. (2004). *J. Synchrotron Rad.* **11**, 278–283.
Clark-Baldwin, K., Tierney, D. L., Govindaswamy, N., Gruff, E. S., Kim, C., Berg, J., Koch, S. A. & Penner-Hahn, J. E. (1998). *J. Am. Chem. Soc.* **120**, 8401–8409.
Co, M. S., Scott, R. A. & Hodgson, K. O. (1981). *J. Am. Chem. Soc.* **103**, 986–988.
Dau, H., Liebisch, P. & Haumann, M. (2005). *Phys. Scr.* T**115**, 844–846.
Dimakis, N. & Bunker, G. (2004). *Phys. Rev. B*, **70**, 195114.
Dimakis, N. & Bunker, G. (2006). *Biophys. J.* **91**, L87–L89.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Errede, B., Kamen, M. O. & Hatefi, Y. (1978). *Methods Enzymol.* **53**, 40–47.
Feiters, M. C., Eijkelenboom, A. P. A. M., Nolting, H.-F., Krebs, B., van den Ent, F. M. I., Plasterk, R. H. A., Kaptein, R. & Boelens, R. (2003). *J. Synchrotron Rad.* **10**, 86–95.
Francia, F., Giachini, L., Boscherini, F., Venturoli, G., Capitanio, G., Martino, P. L. & Papa, S. (2007). *FEBS Lett.* **581**, 611–616.
Giachini, L., Francia, F., Boscherini, F., Pacelli, C., Cocco, T., Papa, S. & Venturoli, G. (2007). *FEBS Lett.* **581**, 5645–5648.
Giachini, L., Francia, F., Mallardi, A., Palazzo, G., Carpene, E., Boscherini, F. & Venturoli, G. (2005). *Biophys. J.* **88**, 2038–2046.
Giachini, L., Francia, F., Veronesi, G., Lee, D. W., Daldal, F., Huang, L. S., Berry, E. A., Cocco, T., Papa, S., Boscherini, F. & Venturoli, G. (2007). *Biophys. J.* **93**, 2934–2935.
Harding, M. M. (1999). *Acta Cryst.* D**55**, 1432–1443.
Harding, M. M. (2000). *Acta Cryst.* D**56**, 857–867.
Harding, M. M. (2002). *Acta Cryst.* D**58**, 872–874.
Harding, M. M. (2004). *Acta Cryst.* D**60**, 849–859.
Harding, M. M. (2006). *Acta Cryst.* D**62**, 678–682.
Hasnain, S. S. (2004). *J. Synchrotron Rad.* **11**, 7–11.
Hasnain, S. S. & Hodgson, K. O. (1999). *J. Synchrotron Rad.* **6**, 852–864.
Hasnain, S. S. & Strange, R. W. (2003). *J. Synchrotron Rad.* **10**, 9–15.
Hough, M. A. & Hasnain, S. S. (1999). *J. Mol Biol.* **287**, 579–692.
Hsin, K., Sheng, Y., Harding, M. M., Taylor, P. & Walkinshaw, M. D. (2008). *J. Appl. Cryst.* **41**, 963–968.
Joly, Y. (2001). *Phys. Rev. B*, **63**, 125120.
Joly, Y. (2003). *J. Synchrotron Rad.* **10**, 58–63.
Kelly, S. D., Kenner, K. M., Fryxell, G. E., Liu, J., Mattigod, S. V. & Ferris, K. F. (2001). *J. Phys. Chem. B*, **105**, 6337–6346.
Levina, A., Armstrong, R. S. & Lay, P. A. (2005). *Coord. Chem. Rev.* **249**, 141–160.
Matsubara, H. (1970). *Methods Enzymol.* **19**, 642–651.
Matthews, B. W., Jansonius, J. N., Colman, P. M., Titani, K., Walsh, K. A. & Neurath, H. (1972). *Nat. New Biol.* **238**, 37–41.

Mijovilovich, A. & Meyer-Klaucke, W. (2003). *J. Synchrotron Rad.* **10**, 64–68.
Muramoto, K., Hirata, K., Shinzawa-Itoh, K., Yoko-o, S., Yamashita, E., Aoyama, H., Tsukihara, T. & Yoshikawa, S. (2007). *Proc. Natl. Acad. Sci. USA*, **104**, 7881–7886.
Newville, M., Ravel, B., Haskel, D., Rehr, J. J., Stern, E. A. & Yacoby, Y. (1995). *Physica B*, **208–209**, 154–155.
Pascarelli, S., Boscherini, F., D'Acapito, F., Hrdy, J., Meneghini, C. & Mobilio, S. (1996). *J. Synchrotron Rad.* **3**, 147–155.
Peariso, K., Huffman, D. L., Penner-Hahn, J. E. & O'Halloran, T. V. (2003). *J. Am. Chem. Soc.* **125**, 342–343.
Ravel, B. & Newville, M. (2005). *J. Synchrotron Rad.* **12**, 537–541.
Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y., Yamasaki, A., Sugimura, T., Kurono, S., Tsujimoto, K., Mizushima, T., Yamashita, E., Tsukihara, T. & Yoshikawa, S. (2007). *EMBO J.* **26**, 1713–1725.
Stern, E. A., Newville, M., Ravel, B., Yacoby, Y. & Haskel, D. (1995). *Physica B*, **208–209**, 117–120.
Strange, R. W., Blackburn, N. J., Knowles, P. F. & Hasnain, S. S. (1987). *J. Am. Chem. Soc.* **109**, 7157–7162.
Strange, R. W., Ellis, M. & Hasnain, S. S. (2005). *Coord. Chem. Rev.* **249**, 197–208.
Trainer, J. A., Getzoff, E. D., Richardson, J. S. & Richardson, D. C. (1983). *Nature (London)*, **306**, 284–287.
Ugliengo, P., Viterbo, D. & Chiari, G. (1993). *Z. Kristallogr.* **207**, 9–23.
Veronesi, G., Whitehead, S. J., Francia, F., Giachini, L., Cotton, N. P., Boscherini, F. Venturoli, G. & Jackson, J. B. (2010). *Biochim. Biophys. Acta.* Submitted.