Ab Initio Self-Consistent X-Ray Absorption Fine Structure Analysis for Metalloproteins

Nicholas Dimakis* and Grant Bunker†
*University of Texas-Pan American, Edinburg, Texas; and †Department of Biological Chemical and Physical Sciences, Illinois Institute of Technology, Chicago, Illinois

ABSTRACT X-ray absorption fine structure is a powerful tool for probing the structures of metals in proteins in both crystalline and noncrystalline environments. Until recently, a fundamental problem in biological XAFS has been that ad hoc assumptions must be made concerning the vibrational properties of the amino acid residues that are coordinated to the metal to fit the data. Here, an automatic procedure for accurate structural determination of active sites of metalloproteins is presented. It is based on direct multiple-scattering simulation of experimental X-ray absorption fine structure spectra combining electron multiple scattering calculations with density functional theory calculations of vibrational modes of amino acid residues and the genetic algorithm differential evolution to determine a global minimum in the space of fitting parameters. Structure determination of the metalloprotein active site is obtained through a self-consistent iterative procedure with only minimal initial information.

Received for publication 8 June 2006 and in final form 11 September 2006.

Address reprint requests and inquiries to Nicholas Dimakis, Tel.: 956-380-8761; Fax: 956-381-2423; E-mail:dimakis@utpa.edu.

Metalloproteins constitute ~30% of all total known proteins in nature (1). Their active sites typically consist of a metal ion center that is coordinated to various amino acid residues. Structural information of metalloproteins is probed by x-ray crystallography, x-ray absorption fine structure (XAFS) (2–3), and a variety of spectroscopic techniques. Although x-ray crystallography is the primary probe for protein structure, its utility in providing accurate metalloprotein structural information is limited to availability and quality of crystalline samples. XAFS provides complementary local structural information about a selected metal ion for molecules in solution as well as crystals and polycrystalline samples (4). Obtaining structural information for Zn sites from conventional spectroscopic methods is difficult or impossible.

Significant advances have been made in recent years in XAFS theory and experimental technology; however, biological XAFS has not achieved its full potential due to limitations on data analysis and modeling of the experimental XAFS spectra. Structural information is usually obtained by XAFS when experimental spectra are fitted by a theoretical model that describes a hypothetical structure by a set of adjustable parameters (e.g., the edge-energy shift ΔE_0, the amplitude reduction factor S_0^2, the absorber- ith scatterer distances R_i, and the mean-square variations of electron single- (SS) and multiple- (MS) half-scattering paths σ_j^2 (5). Clark-Baldwin et al. (6) had described the difficulty of correctly modeling Zn active sites due to the high parameter correlation during XAFS data analysis (e.g., a $ZnCys_3O$ site has been incorrectly modeled with tetrahedral ZnS_4 (7), and $ZnCys_4$ was incorrectly modeled as ZnS_3O (8)) even in highly symmetrical conformations when MS is minimal. This is due to the presence of multiple solutions that are consistent with the experimental XAFS spectra within the uncertainties. This situation substantially worsens when MS dominates, which usually occurs when the x-ray absorbing metal atom is almost collinear with two other scatterers (e.g., metal-histidine complexes). In this case, the number of σ_j^2 parameters significantly exceeds the limited information content in the XAFS data (9), and additional information or assumptions must be provided.

An attempt to reduce the number of MS fitting parameters on active site of metalloproteins was reported by Dimakis and Bunker (10) by expressing the SS and MS σ_j^2 as a direct function of the first shell metal-amino acid residue distance, and the sample temperature thus virtually eliminating these parameters from the fitting procedure. This has been achieved by calculating the phonon normal mode properties on a series of small metal-amino acid model compounds at various metal-first shell scattering atom distances using density functional theory (DFT) (11). However, parameter reduction alone is not sufficient to avoid obtaining an incorrect structure via XAFS especially for cases of unknown metalloproteins where the information about the actual identity of the amino acids and their apparent distances from the known central absorbing metal is limited. We have developed an automatic iterative technique where a series of XAFS $\chi(k)$ spectra from hypothetical structural conformations is directly fitted to the experimental XAFS spectra using 1), the global fitting genetic algorithm ''differential evolution'' (DE) (12) to minimize the mean-square error,

$$e^2 = \sum_i^N (k_i^3(\chi(k)_i^{th} - \chi(k)_i^{exp}))^2/N, \quad (1)$$

between $\chi(k)$ XAFS spectra of hypothetical and experimental spectra, and concurrently 2), eliminate the numerous SS

doi: 10.1529/biophysj.106.090837

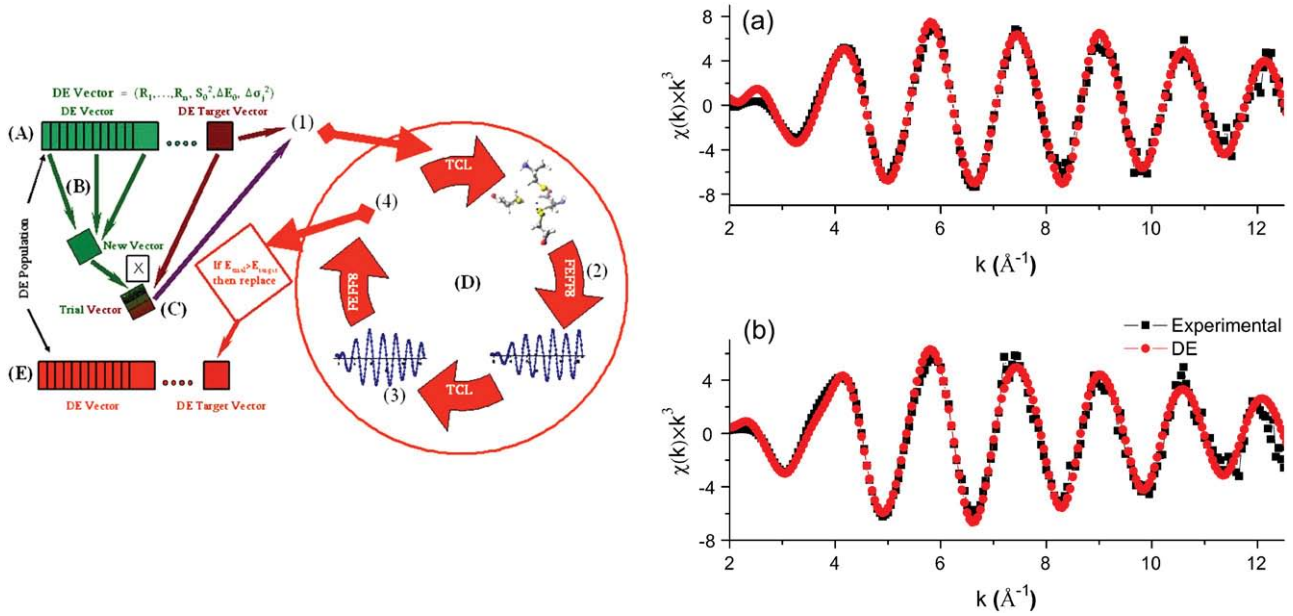

FIGURE 1 (*Left*) **Fitting procedure. (*A*). DE population consists of randomly generated DE parameter vectors {R_i, ΔE_0, S_0^2, $\Delta\sigma_j^2$}. (*B*) A new DE vector is generated from three other DE vectors of the initial population. (*C*) A trial DE vector is produced through a crossover operation. (*D*) 1), This DE vector is sent to the Tool Command Language (TCL) script and a hypothetical structure is generated; 2), χ(k) FEFF8- (13) generated spectra are produced and their scattering paths information is sent back to the TCL script; 3), σ_j^2 are calculated by the TCL script for all-important SS/MS paths; and 4), $\chi(k)$ with DWF information are generated. (*E*) The corresponding error fit is calculated. The process 1–4 is repeated for the target vector. If the error fit e^2 from the trial vector is less than the corresponding error fit from the target vector, the trial vector replaces the target vector in the DE population. (*Right*) Optimal DE four shell fits for a), $ZnCys_4$ and b), $Zn\text{-}Cys_3His$ tetrahedral geometries at 40 K. Fitting is performed over actual unfiltered experimental $\chi(k)$ k^3 XAFS spectra with noise at *k*-range 0–12.5 $Å^{-1}$ for a), 60 and b), 110 scattering paths, respectively.**

and MS σ_j^2 factor parameters from the fitting procedure (Fig. 1, *left*). The process is fully ab initio and only depends on the choice of the absorbing metal.

The method is tested using two Zn metalloprotein active sites: the homogeneous $ZnCys_4$ and the heterogeneous $ZnCys_3His$ ligation, either modeled as tetrahedral compounds. Structural information for these samples had been obtained by Clark-Baldwin et al. (6) using traditional least-squares approximation where 1), the number of available fit parameters was restricted ($\Delta E_0 = 9$ eV, $S^2_{\{0,S\}} = 1.02$, $S^2_{\{0,N\}} = 0.85$), during fitting to avoid wrong identification of the ligation, and 2), MS was ignored. To prove the validity of our technique and to avoid any bias over the selection of the initial parameter guess, the DE vector parameters have been chosen within the following wide range intervals: $R_{Cys} = (2.1–2.6)$ Å, $R_{His} = (1.8–2.5)$ Å, $\Delta E_0 = \pm 20$.eV, $S_0^2 = 0.5–1.5$, and $\Delta\sigma^2 = ((-1.) - 4.) \times 10^{-3}$ Å^2, where $\Delta\sigma^2 = \sigma^2_{j,actual} - \sigma^2_{j,DFT}$ accounting for static σ^2 and miscalculations on the $\sigma^2_{j,DFT}$. Each metal-ligand distance has been treated as an independent parameter. Although there has been no clear argument on how the DE process formally terminates, one can assume that this successfully occurs when the parameters of the final DE population are within an acceptable small range. In our case, we have set the following criteria for the final DE population: $\Delta E_0 = \pm 1.5$ eV, $S_0^2 = \pm 0.05$, average ligand distance $R = \pm 0.002$ Å, whereas no criteria apply to the $\Delta\sigma^2$ parameter.

Allowing parameter variation within the above intervals, the optimal DE vector fits for either sample has been obtained after 316 and 234 steps, respectively; the process may terminate earlier by selecting smaller parameter variation ranges (Fig. 1, *right*). These automatic high quality fits provide e^2 values of 0.655 and 0.670 for the homogenous and heterogeneous structures, respectively, which are higher than the corresponding values reported for the same compounds under conventional least-squares fitting (Table 1). This is due

TABLE 1 Optimal DE vector and error information compared with traditional least-squares fitting from XAFS (6) and x-ray diffraction (XRD) (14); parameters from DE and XAFS are at 40 K

	$ZnCys_4$			$ZnCys_3His$		
Parameters	DE	XAFS	XRD	DE	XAFS	XRD
R_{Cys} (Å)	2.338 (18)	2.30	2.36 (2)	2.338 (13), 2.087	2.28, 2.11	2.31 (6)
ΔE_0 (eV)	10.47	9.00		10.24	9.00	
S_0^2	0.843	1.02		0.888	1.02, 0.85	
$\Delta\sigma^2$ (10^{-3} Å^2)	−0.01	§		0.05	§	
Error e^2	0.655*	0.508‡		0.670*	0.341	
	0.418†			0.499†		

*Fit over raw $\chi(k)$ spectra without filtering and Fourier filtering of †ΔR = 4.5 Å; ‡ΔR = 1.5 Å was used.
§σ^2 = 3.9 (10^{-3} Å^2) as obtained from fitting with experimental EXAFS spectra. This value is reduced to 3.56 (10^{-3} Å^2) due to McMaster correction on the experimental EXAFS spectra. The latter value is very close to $\sigma^2_{DFT} \approx 3.32$ (10^{-3} Å^2) obtained by the DE algorithm.

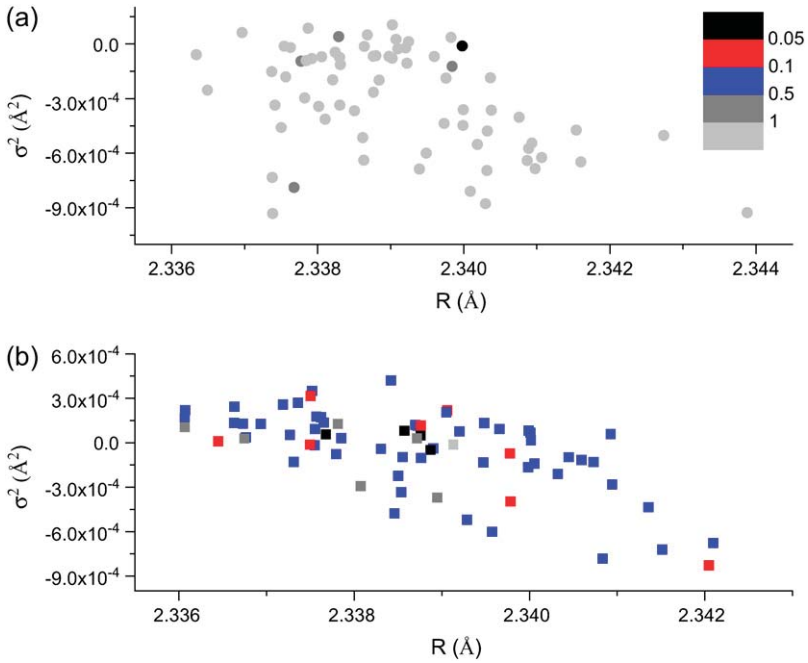

FIGURE 2 Average R_{Cys} variation versus $\Delta\sigma^2$ parameter for the optimal DE population for a), Zn-Cys_4 and b), Zn-Cys_3His, respectively. Colors refer to percentage errors with respect to the e^2 value of the optimal DE vector. Small average metal-ligand distance variation (<1/100 Å) is followed by substantial variation on the $\Delta\sigma^2$ parameter due to high correlation among the individual metal-ligand distances and $\Delta\sigma^2$, thus verifying the presence of multiple solutions within the EXAFS uncertainties. Optimal DE vector corresponds to $\Delta\sigma^2 \approx 10^{-5}$ Å^2 for either sample.

to 1), noise included in the DE fitting procedure (XAFS χ(k) spectra are unfiltered) and 2), increasing number of parameters on a two-shell refinement reduces the e^2 value in the least-squares fitting algorithm. An improvement in the error e^2 does not necessarily lead to improved data description: e^2 may systematically be reduced by performing a Fourier filtering on the XAFS $\chi(k)$ spectra. Average radial first shell distances obtained from the DE algorithm are in agreement with previous XAFS calculations and XRD. Analysis of the optimal DE population reveals the reliability of the structural parameters obtained with our method; focusing on the region of zero $\Delta\sigma^2$, the average first shell radial distance is $<$5/1000 Å (Fig. 2). Additionally, these vectors possess e^2 values within $<$% with respect to the value of the DE optimal vector.

Our method is general: it can be applied to any metalloprotein or hemoprotein as long as the SS and MS σ_j^2 values have been accurately expressed as a function of the first shell distance and sample temperature.

The code will be available upon request from the authors.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

ACKNOWLEDGMENT

We thank Prof. James Penner-Hahn for providing us the XAFS experimental spectra.

REFERENCES and FOOTNOTES

1. Quevillon-Cheruel, S., B. Collinet, C.-Z. Zhou, P. Minard, K. Blondeau, G. Henkes, R. Aufrère, J. Coutant, E. Guittet, A. Lewit-Bentley, N. Leulliot, I. Ascone, I. Sorel, P. Savarin, I. L. de La Sierra Gallay, F. de la Torre, A. Poupon, R. Fourme, J. Janin and H. van Tilbeurgh. 2003. A structural genomics initiative on yeast proteins. *J. Synchrotron Rad.* 10:4–8.
2. Stern, E. A. 1988. Theory of EXAFS. *In* X-ray Absorption. D. C. Koningsberger and R. Prins, editors. Wiley, New York. Chap. 1.
3. Lee, P. A., and J. B. Pendry. 1975. Theory of the extended x-ray absorption fine structure. *Phys. Rev.* B11:2795–2811.
4. Hasnain, S. S., and K. O. Hodgson. 1999. Structure of metal centres in proteins at subatomic resolution. *J. Synchrotron Rad.* 6:852–864.
5. The σ_j^2 parameters that are a direct function of the normal mode properties of the sample and its temperature are incorporated into the XAFS $\chi(k)$ amplitude as exponential decay of the form $e^{-2k^2\sigma_j^2}$ called Debye-Waller factor (DWF), $\hbar k$ being the photoelectron momentum.
6. Clark-Baldwin, K., D. L. Tierney, N. Govindaswamy, E. S. Gruff, C. Kim, J. Berg, S. A. Koch, and J. E. Penner-Hahn. 1998. The limitations of x-ray absorption spectroscopy for determining the structure of zinc sites in proteins. When is a tetrathiolate not a tetrathiolate? *J. Am. Chem. Soc.* 120:8401–8409.
7. Dent, A. J., D. Beyersmann, C. Block, and S. S. Hasnain. 1990. Two different zinc sites in bovine 5-aminolevulinate dehydratase distinguished by extended x-ray absorption fine structure. *Biochemistry*. 29: 7822–7828.
8. Povey, J. F., G. P. Diakum, G. D. Garner, S. P. Wilson, and E. D. Laue. 1990. Metal ion co-ordination in the DNA binding domain of the yeast transcriptional activator GAL4. *FEBS Lett.* 266:142–146.
9. Stern, E. A. 1993. Number of relevant independent points in x-ray-absorption fine-structure spectra. *Phys. Rev.* B48:9825–9827.
10. Dimakis, N., and G. Bunker. 2004. XAFS Debye-Waller factors for Zn metalloproteins. *Phys. Rev.* B70:195114(1)–195114(12).
11. Hohenberg, P., and W. Kohn. 1964. Inhomogeneous electron gas. *Phys. Rev.* 136:B864–B871.
12. Storn, R., and K. Price. 1997. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* 11:341–359.
13. Ankudinov, A. L., B. Ravel, J. J. Rehr, and S. D. Conradson. 1998. Real-space multiple-scattering calculation and interpretation of x-ray-absorption near-edge structure. *Phys. Rev.* B58:7565–7576.
14. Simonson, T., and N. Calimet. 2002. CysxHisy-Zn^{2+} interactions: thiol vs. thiolate coordination. *Proteins.* 49:37–48.