MXAN: a new software procedure to perform geometrical fitting of experimental XANES spectra

M. Benfatto[1], A. Congiu-Castellano[2,4], A. Daniele[2] and S. Della Longa[3,4]

[1]Lab. Naz. Frascati – INFN – P.O. Box 13-00044 Frascati (Italy), [2]Dip. Fisica, Univ. "La Sapienza", Pza A. Moro – 00185 - Roma (Italy), [3]Dip. Medicina Sperimentale, Univ. L'Aquila, via Vetoio – 67100 – L' Aquila (Italy), [4]INFM – Un. Roma I, Pza A. Moro – 00185 - Roma (Italy)

A new software procedure, MXAN, to fit experimental XANES spectra is presented here. The method is based on the comparison between the experimental spectrum and several theoretical calculations generated by changing the relevant geometrical parameter of the site around the absorbing atom. The x-ray photoabsorption cross section is calculated using the general multiple-scattering scheme, utilizing a complex Hedin-Lunqvist energy-dependent potential to describe the exchange correlation interaction. Our method has been applied to the study of geometrical environment of the tetrahedral zinc site of the protein superoxide dismutase (SOD). The experimental Zn K-edge XANES spectrum has been fitted in the space of the first shell coordination parameters (ligand distances and angles) following the behavior of the chi-square as a function of the local distortion from the starting crystallographic structure. The recovered structure is found to be independent on the starting conditions, showing the theoretical uniqueness of the structural solution. Strengths and limitations of the application to real systems are also discussed.

Keywords: X-ray Absorption Spectroscopy, Fitting procedure, superoxide dismutase

1 Introduction

X-ray Absorption Near Edge Structure (XANES) spectroscopy is a powerful method to study the local structure around an absorbing center of various types of matter. Analysis of the spectrum can give both electronic and structural information on the site around an absorbing atom. The fit of the XANES part of the spectrum, i.e. from the rising edge up to few hundreds of eV, in terms of the geometrical structure, has long been an aim of users of this technique, especially in the cases of experimental data with limited k-range where a standard EXAFS (Extended X-ray Absorption Fine Structure) analysis cannot be easily done.

However, the quantitative analysis of the XANES spectra present some difficulties mainly linked to the theoretical approximation in the treatment of the potential and the need for heavy time-consuming algorithms to calculate the absorbing cross section in the framework of full multiple scattering approach.

In this paper we present a software package, named MXAN, to calculate the absorption cross section with the aim to obtain a reliable fit of the experimental data in term of a well defined set of structural parameters. For brevity, we just report here the analysis of the Zn K-edge in the superoxide dismutase protein. We have decided to present this case because of the interest of XANES applications in biology. Other test cases will be presented in a future longer paper.

2 Results and discussion

The method is based on the comparison between the experimental spectrum and several theoretical calculations performed by varying selected structural parameters starting from a well defined initial geometrical configuration around the absorber. The x-ray photoabsorption cross section is calculated using the general multiple-scattering scheme, utilizing a complex Hedin-Lunqvist energy-dependent potential to describe the exchange correlation interaction (Tyson et al. 1992). The set of calculations is done in reasonable time and the optimization in the space of the parameters is achieved by the MINUIT package of the CERN library. The MINUIT package is used to minimize the error function

$$S^2 = n \sum_{i=1,m} w_i [(y^{th}_i - y^{exp}_i) \varepsilon_i^{-1}]^2 / \sum_{i=1,m} w_i$$

where n is the number of independent parameters in the fit, m the number of data points fitted, y^{th}_i and y^{exp}_i the theoretical and experimental values of the XANES spectrum, and ε_i the individual errors in the experimental data set; w_i is a statistical weight. For w_i=constant=1, S^2 becomes the statistical chi-square function χ^2. The solution is found to be independent on the starting conformation, no limitation in the energy range and polarization conditions are present.

Superoxide dismutase (SOD) is a dimeric enzyme. The active site of the bovine SOD contains, in each monomer, one Zn^{2+} ion and one atom of cuprous or cupric Cu ion. The two ions are bridged, in the oxidized state of Cu, by a histidine residue and are distant about 6 Å. The coordination of the Zn atom is a distorted tetrahedron, with ligands formed by the imidazole nitrogen of the bridging histidine, two other imidazole nitrogens, and an oxygen from a residue of aspartate. Data used for the starting configuration are taken from Protein Data Bank code 1SXN. A pictorial view of the Zn site in SOD is shown in Fig. 1, where some of the typical parameters used in the minimization procedure are depicted.

The polar geometrical coordinates of the first ligands are used in this case as the parameters to be optimized by MXAN. To test the right behaviour of the computational procedure and the sensitivity of the method to probe structural parameters we have performed the following theoretical test:

i) The XANES spectrum of the crystal structure of the protein (Zn site of monomer A, using a 20 atom cluster and no correction of core-hole life time) is calculated, and a gaussian noise (chosen by us as equal to 1% of the edge jump) is added. The spectrum is reported as dotted line in Fig.2. It mimics the experimental data to be fitted (it corresponds to structure CS in Table 1).

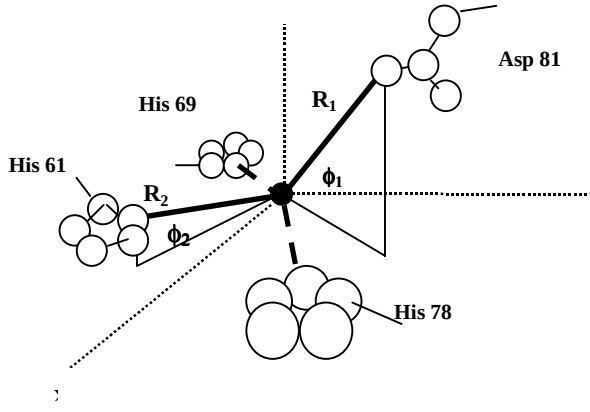

Figure 1 Pictorial view of the geometrical parameters used in the fitting procedure

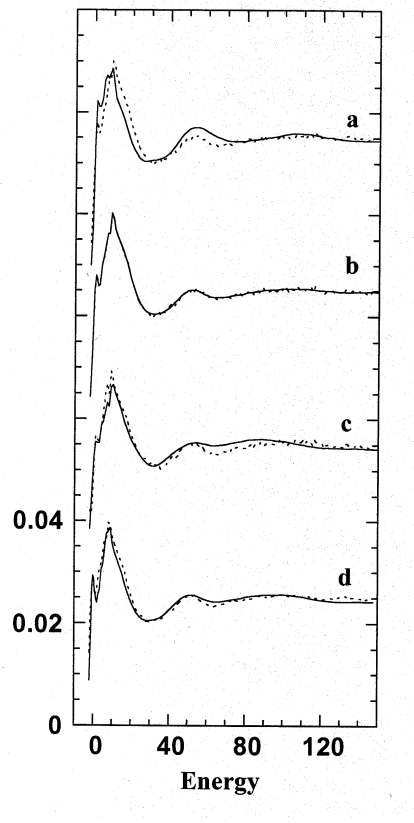

Figure 2 Different theoretical results obtained by changing the fitting strategy. See the main test for the explanation. Dotted line represents the " experimental" data. Solid lines are calculations obtained as best fit of the experimental data

ii) A distortion in the distances and angles of the first shell is applied: in this case the distances d(Zn-O_asp81) and d(Zn-N_his78) go from 1.88 to 2.03 Å and from 1.92 to 2.07 Å respectively; the ϕ polar angles of these two ligands are increased by 15°. The protein atoms covalently linked to these ligands rigidly follow their motions. The other two ligand distances are kept fixed. A sub-space of 4 parameters seems a reasonable choice for the number of parameters (two distances and two angles) that can be investigated in real cases, and the need to safe computer time. A more detailed investigation regarding the number of independent parameters that can be used in real cases is a very debated problem in MS analysis and will be the subject of a forthcoming paper. The calculated spectrum (corresponding to structure DS of table 1) is reported in Fig. 2(a) as solid line.

iii) By using the MXAN procedure we try to recover the CS structure, finding best fit conditions with different errors that depend on the chosen strategy. The results are reported in Fig. 2(b,c,d) and the corresponding structure in Table 1.

As a first strategy, we do not use any statistical weight w_i, and the potential is recalculated at every step. The undistorted structure CS is completely recovered by MXAN (solid line of Fig 2(b)). The statistical error corresponding to the noise introduced is about 0.03 Å for the ligand distances, and of about 4° for the angles. As a second strategy, the potential is kept fixed to that of the DS structure. In such a case, it is impossible to reproduce satisfactorily the spectrum CS. The chi-square increases enormously and systematic errors appear in the evaluation of the structural parameter (of about 0.1 Å for the distances, and of about 10° for the angles). The best fit is the solid line of Fig 2(c). As a third strategy, we have introduced a statistical weight w_i in the fitting procedure, we have chosen a suitable arctangent function center at 30 eV and having a width of 10 eV. In this way, we intend to minimize the potential errors which are relevant in the low energy part of the spectrum. The original structure CS is satisfactorily well recovered, and the best fit is reported as the solid line of Fig 2(d). A summary of the fitting results is reported in table 1.

Table 1

structure	S^2	R_1 (Å)	R_2(Å)	$\Delta\phi_1$(°)	$\Delta\phi_2$(°)
CS	-	1.88	1.92	0.0	0.0
DS	-	2.03	2.07	+15	+15
b	3.2	1.88(2)	1.92(4)	0(4)	0(3)
c	53.8	1.81(2)	2.03(2)	-15(4)	-8(4)
d	12.0	1.84(3)	1.95(4)	-8(5)	2(4)

Finally the MXAN procedure has been applied to the experimental Zn K-edge spectrum of SOD (reported as dotted line in Fig 3). The experimetal data are taken from ref.2 (Murphy et al. 1997). The calculation done using the crystal structure of the Zn site of monomer 1 of the protein is shown as solid line in Fig. 3(a).

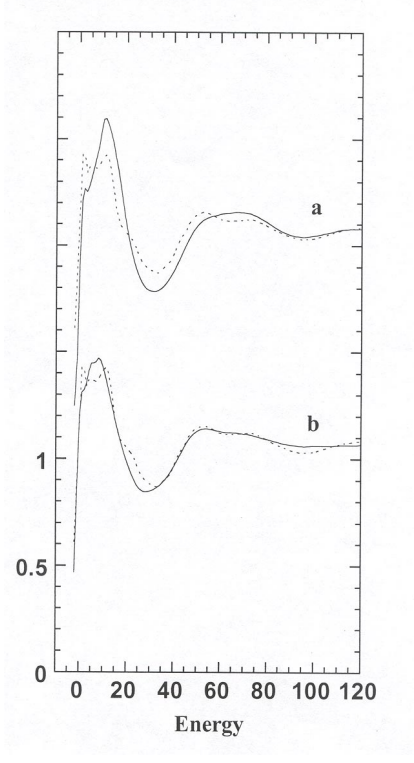

Figure 3 Fitting results of the SOD system at the Zn k-edge. The dotted lines are the experimental data while the solid lines are the theoretical calculations. The calculation obtained by using the crystal structure is reported as solid line in part a. Part b contains the best fit results.

A relevant discrepancy is present between the experimental and theoretical data. However, it should be noted that the diffraction data are given at a resolution of 1.9 Å, r-value = 0.166, corresponding to a precision in the determination of positions of about 0.2 Å (Cruickshank, 1999). A Montecarlo search has been initially done on the nine parameters (4 distances, and 5 angles) that define the first shell coordination geometry in order to recover the experimental spectrum. Only the first ligand distances were finally refined to a minimum. The same arctangent statistical weight has been used here as in the theoretical test, and the potential has been recalculated at every step. The best fit is shown as solid line in Fig. 3(b). As seen from the figure, the best agreement is obtained in the high energy part of the spectrum. This is not surprising because of the use of a weight in the fitting. We note that our procedure does not use the coordination number and Debye-Waller factors as fitting parameters. The final values found by MXAN are fully compatible with the average values of the crystal distances of the Zn sites in monomer 1 and monomer 2, in particular d(Zn-O_asp81)=1.89(3) Å; d(Zn-N_his61)=1.99(5) Å; d(Zn-N_his78)=2.18(5) Å; d(Zn-N_his69)=2.29(6) Å face to 1.9 Å, 1.9 Å, 2.1 Å, 2.1 Å (x-ray structure monomer 1), and 1.7Å, 2.0Å, 2.1Å, 2.3Å (x-ray structure monomer 2) respectively. The values in parenthesis represent the approximate error on the last digit, evaluated by the routine GRADIENT of MINUIT. A recent EXAFS study has determined a distance d(Zn-O)=1.96 and an average distance d(Zn-N) = 2.01 (Murphy et al. 1997).

Entering a bad guess to the procedure produces a best fits far from being satisfactory. For example, by substituting the Zn site structure of SOD with that of the ASV integrase core domain (PDB code 1VSH), containing a tetrahedral Zn site legated to 1 oxygen at 2.0 Å and 3 other oxygen atoms at 2.3 Å results in a very bad "best" fit.

In summary, this paper presents the first attempt to obtain quantitative results from XANES spectroscopy when applied to low symmetry clusters, such as those of biological interest. A precision comparable with EXAFS spectroscopy (about 2%) can be obtained in principle, especially when weighting functions are used to minimize potential errors. The method has been applied to the real case of the protein SOD, obtaining results compatible with crystal and EXAFS structural determination.

References

Cruickshank, D.W.J. (1999) Acta Cryst. **D55**, 583-601.
Murphy, L.M., Strange, R.W. and Hasnain, S.S. (1997) J. Phys. IV, Colloque **C2**, 599-602.
Tyson, T.A., Hodgson, K.O., Natoli, C.R. and Benfatto, M. (1992) Phys. Rev. **B46**, 5997-6019 and references therein.

 Received 31 July 2000 • Accepted 25 October 2000