Journal of
Synchrotron Radiation

ISSN 0909-0495

Received 28 August 2009
Accepted 5 October 2009

Towards a black-box for biological EXAFS data analysis. II. *Automatic BioXAS Refinement and Analysis* (*ABRA*)

Gerd Wellenreuther,‡ Venkataraman Parthasarathy and Wolfram Meyer-Klaucke*

EMBL Hamburg, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany.
E-mail: wolfram@embl-hamburg.de

In biological systems, X-ray absorption spectroscopy (XAS) can determine structural details of metal binding sites with high resolution. Here a method enabling an automated analysis of the corresponding EXAFS data is presented, utilizing in addition to least-squares refinement the prior knowledge about structural details and important fit parameters. A metal binding motif is characterized by the type of donor atoms and their bond lengths. These fit results are compared by bond valance sum analysis and target distances with established structures of metal binding sites. Other parameters such as the Debye–Waller factor and shift of the Fermi energy provide further insights into the quality of a fit. The introduction of mathematical criteria, their combination and calibration allows an automated analysis of XAS data as demonstrated for a number of examples. This presents a starting point for future applications to all kinds of systems studied by XAS and allows the algorithm to be transferred to data analysis in other fields.

Keywords: EXAFS; BioXAS; refinement; metalloproteins; *ABRA*; automation.

1. Introduction

Metal ions are essential for all organisms. They play a pivotal role in biological processes such as respiration, metabolism, photosynthesis, cell division, muscle contraction, nerve impulse transmission and gene regulation. X-ray absorption spectroscopy (XAS) can determine a metal binding site with high resolution and thereby elucidate its function in the system. At present, the evaluation of biological XAS data still requires expert knowledge. This is in strong contrast to other techniques such as protein crystallography, where the pipeline from crystallization to data collection and even model building is strongly supported by increasingly automated software (Maniasetty *et al.*, 2008; Panjikar *et al.*, 2005). Biological XAS serves as the ideal test case for similar automation in XAS, because the type of potential metal ligands is rather limited and at the same time a high demand exists visualized by an increasing number of high-impact projects performed in recent years (Shima *et al.*, 2008; Banci *et al.*, 2005; Liu *et al.*, 2007; Küpper *et al.*, 2008; Pufahl *et al.*, 1997; Masip *et al.*, 2004; Hwang *et al.*, 2000; Harris *et al.*, 2003; Haumann *et al.*, 2008).

The first steps of XAS data processing include dead-time correction, energy calibration, an automatic selection of fluorescence detector channels, as well as extraction of X-ray absorption near-edge structure (XANES) and extended X-ray absorption fine structure (EXAFS). An automated treatment has been implemented in *KEMP* (Korbas, Marsa & Meyer-Klaucke, 2006), *KEMP2* (Wellenreuther & Meyer-Klaucke, 2009) and to a different extent in packages such as *WinXAS* (Ressler, 1998), *Viper* (Klementev, 2001), *EXAFS-PAK* (George & Pickering, 1993) or *ATHENA* (Ravel & Newville, 2005), a graphical user interface based on *IFEFFIT* (Newville, 2001). For the next step, the refinement of the extracted fine structure, several interactive programs are available, including some of the above-mentioned packages as well as *ARTEMIS* (Ravel & Newville, 2005), *Excurve* (Binsted, 1998; Tomic *et al.*, 2004) and *GNXAS* (Westre *et al.*, 1995; Filipponi *et al.*, 1995; Filipponi & Di Cicco, 1995). Although the softwares differ in strategy and approaches, their results are highly similar.[1] Thus, improvements in the quality of EXAFS analysis will in most cases not be triggered by further development of the theory but result from inclusion of additional information. This additional information, *e.g.* based on modelling of Debye–Waller factors (Dimakis & Bunker, 1998, 2004, 2005, 2006; Dimakis *et al.*, 1999, 2008; Bunker *et al.*,

‡ Now at HASYLAB, DESY, Notkestrasse 85, 22607 Hamburg, Germany.

[1] At the EMBO BioXAS training course in 2007 in Hamburg, Germany, M. Newville and W. Meyer-Klaucke compared the refinement of a complex biological iron binding motif (Shima *et al.*, 2008). Within the error margins the resulting parameters for each model were identical for *Excurve* and *Artemis*/*Feff*.

2005), molecular dynamics calculations (D'Angelo *et al.*, 2002) or the introduction of boundary conditions by the bond valence sum method (Newville, 2005), will lower the number of free parameters and increase the accuracy of the EXAFS model. Until now, no algorithm has been published that combines these approaches for automatic refinement of EXAFS spectra.

In a pioneering project, we showed that biological zinc *K*-edge XAS data can automatically be classified into zinc finger proteins and typical catalytic active sites (Wellenreuther & Meyer-Klaucke, 2007). This progress required focusing on typical biological metal ligands/donor atoms (sulfur, oxygen and nitrogen from imidazole groups) and was mainly based on the proper distinction between sulfur coordination and light ligands (here oxygen and nitrogen). The correct discrimination of the latter two light ligands is challenging owing to their similar back-scattering properties. Following this successful pilot project the algorithm has been improved significantly, resulting in the program *ABRA* (*Automatic BioXAS Refinement and Analysis*). The objective has been extended from the quantification of sulfur ligands in zinc-binding sites to the determination of the most probable metal-binding motif for all 3*d* metals based on first-shell contributions and multiple scattering within the imidazole unit.

ABRA achieves this ambitious goal for typical metal binding sites by incorporating prior knowledge of structural properties of metalloproteins. For this purpose, *ABRA* judges the quality of a refinement by criteria that an expert researcher might use. Typically, structural models differ in the reproduction of experimental data, summarized by the goodness of fit, but also in the derived parameters such as individual bond lengths, Debye–Waller factors and Fermi energy shifts.

Even for a subset of structural models, typically at least ten, a manual refinement takes considerable time, while being highly repetitive, limiting the effectiveness in everyday data analysis. Such repetitive tasks can be automated. In fact, nowadays *all* potential binding motifs (some 100 to 1000 models) can be refined automatically within one or two hours, saving the time of EXAFS experts. Unsuitable models are filtered out on the basis of physical and chemical criteria. The resulting fits of the remaining potential binding motifs provide an ideal starting point for a systematic analysis and comparison (meta-analysis), without running into the danger of biasing towards a favourite model.

2. An algorithm for automated refinement of EXAFS data

The number of different ligand types and their coordination numbers span the space of structural models, *e.g.* a generic biological metal binding site is here expressed as $S_xHis_yO_z$, leaving out the limited number of other ligands as CN (Korbas, Vogt *et al.*, 2006; Shima *et al.*, 2008) or prosthetic groups such as heme (Labhardt & Yuen, 1979) as well as contributions from remote aromatic ligands such as tryptophan (Xue *et al.*, 2008). Sampling this model space with equidistant variations of the coordination numbers yields the *pool of models* (typically comprising 400 structural models of the generic type $S_xHis_yO_z$ with different x, y and z and $x + y + z \in$ [minimal number of ligands; maximal number of ligands]).

Considering the inherent error margin of 20 to 25% in coordination numbers determined by EXAFS data analysis we strongly encourage the sampling of coordination space at most in steps of 0.5 of each donor group, even in the case of mononuclear metal sites. This will ensure proper sampling, and *ABRA*'s meta-analysis will calculate non-integer coordination numbers anyway. Consequently, the results of the meta-analysis approximate the location of the optimal structural model in coordination space. For example, a result of 3.3 S and 0.9 O strongly favours the structural binding motif of 3 S and 1 O in the case of a mononuclear metal site, while in the case of a dinuclear metal site 3.5 S and 1 O should also be considered (in this case one of the two metal binding sites might have four and the other five donor atoms). For multinuclear metal binding sites a finer sampling is beyond the scope of an initial analysis, because it requires additional input, *i.e.* the sequence as in Peroza *et al.* (2009).

For practical purposes, available computer power is imposing a lower limit of around 0.25 step width for sampling the coordination space, and owing to the strong correlation of coordination numbers with Debye–Waller factors a finer sampling will in our experience not increase the accuracy of the extracted data.

To ensure proper sampling of coordination space throughout the analysis the coordination numbers have to remain constant during each refinement of individual models. Afterwards the results are analysed: the structural models are sorted according to their *total score*, which is based on several criteria. Finally, the meta-analysis calculates on the basis of the top-scoring models an average best model including an estimation of the errors of its coordination numbers. *ABRA* is not reducing the number of models in the final data pool! While an early removal of models would speed up the evaluation, it is difficult to define general absorber-independent removal criteria. Thus all structural models will be fitted to the data, allowing users to check the performance of their working hypothesis.

Central to the algorithm is the evaluation of the quality of individual refinements. This evaluation is based on several criteria, *e.g.* the reduced χ^2, the element-specific bond distances *etc.* In general, one rejects any model failing in a single criterion, *e.g.* a model with abnormal bond lengths or Debye–Waller factors. Consequently, *ABRA*'s criteria are designed to be mutually obligatory; significant failure in *any* criterion implies the overall failure of the corresponding model. Accordingly criteria are expressed as scores ranging from 0.0 (complete failure of a criterion) to 1.0 (perfect match). To reject models being low in any individual score, *ABRA*'s total score is calculated as a weighted geometrical mean over all criteria,

$$\text{total score} = \prod_i C_i^{w_i}. \quad (1)$$

Herein C_i represents the score for criterion i, and w_i the *criterion weights*, determined on the basis of reference data sets, with

$$\sum w_i = 100\%. \quad (2)$$

In this concept the automatic XAS-refinement is reduced to the formulation of proper criteria and the adjustment of the corresponding criterion weights using a set of model spectra. The following sections introduce the criteria and the calculation of their optimal weights, while details on implementation and usage of *ABRA* are given in the supplementary material[2].

2.1. Definition of criteria

So far, the following criteria have been introduced in *ABRA*: goodness-of-fit, bond length, bond valence sum, Fermi energy and Debye–Waller factor. Their definitions and weightings in *ABRA*'s scoring routine are defined as follows.

Goodness-of-fit. The goodness-of-fit quantifies the quality of a refinement. *Excurve* (Binsted *et al.*, 1992; Binsted & Hasnain, 1996; Binsted, 1998; Gurman *et al.*, 1984, 1986) minimizes a special fit-index $\Phi_{\rm EXAFS}$ and additionally provides both *R*-factor and reduced χ^2 (Lytle *et al.*, 1989). In *ABRA* the first criterion is based on the reduced χ^2 because (i) it progressively penalizes deviations of theory from measurement and (ii) it takes into account the degree of over-determination in the refinement,

$$\chi^2_{\rm red} = \left[\frac{1}{(N_{\rm ind} - N_{\rm pars})}\right]\left(\frac{N_{\rm ind}}{N}\right)\sum_i^N k^6\left[\chi_i^{\rm exp}(k) - \chi_i^{\rm th}(k)\right]^2, \quad (3)$$

where the number of relevant independent points (Lytle *et al.*, 1989) is calculated as

$$N_{\rm ind} = \frac{2\Delta k \Delta R}{\pi} + 1. \quad (4)$$

Herein k^3-weighting, typical for the analysis of biological EXAFS data, is applied and a k-independent statistical error is assumed. Frequently, no exact experimental errors are provided. Therefore, the reduced χ^2 depends on the data quality and cannot serve as an absolute criterion. In model datasets we obtained values of the order of 1.0 for reasonable fits, while values for unreasonable models typically fell in the range from 5.0 to 20.0 and more. Therefore the results are projected on the interval [0.0, 1.0] with a value of 0.0 corresponding to a reduced $\chi^2 \geq 10.0$ (= complete failure) and a value of 1.0 corresponding to reduced $\chi^2 \leq 1.0$ (= complete success).

Bond length. In order to exclude models resulting in chemically unreasonable bond lengths, *ABRA* compares each distance with an internal database (see §2.2 below). Based on the analysis of the Cambridge Structural Database (CSD; Allen, 2002) and the Brookhaven Protein Data Bank (PDB; Berman *et al.*, 2000) the mean distance and its standard deviation is known for most metal–ligand pairs. Assuming a normal distribution of the distances, the probability of an individual bond distance can be estimated, and used as a criterion as done in our pilot study (Wellenreuther & Meyer-Klaucke, 2007). However, the bond distance of a ligand depends on the overall coordination number, ligand types, as well as metal oxidation and spin state (Harding, 1999, 2000, 2001, 2002, 2004, 2006; Paulsen *et al.*, 2001). The latter dependency is frequently neglected and thus no suitable generalized data are available. To overcome the problem of bimodal bond-length distributions for two spin states caused by different ionic radii we exclude at present high-spin compounds with very long bond lengths [*e.g.* octahedral high-spin (Paulsen *et al.*, 2001)]. As a potential workaround we intend implementing a treatment similar to the one for oxidation states, where both scenarios are calculated and compared (see below).

The remaining distribution is therefore assumed to be unimodal. Given the fact that differences in bond lengths are not statistical in nature, the usage of a Gaussian distribution suggesting the existence of a 'true mean' bond distance would be a crude approximation. Thus, we introduced a uniform distribution with smoothed edges (shown for Zn—S in Fig. 1): all bond distances of ~2.28–2.34 Å are 'correct' for Zn—S, and the score quickly drops only outside this range. In contrast, a Gaussian distribution would either be too sharp, penalizing correct distances (see Fig. 1; Gaussian with σ = 0.025 Å), or too wide, allowing unrealistic distances (Gaussian with σ = 0.05 Å). Thus an individual rectangle with smoothed edges is defined for each 3*d*-metal–ligand pair, providing a score for every individual bond distance. For the centre of the rectangle we used the ideal target distances determined by Harding (2006) (*e.g.* 2.31 Å for Zn—S). The full width of any first-shell ligand was set to 0.05 Å; correspondingly all Zn—S distances ranging from 2.285 to 2.335 Å achieve a full score of 1.0 (see Fig. 1). Longer distances, *e.g.* metal–metal distances, were given much more relaxed ranges using a full width of 0.6 Å, thereby keeping them distinct from shorter first-shell distances. On this basis the total bond-length criterion is

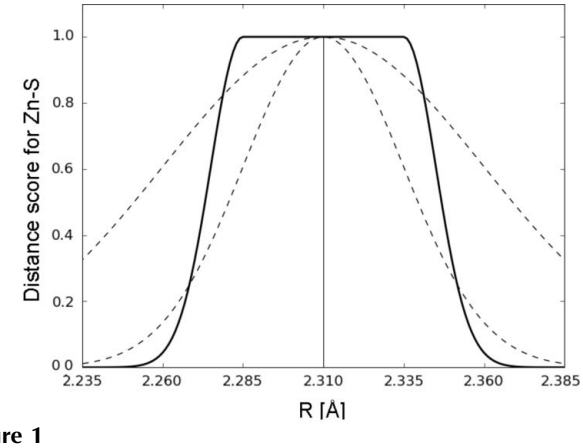

Figure 1
Comparison of smoothed rectangle (solid line) with Gaussian distribution (dashed lines with σ = 0.05 Å and 0.025 Å) for defining the target distances for the Zn—S bond.

[2] Supplementary data for this paper are available from the IUCr electronic archives (Reference: HI5604). Services for accessing these data are described at the back of the journal.

calculated as the geometrical mean, weighted by the coordination number of each ligand type. Again multiplication ensures that a single questionable ligand distance causes a low value of the total score, and thus the corresponding structural model is rejected.

Bond valence sum. Neither goodness-of-fit nor individual bond distances allow full judgement on how sensible the *total* coordination of a model is. Bond valence sum (BVS) analysis is a simple tool for checking this (Brown & Altermatt, 1985; Thorp, 1992, 1998; Liu & Thorp, 1993),

$$\mathrm{BVS} = \sum \exp\left(\frac{R_i' - R_i}{0.37}\right). \quad (5)$$

It considers the differences in first-shell distances R_i and tabulated BVS distance R_i' and assumes the valences for the first-shell ligands summing up to the oxidation state of the metal $\pm$ 0.25 (Brown & Altermatt, 1985). These limits of uncertainty in the BVS also allow for the possibility of weak additional bonds neglected in these initial models, *i.e.* intermediate binding modes between monodentate and bidentate carboxylates. From BVS a score is calculated based on a Gaussian distribution around the expected oxidation state with a standard deviation of $\sigma = 0.5$. For unknown oxidation states for each structural model all corresponding bond valence sums are calculated, and the one with the highest score is chosen for the computation of the total score. This enables *ABRA* to estimate the oxidation state of a sample based on the distances extracted from the EXAFS spectrum. Obviously, mixed oxidation states caused by the presence of two binding sites stabilizing different metal oxidation states, *e.g.* 1 $\times$ Fe(II) and 1 $\times$ Fe(III), or half oxidation/reduction of the metal ions should be indicated by a half-integer oxidation state estimate in the final results (here $\sim$2.5).

Fermi energy. Despite their different phases, backscattering from S ligands can be 'successfully' modelled by light ligands and *vice versa*. Here, the wrong phases are 'corrected' by shifting the Fermi energy. *ABRA* detects these shifts and penalizes them so that Fermi energies in the interval from -7.5 to -2.5 eV obtain a score of 1.0, and outside this range (σ = 1 eV) the score drops to zero. Note that this strategy requires a consistent definition of the edge position as implemented in some data reduction programs (Korbas, Marsa & Meyer-Klaucke, 2006). For Fe, 7120 eV was used as E_0. For other elements, 3 eV were added to the edge positions tabulated for corresponding metal foils (Thompson *et al.*, 2001).

Debye–Waller factor. Unrealistic values of the Debye–Waller factors may indicate incorrect structural models (George *et al.*, 1999). Too small Debye–Waller factors artificially magnify a ligand contribution, whereas too large Debye–Waller factors do the reverse. During the automatic refinements in *DL_EXCURV* the lower and upper limits for the Debye–Waller factor ($2\sigma^2$) are set to 0.003 Å^2 and 0.030 Å^2, respectively. *ABRA* considers all Debye–Waller factors in the range from 0.004 Å^2 to 0.025 Å^2 as reasonable. Any model with a first-shell Debye–Waller factor outside *ABRA*'s margins is rejected. This criterion works very well for joint refinements of all first-shell Debye–Waller factors. Whether it has to be slightly lowered for individual refinements of Debye–Waller factors, time will tell.

Thus, the Debye–Waller criterion is a true Boolean decision leading to either outright rejection of the questionable model or its acceptance. All other criteria (goodness-of-fit, bond length, Fermi energy and BVS) are taken into account depending on their individual criterion-weight for the calculation of the total score.

2.2. Databases

Ideal bond lengths were based on M. M. Harding's revised metal–donor atom target distances, which have been obtained by filtering both the PDB and CSD. From the PDB of March 2005 all datasets with resolution $\leq$ 1.25 Å were used to generate a database of metal clusters. Typical bond distances varied by less than 0.10 Å, and outliers by more than 0.40 Å were rejected (Harding, 2006). Harding also queried the CSD in November 2005 for metal clusters, only considering datasets with a crystallographic R-factor < 0.065 and again excluding outliers [for further details, see Harding (1999)].

These target distances are used in *ABRA*, with one exception: the target bond distance of Fe–His of 2.16 Å, even higher than the target bond length for the Fe(II) high-spin binding site considered here, was replaced by the PDB value of 2.03 Å (Harding, 2006), in line with low-spin iron states and many Fe(II) high-spin binding sites. For oxygen donors those values covering carboxylates were taken, which are typically close to those tabulated for histidine groups. Individual oxygen donors might be present at shorter distances (Wolter *et al.*, 2000; Duda *et al.*, 2003), which will result, for the joint refinement of histidine and oxygen groups, that we strongly suggest, in rather large Debye–Waller factors prompting at present an additional manual refinement.

Bond valence sum parameters were taken from Brese & O'Keeffe (1991) and O'Keeffe & Brese (1992). At present, the internal database covers Mn(II), Mn(III), Fe(II), Fe(III), Co(II), Co(III), Cu(I), Cu(II) and Zn(II). It reflects the dependence on the oxidation state if those values were available. An additional database for Fe–S clusters has been established, because the bond distances for sulfide differ considerably from those of sulfur-containing amino acids.

2.3. Meta-analysis

Owing to very similar scattering properties it is difficult to differentiate between the light ligands nitrogen and oxygen. Thus, several models obtain similar scores, and the selection of a single top-scoring model can hardly be justified. It would be a mistake to claim that *ABRA*'s top model always models the data best.

Moreover, in the absence of a covariance matrix including coordination numbers, *ABRA* mimics an expert: confronted with 400 different structural models he/she might look at the top-scored ones, trying to identify a common pattern. The crucial part herein is to define *top-scoring*: which models are virtually indistinguishable from the best model? The easiest

Table 1
Comparison of published Zn-EXAFS data with *ABRA*'s results (1σ errors on the last digit extracted by meta-analysis are given in parentheses).

S, His and O are the zinc donor groups and their numbers are given for the different samples and structural models. The sum of His and O donor groups given by the meta-analysis is abbreviated as Low-*Z* ligands.

	Number of ligands							
	Published results			Best model by *ABRA*			Average of good models (meta-analysis)	
	S	His	O	S	His	O	S	Low-*Z*
Zinc finger proteins								
HPV E7	4.0			4.0			4.0 (4)	0.2 (2)
GCM	3.0	1.0		3.0	1.0		3.0 (0)	1.2 (2)
ZnF-UBP	2.67	1.33		2.67	1.00	0.33	2.8 (3)	1.2 (3)
Catalytic active sites								
Ec ZiPD		2.0	2.5		3.0	1.5	0.0 (0)	4.2 (2)
At Glx2-2		2.5	2.5		4.0	0.5	0.0 (0)	4.2 (2)
Bc bla	0.5	3.0	1.0		3.0	1.0	0.1 (2)	4.0 (4)
Model compounds								
Bis(acetato)bis(imidazol)zinc(II)		2.0	2.0		2.0	2.0	0.0 (0)	4.2 (2)
Tetrakis(imidazole)zinc(II)-perchlorate		4.0			3.5	0.5	0.0 (0)	4.3 (2)

criterion for a selection is a threshold. Typically, the two best models differ by about a percent in the total score. Therefore, a relative *meta-analysis threshold* of 2.5% marks ~5–20 models out of a pool of 400 as 'top scoring'. The final step of the meta-analysis is to average these 'top-scoring' models, which yields the mean model with error margins for each fit parameter, now including the coordination numbers. Typically, the top-scoring model lies within these error margins.

In rare cases the averaging process leads to errors being exactly equal to zero for two reasons: (i) the coordination of a certain ligand is constant for all 'top-scoring' models or (ii) the second best model is already below the threshold. In the latter case the error for *all* coordination numbers is set to zero, indicating *ABRA*'s confidence in the overall top-scoring model. The occurrence of zero error margins is tuned by the meta-analysis threshold. Higher thresholds give less often error margins of zero for coordination numbers, but require incorporating increasingly worse models into the meta-analysis. Values significantly above 2.5% result in uncertainties larger than the systematic errors inherent to EXAFS-data analysis (Sayers, 2000). Based on our experience, 2.5% is a good choice for the meta-analysis threshold, resulting in zero error estimates typically in justified cases.

Obviously, the error margins determined by the meta-analysis are purely statistical, so the user has to further take into account any systematic errors, *e.g.* those intrinsic to EXAFS.

2.4. Determination of optimal criterion-weights

The influence of each criterion on the total score is determined by its criterion-weighting factors $w_{\rm Chi}$, $w_{\rm BVS}$, $w_{\rm Bond}$ and $w_{\rm FE}$ in equation (1). The quality of *ABRA*'s prediction varies depending on the set of criterion-weights. In general, the optimal criterion-weights might differ for individual absorption edges. Consequently, the *optimal set of criterion weights* is determined individually for two absorption edges. This requires characteristic and well grounded datasets for both absorption edges to properly 'train' *ABRA*. In short, initially each dataset is refined with *ABRA* once, and then scored with *different* sets of criterion weights. For each set *ABRA*'s prediction is compared against the expected targets, yielding a measure (defined in detail in the following paragraph) for the performance with this set of weights. Its minimum marks the optimal set of criterion-weights.

Using this approach, the optimal sets of criterion-weights were established individually for the Zn and Fe absorption edges; the datasets and *ABRA*'s optimal results are given in Tables 1 and 2, respectively. From these datasets the coordination numbers ($N_{p\rm S}$, $N_{p\rm His}$, $N_{p\rm O}$) as well as their total number of light ligands ($N_{p\rm ll}$) and their oxidation state ($N_{p\rm Ox}$) served as target parameters for the *deviation from published models* ε,

$$\varepsilon = \sum_{\rm datasets} \sum_{\rm parameters} c_i \left(N_{pi} - N_{Ai}\right)^2, \qquad (6)$$

with N_{Ai} being the result of the meta-analysis and c_i the *penalization weights*. The penalization weights are required to balance *ABRA*'s capability to correctly determine individual ligands, and the oxidation state (*e.g.* a high penalization-weight $c_{\rm S}$ pushes the criterion weights to ensure a proper estimation of the sulfur coordination, while worsening the performance in all other fields). The penalization weights c_i were set to

$$c_{\rm S} = 1.0, \quad c_{\rm His} = c_{\rm O} = 0.25, \quad c_{\rm ll} = 0.75, \quad c_{\rm Ox} = 1.0. \qquad (7)$$

Thus a deviating coordination number of S has the same effect as a missing light ligand ($N_{\rm ll}$ and either $N_{\rm His}$ or $N_{\rm O}$ wrong by one leads to a total penalization $c_{\rm His/O} + c_{\rm ll} = 1.0 = c_{\rm S}$). Any imidazole modelled as an oxygen adds $c_{\rm His} + c_{\rm O} = 0.5$ to ε, the deviation from published models. An incorrect oxidation state [*e.g.* Fe(II) instead of Fe(III)] is counted in the same way with $c_{\rm Ox} = 1.0$. Thereby, the optimization procedure has a major focus on the proper determination of coordination numbers for sulfur and light ligands and only to a smaller extent on the oxidation state. The correct determination of the fit parameters is ensured by optimizing the *criterion weights*.

Table 2
Comparison of published Fe-EXAFS data with *ABRA*'s results (1σ errors on the last digit extracted by meta-analysis are given in parentheses).

S, His and O are the zinc donor groups and their numbers are given for the different samples and structural models. The sum of His and O donor groups given by the meta-analysis is abbreviated as Low-*Z* ligands. The iron oxidation state resulting from the bond valance sum analysis is given in the column OS.

	Number of ligands										
	Published results				Best model by *ABRA*				Average of good models (meta-analysis)		
	S	His	O	OS	S	His	O	OS	S	Low-*Z*	OS
Ss ABCE1	4			III	4.0			III	4.2 (2)	0.0 (0)	3.0 (0)
Pa RM2-4$_{ox}$	4			III	4			III	4.2 (2)	0.0 (0)	3.0 (0)
Pa RM2-4$_{red}$	4			II	3.5			II	4.0 (5)	0.0 (0)	2.5 (5)
Hs TH1$_{nat}$		2	3	II		0.5	4.5	II	0.0 (0)	4.8 (2)	2.0 (0)
Hs TH1$_{ox}$		2	3	III		2.5	3.5	III	0.2 (3)	5.1 (8)	2.8 (4)
Fe-Pvd			6	III		1	5	III	0.0 (0)	5.8 (2)	3.0 (0)

During the determination of optimal weights the *criterion-weight-space* was sampled in steps of 0.05, and in steps of 0.01 for the subspace with $w_{Chi} \geq 0.5$ for both absorption edges. The coarse sampling for $w_{Chi} < 0.5$ is fully sufficient since w_{Chi} turned out to be always the major component of the optimal set of criterion weights.

Based on sum-rule (2), three criterion-weights would have two degrees of freedom, so that they can be visualized in so-called ternary plots (Philpotts, 1990; Möbius, 1827). Here, each point inside the equilateral triangle corresponds to a combination of three weights: [33%, 33%, 33%] describing the centre and a permutation of [100%, 0%, 0%] each corner. Since *ABRA* is currently using four criteria, this requires the extension from a two-dimensional ternary triangle to a three-dimensional quaternary tetrahedron. Owing to the obvious problems of displaying complex three-dimensional data in a two-dimensional projection, we have chosen to display only slices through this tetrahedron. For Zn three slices for given values of w_{Chi} through the corresponding quaternary tetrahedron are shown in Fig. 2. Herein, the red areas indicate low values of ε and therefore the best overall performance of *ABRA*, while yellow illustrates poorer performance. A trend towards lower scores is evident for higher values of w_{BVS}. The absolute minimum for Zn was found for w_{Chi} = 87%, w_{BVS} = 9%, w_{Bond} = 3% and w_{FE} = 1%, and is indicated in Fig. 2(*b*) by the small black circle. Combinations of criterion-weights with at least one vanishing component lie on the sides of the tetrahedron; their scores are typically much worse as seen from their light blue colour. Since the colour scales were limited to $\varepsilon < 75$, the importance of using all four criteria is underestimated; ε can grow to up to 100–1000 for only three criteria compared with a value of $\sim$12 in the minimum.

In order to judge the improvement of each criterion, the lowest values of ε are plotted as a function of individual criterion-weights (Fig. 3). The position of the absolute minimum is indicated by the dashed lines. The shapes of the minima are visualized by enlarged circles representing an increase of ε by less than 5%. In Fig. 3(*a*) the steep slope for w_{Chi} below 100% represents the tremendous improvement of *ABRA*'s performance upon utilization of the other three criteria, reducing ε by a factor of six. The impact of individual criteria can be observed in the other plots: without the BVS criterion (Fig. 3*c*) the deviation from published models ε is not

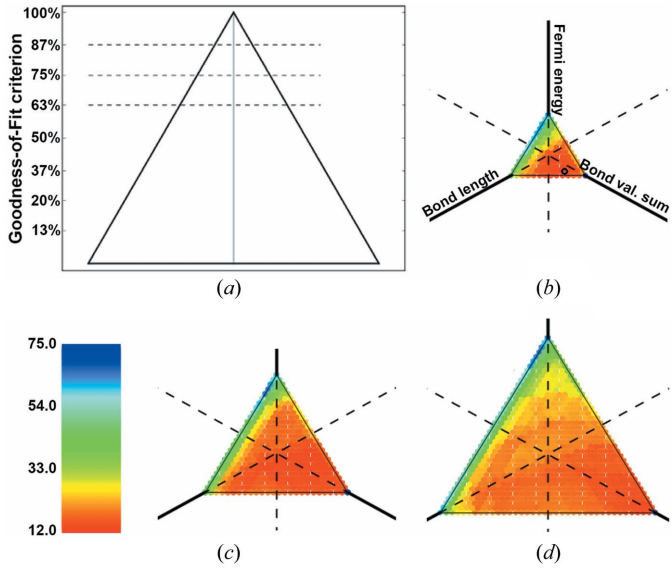

Figure 2
Estimation of optimal weights for Zn data refinement. Three slices through the quaternary tetrahedron for fixed values of the goodness-of-fit criterion (*a*) are plotted, resulting in three ternary plots (*b*–*d*) showing the weighted sum of deviation ε from published results [equations (6) and (7)]. The colour scale was cut off at a maximum ε of 75. The position of the absolute minimum is designated by a small black circle in (*b*).

better than 35, which is reduced by a factor of three for w_{BVS} = 9%. The other two criteria, bond length (see Fig. 3*b*) and Fermi energy (Fig. 3*d*), have a smaller but noticeable influence on the score, and their minima are well localized within a few percent points.

The optimization procedure for Fe results in a similar morphology of ε (Fig. 4). The absolute minimum can again be found in the region corresponding to high BVS and small but non-zero Fermi energy criterion weights w_{Chi} = 62%, w_{BVS} = 19%, w_{Bond} = 15% and w_{FE} = 4%. Still the goodness-of-fit criterion is most important, but all other weights have increased. Again the minima are rather wide and ε improves by a factor of five upon introduction of the three additional criteria (Fig. 5). Although the positions of the minima differ considerably for Fe and Zn, a detailed inspection shows that for both weighting schemes the resulting ε varies only slightly. Thus general trends can be extracted from these two cases, which will lead to criterion weight applicable to all other

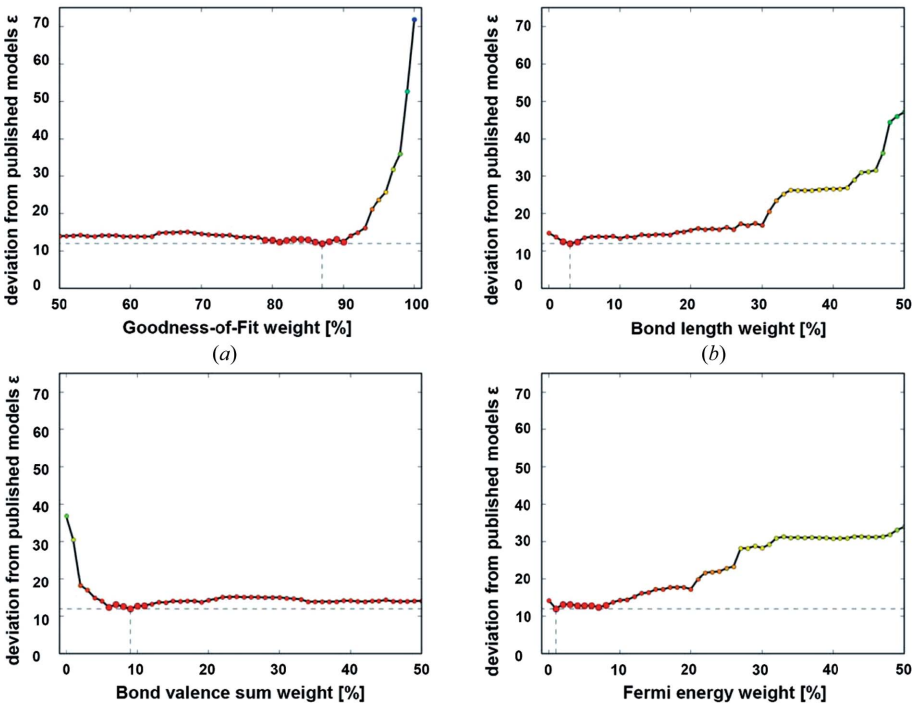

Figure 3
Estimation of optimal weights for Zn. The sum of deviations from published models ε [equation (6)] for each combination of criterion-weights is plotted over one of the four weights. The lowest scores for each value of the criterion-weights are connected by a line. The combinations yielding ε not larger than +5% of the overall minimum are enlarged, indicating the dimension of the minimum with respect to the corresponding criterion-weight.

absorber elements: the Fermi energy criterion has for Fe and Zn the smallest weight, 1% and 4%, respectively. Based on the shape of the minimum for Zn we use 4% for all other elements. The bond length weight does vary considerably between 3% and 15% for Fe, but not for Zn. Thus we set it to 13%. In contrast, the BVS weight changes very smoothly for both metals. Here an average value (14%) seems to be justified. This leaves a weighting of 69% on the goodness-of-fit, which seems to be reasonable for both absorber elements.

Furthermore, with the help of this optimization procedure, additional criteria were tested, *e.g.* goodness-of-fit of the Fourier-transformed spectrum. In the optimization procedure the corresponding weight was refined to zero, indicating that it does not improve the quality of the analysis.

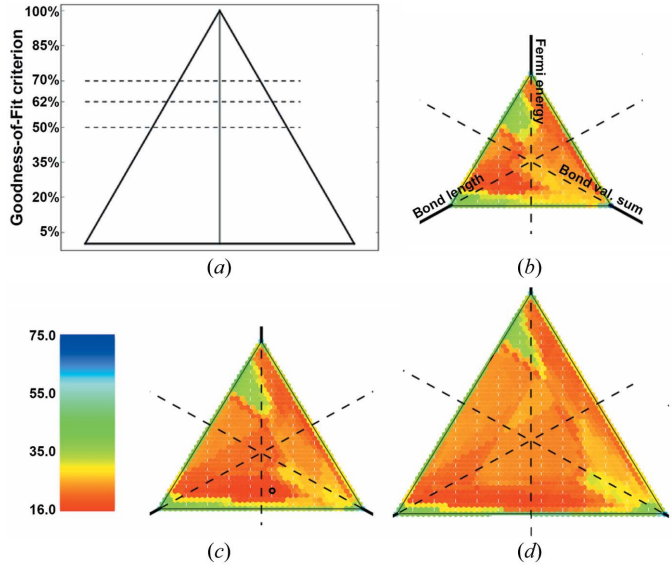

Figure 4
Estimation of optimal weights for Fe data refinement. Three slices through the quaternary tetrahedron for fixed values of the goodness-of-fit criterion (*a*) are plotted, resulting in three ternary plots (*b–d*) showing the weighted sum of deviation ε from published results [equations (6) and (7)]. The colour scale was cut off at a maximum ε of 75. The position of the absolute minimum is designated by a small black circle in (*c*).

3. Results

3.1. Zinc data

The most important benchmark is *ABRA*'s performance regarding the model datasets. For Zn these are three structurally different Zn-finger proteins, oncoprotein E7 of human papilloma viruses (*HPV* E7), glial cells missing domain of mGCMa (GCM), and HDAC6 ZnF-UBP domain (ZnF-UBP) (Ohlenschläger *et al.*, 2006; Cohen *et al.*, 2002; Boyault *et al.*, 2006), three Zn-dependent enzymes, *Escherichia coli* ZiPD (*Ec* ZiPD), *Arabidopsis thaliana* glyoxalase II (*At* Glx2-2), and Bacillus cereus, strain 569/H/9, metallo beta-lactamase (*Bc* bla) (Vogel *et al.*, 2004; Schilling *et al.*, 2003; Paul-Soto *et al.*, 1999), and two model compounds (Feiters *et al.*, 2003) as summarized in Table 1. Our primary goal is to properly determine the structural binding motif in all cases. Owing to the limited accuracy of EXAFS data analysis in coordination numbers (Teo & Joy, 1981), a determination is defined as *proper* if all ligands are identified and correctly quantified, allowing only an error margin of ± 0.5 with respect to the published values.

The published S-, His- and O-coordination numbers are given in the first column of Table 1, followed by *ABRA*'s top-ranking model in the middle column. Comparison of *ABRA*'s results with the published values shows that the S-coordination is only off once by 0.5 (in the case of the metallo-ß-lactamase from Bacillus cereus II); in this case the His- or the O-coordination is still correct. The individual coordinations of His and O as determined by *ABRA* are often off, but the total coordination numbers for light ligands are all correctly determined within ± 0.5. The reason for this is the EXAFS-intrinsic difficulty in distinguishing nitrogen and oxygen.

The last pair of columns of Table 1 show the results of *ABRA*'s meta-analysis, providing the number of S and light ligands, in addition to their estimated error margins. The

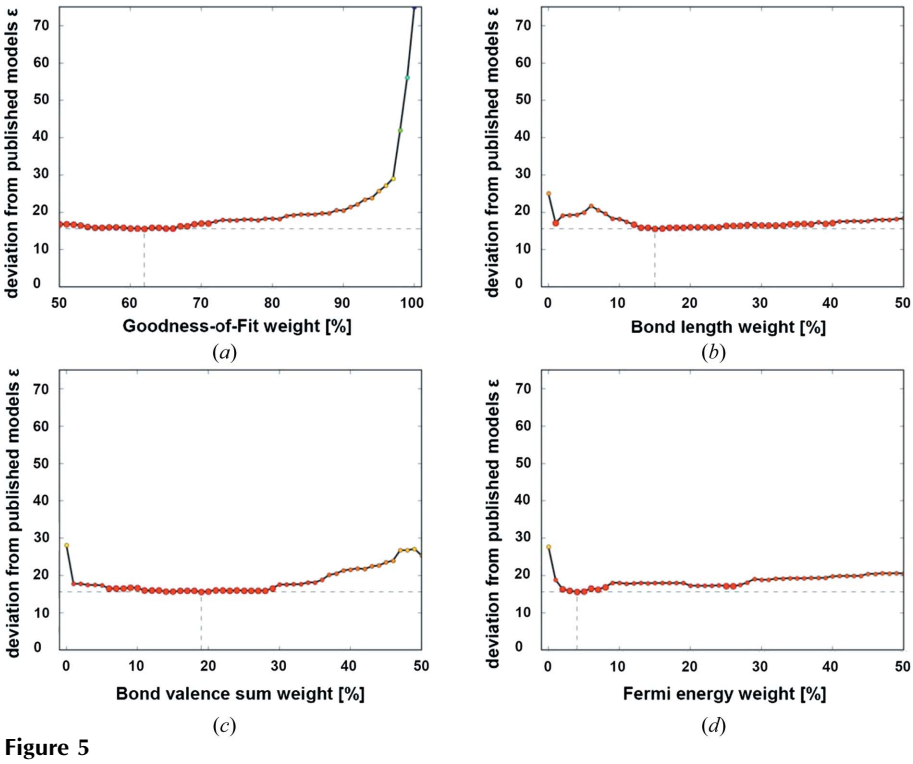

Figure 5
Estimation of optimal weights for Fe. The sum of deviations from published models ε [equation (6)] for each combination of criterion-weights is plotted over one of the four weights. The lowest scores for each value of the criterion-weights are connected by a line. The combinations yielding ε not larger than +5% of the overall minimum are enlarged, indicating the dimension of the minimum with respect to the corresponding criterion-weight.

sulfur-coordination is correctly calculated in the meta-analysis. The total coordination of light ligands is never off by more than 0.5. This allows us to conclude that the determination of structural binding motifs for the model datasets for Zn is carried out with reasonably good accuracy (Figs. S2–S9, Tables S1–S8 of supplementary material).

The meta-analysis provides error bars for the individual coordination numbers. These error margins depend on the meta-analysis threshold defined in §2.3. For the reference datasets in seven out of eight cases the published data are within 1σ margins of the meta-analysis for the S-coordination numbers and all are within 2σ margins. The number of light ligands is consistent for six out of eight models. Therefore we note that the errors provided by *ABRA*'s meta-analysis are typically meaningful, and the meta-analysis threshold of 2.5% is well defined.

3.2. Iron data

The optimization of weights for Fe has been carried out using datasets of an iron–sulfur cluster, *Sulfolobus solfataricus* ABCE1, *Ss* ABCE1 (Barthelme *et al.*, 2007), two different oxidation states of rubredoxin, *Pa* RM2-4$_{ox}$ and *Pa* RM2-4$_{red}$ from *Pyrococcus abyssi* (Wegner *et al.*, 2004), and *homo sapiens* tyrosine hydroxyalase in native and reduced state, *hs* TH1$_{nat}$ [oxidation state Fe(II)] and *hs* TH1$_{ox}$ [oxidized with H_2O_2 to Fe(III)] (Meyer-Klaucke *et al.*, 1996), as well as ferric pyoverdine (Fe-Pvd) (Wirth *et al.*, 2007). The results using optimal weights are given in Table 2. All published Fe-EXAFS data were refined without defining the oxidation state of the sample. This makes the analysis of the refinements considerably more difficult than for Zn-EXAFS.

The dataset of the Fe–S cluster was refined and analyzed using *ABRA*'s Fe–S database. The only difference of this and the default database is a different ideal Fe–S distance. The S coordination from *ABRA*'s top-ranking models deviates only by 0.5 in one case; in all other cases the determination is right on the spot. The His- and O-coordination numbers differ in three cases, but the overall number of light ligands is typically correct. The oxidation states are always determined correctly.

The results from the meta-analysis for the number of S and light ligands are within ± 0.5 of the published results. The oxidation states were correctly determined in four out of six cases unanimously (meta-analysis error of 0), with a clear tendency in the fifth case (2.8 ± 0.4) and indications for mixed valences in the sixth case (2.5 ± 0.5). Thus, *ABRA* determines structural binding motifs for Fe proteins (Fig. S10–S15, Tables S9–S14) and successfully estimates the oxidation state. The meta-analysis results for coordination numbers of both sulfur and light ligands are always within 1σ margins. In summary, the error margins provided by the meta-analysis of *ABRA* present a good estimate.

3.3. Further applications

In order to check the performance of the program, two examples, taken from the literature, were re-analyzed: *Saccharomyces cerevisiae* GAL4 (GAL4) initially interpreted as a S_3O_1-motif (Povey *et al.*, 1990), where later the binding site was identified as a tetrathiolate (S_4) (Clark-Baldwin *et al.*, 1998); conversely, the S_3O_1-site in bovine liver aminolevulinate dehydratase (ALAD) was based on difference spectra modelled as a tetrathiolate (Dent *et al.*, 1990). Several reasons might have led to these results that did not stand the test of time, *e.g.* sample-related problems (wrong Zn stoichiometry and/or presence of metal chelators), limitations in theory used at that time, or an error in data analysis. While *ABRA* surely cannot substitute a detailed characterization of the sample [especially a quantification of the metal content (Garman & Grime, 2005; Luz *et al.*, 2005; Wellenreuther *et al.*, 2008)], it should assist the users in data interpretation, avoiding bias and oversight. Both datasets were extracted from the publications by *DataThief III* (Tummers, 2006). The energy axis had to be rescaled to assure consistency by −20 and −10 eV for ALAD

Table 3
Comparison of *ABRA*'s results for different structural models for the ALAD dataset.

These data were initially published as a tetrathiolate (Dent *et al.*, 1990). A revised model was published later (Clark-Baldwin *et al.*, 1998). The model proposed by *ABRA* as a result of the meta-analysis suggests the presence of more than one light ligand. ΔEF = Fermi energy shift.

Model	*N*	*M*—*L*	*R* (Å)	$2\sigma^2$ (Å^2)	ΔEF (eV)	*ABRA* score
Top-ranking model $S_{2.0}O_{1.5}$	2.0	Zn—S	2.277 (4)	0.0045 (8)	−8.7 (8)	69.49%
	1.5	Zn—O	2.03 (1)			
Meta-analysis result $S_{2.1}O_{1.6}$	2.1 (2)	Zn—S	2.27 (1)	0.0054 (8)	−7.8 (8)	69.49–67.80% (7 models)
	1.6 (2)	Zn—(O,His)	2.03 (1)	0.0054 (8)		
Revised model	3.0	Zn—S	2.261 (4)	0.0087 (7)	−5.7 (8)	65.33%
	1.0	Zn—O	2.01 (2)			
Initial model	4.0	Zn—S	2.244 (4)	0.0070 (8)	−2.7 (7)	50.28%

Table 4
Comparison of *ABRA*'s results for the GAL4 dataset which was initially published as a S_3O_1 (Povey *et al.*, 1990).

The crystal structure suggested the presence of four sulfur ligands; this is not consistent with the EXAFS data. ΔEF = Fermi energy shift.

Model	*N*	*M*—*L*	*R* (Å)	$2\sigma^2$ (Å^2)	ΔEF (eV)	*ABRA* score
Top-ranking model $S_{2.5}O_{1.0}$	2.5	Zn—S	2.300 (3)	0.0096 (4)	−6.3 (5)	92.18%
	1.0	Zn—O	1.984 (9)			
Meta-analysis result $S_{2.7}O_{1.0}$	2.8 (3)	Zn—S	2.30 (1)	0.0110 (4)	−6.0 (5)	92.18–89.90% (12 models)
	1.0 (5)	Zn—(O,His)	1.99 (1)	0.0106 (6)		
Initial model	3.0	Zn—S	2.298 (3)	0.0112 (4)	−5.6 (4)	91.93%
	1.0	Zn—O	1.97 (1)			
Model based on crystal structure	4.0	Zn—S	2.295 (3)	0.01429 (1)	−5.5 (4)	88.72%

and GAL4, respectively. Afterwards the datasets were binned with $\Delta k = 0.05$ Å^{-1} (Fig. S16, Table S15).

For ALAD, published by Dent *et al.* (1990) as a tetrathiolate, *ABRA*'s analysis gives a top-scoring model of $S_{2.0}O_{1.5}$ with a total score of 69.5% (see Table 3). The score is significantly below 100% owing to the substantial residual at high *k*-values (see Fig. S17, Table S16). *ABRA*'s meta-analysis identifies 2.1 (2) sulfur and 1.6 (2) light ligands. The tetrathiolate model achieved a total score of 50.3%; its rejection was based on the very short Zn—S distance of only 2.244 (4) Å, deviating from the target distances. On the contrary, the later favoured S_3O_1 model achieved a total score of 65.3% and is therefore included in the meta-analysis. Despite the limited data range (Δk = 3–12 Å^{-1}), *ABRA* excludes a tetrathiolate binding motif for ALAD, and instead suggests a mixed coordination with S and O, in agreement with models published before (Hasnain *et al.*, 1985) and later based on XAS (Clark-Baldwin *et al.*, 1998) or protein crystallography (Erskine *et al.*, 1999).

ABRA's refinement of GAL4 yields $S_{2.5}O_{1.0}$ as the top-scoring model with a total score of 92.2% (see Table 4). The meta-analysis results in an average best model of 2.8 (3) S and 1.0 (5) light ligands calculated from the 12 top-scoring models. Both findings support the initial S_3O_1 model. The S_3O_1 model is included in the 12 top-scoring models, ranking at #3 with a total score of 91.9%. Such small separations in the score are a good example of where the benefits of *ABRA*'s meta-analysis come into play. The S_4 model is rejected owing to a score of 88.7%, caused by a decrease of the goodness-of-fit and the BVS criterion mainly. In conclusion, we can confirm the interpretation of the data published by Povey *et al.* (1990), which might suggest that the data do not resemble that of the GAL4 protein properly (Marmorstein *et al.*, 1992).

4. Discussion

We have shown that the algorithm implemented in the *ABRA* software is able to automatically refine and analyze biological EXAFS data in the cases of Zn and Fe X-ray absorption spectra. The predicted model is typically precise and the errors given by the meta-analysis provide a good estimate for the uncertainties of the coordination numbers. Normally the sulfur-coordination is determined within ±0.5 ligands, whereas the ligands O and histidine are determined with a lower reliability, owing to their comparable scattering properties and similar bond distances. In the initial refinement it is therefore better to group the lighter ligands in the analysis. The determination of contributions from light and sulfur ligands is comparable. The present implementation reflects a refinement strategy followed by many scientists, but leaves out multiple scattering *via* the central absorber atom, for which so far significant contributions have only been identified for a limited number of metalloproteins (*e.g.* Ha *et al.*, 2007; Hollenstein *et al.*, 2009). The software can help to properly model XAS data in an objective and automated manner. For ALAD, *ABRA* has indicated the correct binding motif, and strongly discouraged the tetrathiolate model. For GAL4, the initial interpretation of the spectra is supported, which might either indicate a problem with the sample or suggest different binding pattern under different conditions (Kraulis *et al.*, 1992). In both cases, the automated analysis clearly differentiated between S_4 and S_3O_1 models, even for sub-optimal datasets. The analysis, so far utilized in a number of cases, always provided a conservative estimate for the models consistent with the data, and served as a solid basis for additional interactive refinements that were required for multi-

nuclear proteins (Wellenreuther *et al.*, 2009; Peroza *et al.*, 2009).

In summary, automatic refinement of XAS data is possible on the basis of scoring algorithms. The introduction of criteria in addition to the least-squares refinement results in a more reliable identification of the metal binding motif as highlighted in Figs. 3(*a*) and 4(*a*). In future, a fine tuning of the criterion-weights for each absorber element might be included. Here, the bottleneck is (*a*) the comprehensive set of adequate reference datasets, (*b*) the accurate spin- and oxidation state-dependent database for individual metal ligand distances, and (*c*) the demands for computing-power during the analysis of all reference datasets. Owing to the fact that the criterion-weights for Zn- and Fe-EXAFS analysis differ only slightly, we are at present using an intermediate set for different 3*d* elements, *e.g.* Mn, Co and Cu. In the present version the joint refinement of Debye–Waller factors for ligands at similar distances to the absorber atom ensures that no over-interpretation of data takes place. As soon as predefined relative Debye–Waller factors become available for an increasing number of metal binding sites in proteins (Dimakis & Bunker, 2004), the accuracy of automated EXAFS data refinement will further improve. The meta-analysis has provided, so far, a set of models among which we could identify our best model. This interactive step seems to be required in cases where the conservative assumptions are not fully justified and are typically indicated by the absence of a satisfying best model. In other cases, the expert knowledge already incorporated in *ABRA* allowed us to directly use its top model after an interactive cross-check. Therefore we are very optimistic that *ABRA* helps non-XAS experts to refine and analyze their own datasets and in addition might serve as a quality control for XAS data analysis.

Currently, *ABRA* is running on the computer cluster of EMBL Hamburg (http://cluster.embl-hamburg.de/exafs/exafs_new.html); the access is free. Other versions based on data-analysis packages such as *Feffit/IFeffit* (Newville, 2001) could be easily implemented along these lines (Ravel & Newville, 2009). Moreover, this algorithm can be applied to other techniques and thus help further automating data analysis.

We gratefully acknowledge funding of the BIOXHIT project by the European Commission under FP6 contract LSHG-CT-2003-503420 and fruitful discussions with Marjorie M. Harding (University Edinburgh, UK). We also thank Gianpietro Previtali from the computer group of the EMBL Hamburg Outstation for supporting the computer infrastructure.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Banci, L., Bertini, I., Ciofi-Baffoni, S., Katsari, E., Katsaros, N., Kubicek, K. & Mangani, S. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 3994–3999.
Barthelme, D., Scheele, U., Dinkelaker, S., Janoschka, A., MacMillan, F., Albers, S.-V., Driessen, A. J. M., Salomone-Stagni, M., Bill, E., Meyer-Klaucke, W., Schünemann, V. & Tampé, R. (2007). *J. Biol. Chem.* **282**, 14598–14607.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucl. Acids Res.* **28**, 235–242.
Binsted, N. (1998). *EXCURV98.* CCLRC Daresbury Laboratory, Warrington, UK.
Binsted, N. & Hasnain, S. S. (1996). *J. Synchrotron Rad.* **3**, 185–196.
Binsted, N., Strange, R. W. & Hasnain, S. S. (1992). *Biochemistry*, **31**, 12117–12125.
Boyault, C., Gilquin, B., Zhang, Y., Rybin, V., Garman, E., Meyer-Klaucke, W., Matthias, P., Müller, C. W. & Khochbin, S. (2006). *EMBO J.* **25**, 3357–3366.
Brese, N. E. & O'Keeffe, M. (1991). *Acta Cryst.* B**47**, 192–197.
Brown, I. D. & Altermatt, D. (1985). *Acta Cryst.* B**41**, 244–247.
Bunker, G., Dimakis, N. & Khelashvili, G. (2005). *J. Synchrotron Rad.* **12**, 53–56.
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.
Cohen, S. X., Moulin, M., Schilling, O., Meyer-Klaucke, W., Schreiber, J., Wegner, M. & Müller, C. W. (2002). *FEBS Lett.* **528**, 95–100.
D'Angelo, P., Barone, V., Chillemi, G., Sanna, N., Meyer-Klaucke, W. & Pavel, N. V. (2002). *J. Am. Chem. Soc.* **124**, 1958–1967.
Dent, A. J., Beyersmann, D., Block, C. & Hasnain, S. S. (1990). *Biochemistry*, **29**, 7822–7828.
Dimakis, N., Al-Akhras, M. & Bunker, G. (1999). *J. Synchrotron Rad.* **6**, 266–267.
Dimakis, N. & Bunker, G. (1998). *Phys. Rev. B*, **58**, 2467–2475.
Dimakis, N. & Bunker, G. (2004). *Phys. Rev. B*, **70**, 195114.
Dimakis, N. & Bunker, G. (2005). *Phys. Scr.* T**115**, 175–178.
Dimakis, N. & Bunker, G. (2006). *Biophys. J.* **91**, L87–L89.
Dimakis, N., Farooqi, M. J., Garza, E. S. & Bunker, G. (2008). *J. Chem. Phys.* **128**, 115104.
Duda, D., Govindasamy, L., Agbandje-McKenna, M., Tu, C., Silverman, D. N. & McKenna, R. (2003). *Acta Cryst.* D**59**, 93–104.
Erskine, P. T., Norton, E., Cooper, J. B., Lambert, R., Coker, A., Lewis, G., Spencer, P., Sarwar, M., Wood, S. P., Warren, M. J. & Shoolingin-Jordan, P. M. (1999). *Biochemistry*, **38**, 4266–4276.
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.
Filipponi, A. & Di Cicco, A. (1995). *Phys. Rev. B*, **52**, 15135–15149.
Filipponi, A., Di Cicco, A. & Natoli, C. R. (1995). *Phys. Rev. B*, **52**, 15122–15134.
Garman, E. F. & Grime, G. W. (2005). *Prog. Biophys. Mol. Biol.* **89**, 173–205.
George, G. N., Hilton, J., Temple, C., Prince, R. C. & Rajagopalan, K. V. (1999). *J. Am. Chem. Soc.* **121**, 1256–1266.
George, G. N. & Pickering, I. J. (1993). *EXAFSPAK: A Suite of Computer Programs for Analysis of X-ray Absorption Spectra.* Stanford Synchrotron Radiation Laboratory, Stanford, CA, USA.
Gurman, S. J., Binsted, N. & Ross, I. (1984). *J. Phys. C*, **17**, 143–151.
Gurman, S. J., Binsted, N. & Ross, I. (1986). *J. Phys. C*, **19**, 1845–1861.
Ha, S. W., Korbas, M., Klepsch, M., Meyer-Klaucke, W., Meyer, O. & Svetlitchnyi, V. (2007). *J. Biol. Chem.* **282**, 10639–10646.
Harding, M. M. (1999). *Acta Cryst.* D**55**, 1432–1443.
Harding, M. M. (2000). *Acta Cryst.* D**56**, 857–867.
Harding, M. M. (2001). *Acta Cryst.* D**57**, 401–411.
Harding, M. M. (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.
Harris, H. H., Pickering, I. J. & George, G. N. (2003). *Science*, **301**, 1203.
Hasnain, S. S., Wardell, E. M., Garner, C. D., Schlosser, M. & Beyersmann, D. (1985). *Biochem. J.* **230**, 625–633.
Haumann, M., Grundmeier, A., Zaharieva, I. & Dau, H. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 17384–17389.

Hollenstein, K., Comellas-Bigler, M., Bevers, L. E., Feiters, M. C., Meyer-Klaucke, W., Hagedoorn, P. L. & Locher, K. P. (2009). *J. Biol. Inorg. Chem.* **14**, 663–672.
Hwang, J., Krebs, C., Huynh, B. H., Edmondson, D. E., Theil, E. C. & Penner-Hahn, J. E. (2000). *Science*, **287**, 122–125.
Klementev, K. (2001). *J. Phys. D*, **34**, 209–217.
Korbas, M., Marsa, D. F. & Meyer-Klaucke, W. (2006). *Rev. Sci. Instrum.* **77**, 063105.
Korbas, M., Vogt, S., Meyer-Klaucke, W., Bill, E., Lyon, E. J., Thauer, R. K. & Shima, S. (2006). *J. Biol. Chem.* **281**, 30804–30813.
Kraulis, P. J., Raine, A. R., Gadhavi, P. L. & Laue, E. D. (1992). *Nature (London)*, **356**, 448–450.
Küpper, F. C. *et al.* (2008). *Proc. Natl Acad. Sci. USA*, **105**, 6954–6958.
Labhardt, A. & Yuen, C. (1979). *Nature (London)*, **277**, 150–151.
Liu, T., Ramesh, A., Ma, Z., Ward, S. K., Zhang, L., George, G. N., Talaat, A. M., Sacchettini, J. C. & Giedroc, D. P. (2007). *Nat. Chem. Biol.* **3**, 60–68.
Liu, W. & Thorp, H. H. (1993). *Inorg. Chem.* **32**, 4102–4105.
Luz, H. N. d., Spemann, D., Meyer-Klaucke, W. & Tröger, W. (2005). *Nucl. Instrum. Methods Phys. Res. B*, **231**, 308–314.
Lytle, F. W., Sayers, D. E. & Stern, E. A. (1989). *Physica B*, **158**, 701–722.
Maniasetty, B. A., Turnbull, A. P., Panjikar, S., Bussow, K. & Chance, M. R. (2008). *Proteomics*, **8**, 612–625.
Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S. C. (1992). *Nature (London)*, **356**, 408–414.
Masip, L., Pan, J. L., Haldar, S., Penner-Hahn, J. E., DeLisa, M. P., Georgiou, G., Bardwell, J. C. & Collet, J. F. (2004). *Science*, **303**, 1185–1189.
Meyer-Klaucke, W., Winkler, H., Schünemann, V., Trautwein, A. X., Nolting, H. F. & Haavik, J. (1996). *Eur. J. Biochem.* **241**, 432–439.
Möbius, A. F. (1827). *Der Barycentrische Calcül.* Leipzig: Verlag von Johann Ambrosius Barth.
Newville, M. (2001). *J. Synchrotron Rad.* **8**, 322–324.
Newville, M. (2005). *Phys. Scr.* T**115**, 159–161.
Ohlenschläger, O., Seiboth, T., Zengerling, H., Briese, L., Marchanka, A., Ramachandran, R., Baum, M., Korbas, M., Meyer-Klaucke, W., Durst, M. & Görlach, M. (2006). *Oncogene*, **25**, 5953–5959.
O'Keeffe, M. & Brese, N. E. (1992). *Acta Cryst.* B**48**, 152–154.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Paulsen, H., Grünsteudel, H., Meyer-Klaucke, W., Gerdan, M., Grünsteudel, H. F., Chumakov, A. I., Rüffer, R., Winkler, H., Toftlund, H. & Trautwein, A. X. (2001). *Eur. Phys. J. B*, **23**, 463–472.
Paul-Soto, R., Bauer, R., Frere, J.-M., Galleni, M., Meyer-Klaucke, W., Nolting, H., Rossolini, G. M., de Seny, D., Hernandez-Valladares, M., Zeppezauer, M. & Adolph, H.-W. (1999). *J. Biol. Chem.* **274**, 13242–13249.
Peroza, E. A., Al Kaabi, A., Meyer-Klaucke, W., Wellenreuther, G. & Freisinger, E. (2009). *J. Inorg. Biochem.* **103**, 342–353.
Philpotts, A. (1990). *Principles of Igneous and Metamorphic Petrology*, 1st ed. Englewood Cliffs: Prentice Hall.
Povey, J. F., Diakun, G. P., Garner, C. D., Wilson, S. P. & Laue, E. D. (1990). *FEBS Lett.* **266**, 142–146.
Pufahl, R. A., Singer, C. P., Peariso, K. L., Lin, S. J., Schmidt, P. J., Fahrni, C. J., Culotta, V. C., Penner-Hahn, J. E. & O'Halloran, T. V. (1997). *Science*, **278**, 853–856.
Ravel, B. & Newville, M. (2005). *J. Synchrotron Rad.* **12**, 537–541.
Ravel, B. & Newville, M. (2009). Private communication.
Ressler, T. (1998). *J. Synchrotron Rad.* **5**, 118–122.
Sayers, D. E. (2000). *Report of the International XAFS Society Standards and Criteria Committee*, http://www.i-x-s.org/OLD/subcommittee_reports/sc/SC00report.pdf.
Schilling, O., Wenzel, N., Naylor, M., Vogel, A., Crowder, M., Makaroff, C. & Meyer-Klaucke, W. (2003). *Biochemistry*, **42**, 11777–11786.
Shima, S., Pilak, O., Vogt, S., Schick, M., Salomone-Stagni, M., Meyer-Klaucke, W., Warkentin, E., Thauer, R. K. & Ermler, U. (2008). *Science*, **321**, 572–575.
Teo, B. & Joy, D. (1981). *EXAFS Spectroscopy: Techniques and Applications.* New York: Plenum.
Thompson, A., Attwood, D., Gullikson, E., Howells, M., Kortright, J., Robinson, A., Underwood, J., Kim, K., Kirz, J., Williams, G. & Scofield, J. (2001). *X-ray Data Booklet*, 2nd ed. Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
Thorp, H. H. (1992). *Inorg. Chem.* **31**, 1585–1588.
Thorp, H. H. (1998). *Inorg. Chem.* **37**, 5690–5692.
Tomic, S., Searle, B. G., Wander, A., Harrison, N. M., Dent, A. J., Mosselmans, J. F. W. & Inglesfield, J. E. (2004). CCLRC Technical Report DL-TR-2005-001, ISSN 1362-0207.
Tummers, B. (2006). *DataThief III*, http://www.datathief.org/.
Vogel, A., Schilling, O. & Meyer-Klaucke, W. (2004). *Biochemistry*, **43**, 10379–10386.
Wegner, P., Bever, M., Schünemann, V., Trautwein, A. X., Schmidt, C., Bonisch, H., Gnida, M. & Meyer-Klaucke, W. (2004). *Hyperfine Interact.* **156**, 293–298.
Wellenreuther, G., Cianci, M., Tucoulou, R., Meyer-Klaucke, W. & Haase, H. (2009). *Biochem. Biophys. Res. Commun.* **380**, 198–203.
Wellenreuther, G., Fittschen, U. E. A., Achard, M. E. S., Faust, A., Kreplin, X. & Meyer-Klaucke, W. (2008). *Spectrochim. Acta B*, **321**, 572–575.
Wellenreuther, G. & Meyer-Klaucke, W. (2007). *AIP Conf. Proc.* **882**, 322–324.
Wellenreuther, G. & Meyer-Klaucke, W. (2009). *J. Phys. Conf. Ser.* **190**, 012033.
Westre, T. E., Di Cicco, A., Filipponi, A., Natoli, C. R., Hedman, B., Solomon, E. I. & Hodgson, K. O. (1995). *J. Am. Chem. Soc.* **117**, 1566–1583.
Wirth, C., Meyer-Klaucke, W., Pattus, F. & Cobessi, D. (2007). *J. Mol. Biol.* **368**, 398–406.
Wolter, T., Meyer-Klaucke, W., Müther, M., Mandon, D., Winkler, H., Trautwein, A. X. & Weiss, R. (2000). *J. Inorg. Biochem.* **78**, 117–122.
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). *Nat. Chem. Biol.* **4**, 107–109.