2002

Group-fitted ab initio single- and multiple-scattering EXAFS Debye-Waller factors

Nicholas Dimakis
The University of Texas Rio Grande Valley

Grant Bunker

PHYSICAL REVIEW B, VOLUME 65, 201103(R)

Group-fitted *ab initio* single- and multiple-scattering EXAFS Debye-Waller factors

Nicholas Dimakis and Grant Bunker
Illinois Insitute of Technology, Chicago, Illinois 60616
(Received 23 July 2001; published 10 May 2002)

X-ray absorption fine structure (XAFS) spectroscopy is one of the few direct probes of the structure of metalloprotein binding that is equally applicable to proteins in crystals, solutions, and membranes. Despite considerable progress in the calculation of the photoelectron scattering aspects of XAFS, calculation of the vibrational aspects has lagged because of the difficulty of the calculations. We report here initial results that express single- and multiple-scattering Debye-Waller factors as polynomial functions of first shell radial distance for metal-peptide complexes, enabling quantitatively accurate full multiple-scattering XAFS data analysis of active sites of unknown structure at arbitrary temperatures without the use of *ad hoc* assumptions.

DOI: 10.1103/PhysRevB.65.201103 PACS number(s): 61.10.Ht

X-ray absorption fine structure (XAFS) spectroscopy[1,2] is a technique that allows one to obtain structural and electronic information on metalloproteins in crystalline or solution form. Although there has been impressive progress in quantitative calculation of the photoelectron scattering aspects of XAFS, calculation of the essential vibrational aspects has lagged because of the difficulty of performing sufficiently accurate calculations. We report here steps towards a practical and broadly applicable approach to apply density functional theory (DFT) (Refs. 3–5) to XAFS multiple-scattering data analysis of metalloproteins.

In the simplest case, vibrational effects appear in the extended XAFS (EXAFS) equation in the form of $e^{-2k^2\sigma_j^2}$, where σ_j^2 is the mean square variation (MSV) of the jth photoelectron scattering path; these are functions of the normal mode phonon eigenfrequencies and eigenvectors of the system. This exponential term is known as the Debye-Waller factor (DWF) and it accounts for the disorder, both thermal and structural, of the relevant scattering path. Depending on the geometry of the structure under investigation, the actual number of single-scattering (SS) and multiple-scattering (MS) paths can easily be on the order of several hundred in a low symmetry structure. Since experimental data can only support a limited number of fitted parameters, i.e., $2\Delta k\Delta R/\pi+2\simeq 20-30$,[6] where Δk and ΔR are the k- and R-space data ranges, these factors must be known from an alternative source to quantitatively fit experimental data without making unjustified assumptions.

There is considerable interest in the active sites of metalloproteins in their various natural states and also in crystalline forms. Recently *ab initio* calculations have been presented of the DWF's of small molecules[7] using various approaches: DFT, semiempirical quantum, and force field models (FFM's). The equation of motion approach was also tested for the full Zn-tetraimidazole structure[8–12] as an alternative to the single parameter Einstein or Debye models, which fail to describe either single-scattering or multiple-scattering paths in systems with heterogeneous bond strengths. Loeffen and Pettifer (Ref. 10) calculated vibrational properties using a FFM derived from inelastic-neutron-scattering data, and combined them with electronic-scattering functions calculated with FEFF6.[13] They attributed the lack of quantitative agreement with experiment to deficiencies in the muffin tin potentials used by FEFF6. In this paper we find, in contrast, that we can obtain quantitative agreement with Zn-tetraimidazole solution spectra through use of the accurate eigenvalues and eigenvectors calculated by DFT, even without self-consistent electronic-scattering potentials. The need for self-consistent potentials is manifested primarily through the vibrational properties rather than the electronic-scattering properties. Ideally these would be calculated on the same basis in the same program, but until this is successfully implemented a hybrid DFT/FEFF approach appears to be viable as described below.

Active sites of metalloproteins typically consist of a central metal atom that is coordinated to amino acid residues from the protein. Previously[14] we have found that a reduced model cluster can be used to calculate DWF's by DFT methods with good accuracy. Our approach is to model those vibrational properties relevant to XAFS by applying such methods to the most common amino acid side groups found in metal sites (e.g., imidazole, carboxyl, phenol). In this paper, we present an approach in which DWF's are accurately parametrized with simple analytical expressions involving the first shell radial distance R (which varies with the coordination number and the types of binding groups) and temperature. This allows experimenters to use these functions in a fitting procedure to analyze EXAFS data, without the need to perform complex DFT calculations themselves. In practice this will lead to tables of equations that for a pair of metal-group DWF's will be given as a function of temperature and R for all possible paths. We here describe this approach for Zn bound to imidazole ligands. This model is first tested against the well-characterized structure of tetrahedral Zn-tetraimidazole,[10–12] and its behavior in high-coordinated group (imidazole) environments is determined. This structure was chosen because of its biological relevance and because of the presence of extensive multiple scattering[15] owing to the forward-scattering enhancement, which occurs when the absorbing and scattering atoms are in nearly collinear geometry. We then test this method on computationally modeled Zn-histidine complexes that coordinate in two distinct and biologically relevant ways.

A simplified cluster consisting of one imidazole ring, a Zn absorbing atom, and a hydrogen bound to the Zn was built and SCF energy/geometry optimized using Unichem/DGAUSS 5.0 (Ref. 16) under DFT on the local spin density

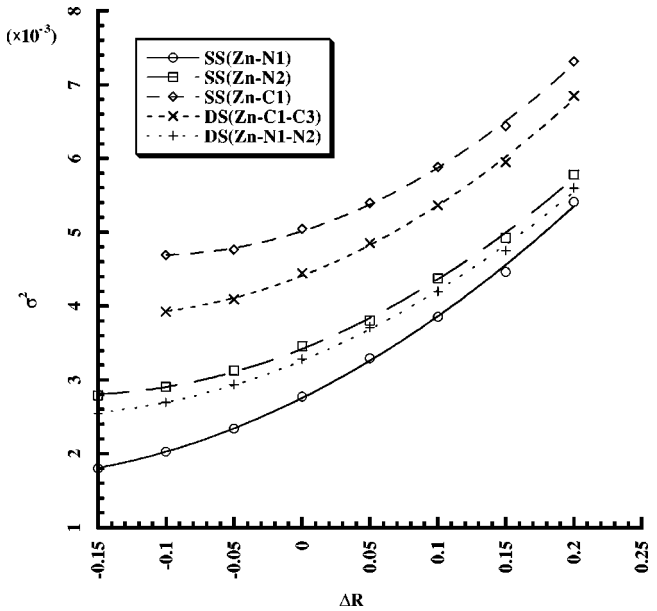

FIG. 1. Various DFT-calculated σ^2 SS and MS vs linearly translated ΔR_{Zn-N^1} distance from the absorbing Zn atom at 120 K. Lines represent second order polynomial fits for the ΔR dependence. At $\Delta R=-0.15$ Å one negative normal mode frequency appeared in the spectrum. This is an indication that such short bond lengths will lead to unrealistic results and for this reason in some of the paths this point has been excluded. The zero R refers to 1.956 Å.

(LSD) using the form of Vosko, Wilk, and Nusair (VWN) (Ref. 17) and double zeta (DZV) (Ref. 18) basis sets, which are specifically optimized for DFT use. Nonlocal, gradient-corrected spin density functionals were not employed here, since for these particular types of structures the LDA as described above provided sufficient accuracy. The high spin configuration was used since this leads to the ground state for this cluster. Normal mode calculations of phonon eigenfrequncies and eigenvectors followed and these were used as input to a computer program DEBYE written by the authors to calculate MSV's for all relevant multiple-scattering paths. Care was taken to ensure that no negative frequencies were observed in the normal mode spectrum, i.e., the cluster had converged to an energy minimum rather that to a saddle point. Since in most active sites the first shell radial distance is well resolved from both the higher shell and multiple-scattering contributions, it is useful to regard the higher shell SS and MS DWF's as functions of the first neighbor distance. We then perturb the Zn-N distance and we recalculate SS and MS as shown in Fig. 1.

We find that these surfaces (functions of R and T) for all paths can be accurately approximated by second order polynomials as a function of the perturbed radial distance ($R^2 \simeq 1$ for fitting). (Although three-atom triple-scattering paths have little contribution in the XAFS spectrum due to longer photoelectron path length their DWFs can still be given by Eqs. (1) and (2) with no loss of accuracy.) In all perturbed structures the imidazole ring remained frozen in order to avoid artificial ring deformations.

Summarizing all the above and using fitting techniques, the general formula that describes SS and MS DWF's has the form

$$\sigma^2(\Delta R,T)=\sigma^2(R_0,T)+A(T)\Delta R+B(T)\Delta R^2, \qquad (1)$$

where $A(T)=A_0+A_1T+A_2T^2+A_3T^3$ and $B(T)=B_0+B_1T+B_2T^2+B_3T^3$ are third order polynomials, and A_i, B_i, $i=0,1,..,3$ are temperature-independent coefficients. The MSV $\sigma^2(R_0,T)$ is a temperature-dependent factor that can be approximated as

$$\sigma^2(R_0,T)=\sigma^2(R_0,0)+C_0T+C_1T^2+C_2T^3. \qquad (2)$$

Coefficients A, B, C and the frozen $\sigma^2(R_0,0)$ are given at Table I. The above equations can be used to either calculate these factors from them directly, or input them in the FEFFIT (Ref. 19) program as fitting equations for temperatures from 1 up to 400 K in Fig. 2.

In the imidazole ring, atoms are arranged as follows: $N^{(1)}$ binds to the metal, and $C^{(1)}$, $C^{(2)}$ are second nearest neighbor carbons. $N^{(2)}$, $C^{(3)}$ are third nearest neighbor atoms. $C^{(2)}$ binds to $N^{(1)}$ and $N^{(2)}$.

In XAFS the average first shell radial distance in most cases can be easily determined by conventional methods at an accuracy of 0.01–0.02 Å. This information can be inputted in the equations and with the corresponding temperature all DWF's for any given path can be calculated. It is well known that not all paths are equally important. Beyond the first shell Zn-$C^{(1),(2)}$ SS and double scattering (DS), and Zn-$N^{(1)}$-$N^{(2)}$, Zn-$N^{(1)}$-$C^{(3)}$ DS paths are the most important in terms of the amplitude contribution in the XAFS spectrum. The large contribution of the first two paths to the XAFS spectrum is because of the short photoelectron path range, and the latter paths due to the almost collinear geometry of Zn-$N^{(1)}$-$N^{(2)}$,$C^{(3)}$ (''focusing effect''). In Table I we present all important paths, even the ones whose contributions are at 4% with respect to the first shell amplitude.

If care is taken, another quantity that can be derived with good accuracy from experimental data is the first shell DWF. Depending on data quality and temperature, the first shell DWF can be extracted by fitting experimental data with FEFF at an accuracy of $\pm 2-7\times10^{-4}$ Å^2. As will be discussed later, this information can be used to further enhance the quality of the calculated factors using coefficients from Table I.

In the case of Zn, tetrahedral structures are among the more highly coordinated structures to appear in active sites of metalloproteins. The higher the group-coordination number, the higher the error on the calculated DWF. For example, structures with two imidazoles will be predicted more accurately than four imidazole ones. Using the factors from Table I, MSV's for the most important paths are calculated and compared with the corresponding DFT-calculated factors (FSDFT's) for the whole tetraimidazole structure. The latter factors are considered as an accurate reference because the normal mode vibrational frequencies calculated under the DFT-DZ basis set are about 98% of the corresponding experimental infrared and/or Raman spectra for a given structure. Experimentally, Zn-tetraimidazole complexes have been crystallized with perchlorate ($r_{\text{Zn-N}^{(1)}}=2.00$ Å) and

TABLE I. Calculated *A*, *B*, and *C* coefficients for SS and MS DWF's for the Zn-imidazole group. TS paths Zn-$N^{(1)}$-$N^{(2)}$ and Zn-$N^{(1)}$-$C^{(3)}$ due to their linearity have DWF's that are the same as the corresponding double scattering for the path. For the low contribution TS Zn-$N^{(1)}$-$C^{(1),(2)}$ DWF's the corresponding DS factors may be used.

Path	Coefficient											
	A_0	A_1 $\times10^{-3}$	A_2 $\times10^{-4}$	A_3 $\times10^{-7}$	B_0	B_1 $\times10^{-2}$	B_2 $\times10^{-4}$	B_3 $\times10^{-6}$	C_0 $\times10^{-3}$	C_1 $\times10^{-5}$	C_2 $\times10^{-8}$	σ_0^2
Zn-$N^{(1)}$	6.357	−4.055	2.615	−3.15	5.001	3.312	7.83	−1.16	−1.88	3.17	−3.16	2.610
Zn-$N^{(2)}$	4.662	−4.454	2.547	−3.22	6.205	5.716	5.70	−0.79	−2.65	4.73	−4.86	3.186
Zn-$C^{(1)}$	4.063	−5.713	2.018	−2.50	8.069	7.297	7.57	−1.05	−2.80	9.28	−10.54	4.217
Zn-$C^{(2)}$	3.858	−8.033	1.777	−2.17	6.898	5.924	6.687	−0.93	−2.96	7.86	8.66	4.103
Zn-$C^{(3)}$	5.141	−3.230	2.811	−3.56	6.984	6.079	6.69	−0.93	−2.73	5.72	−6.05	3.548
Zn-$N^{(1)}$-$N^{(2)}$	4.830	−4.929	2.592	−3.27	5.773	4.837	5.72	−0.79	−2.50	4.10	−4.08	3.074
Zn-$N^{(1)}$-$C^{(3)}$	5.358	−4.070	2.879	−3.63	6.363	5.348	6.33	−0.88	−2.66	4.70	−4.76	3.383
Zn-$C^{(2)}$-$N^{(1)}$	4.846	−5.573	2.24	−2.75	5.966	4.244	6.08	−0.85	−2.29	4.33	−4.38	3.418
Zn-$C^{(1)}$-$N^{(1)}$	5.036	−4.507	2.425	−3.00	6.901	4.820	6.84	−0.94	−2.23	4.94	−5.26	3.314
Zn-$C^{(1)}$-$C^{(3)}$	4.754	−6.053	2.497	−3.13	6.545	6.989	6.72	−0.94	−2.82	7.09	−7.76	3.898
Zn-$C^{(2)}$-$N^{(2)}$	4.334	−6.437	2.144	−2.65	6.606	3.750	7.88	−1.11	−2.87	5.89	−6.28	3.486
Zn-$C^{(2)}$-$N^{(2)}$-$N^{(1)}$	4.441	−5.085	2.315	−2.89	6.576	4.60	6.253	−0.86	−2.61	4.82	−5.02	3.176
Zn-$C^{(1)}$-$C^{(3)}$-$N^{(2)}$	4.496	−1.836	2.391	−3.06	6.347	4.918	5.92	−0.82	−2.30	4.77	−4.57	2.965

tetrafluoroborate ions ($r_{\text{Zn-N}^{(1)}}=1.983$ Å). Without any constraints, FSDFT Zn-tetraimidazole optimized geometry approaches the tetrafluoroborate compound.

Factors calculated using the whole structure are about 1–3 % in error with respect to experimental data. A FSDFT calculation at 20 K gives MSV for Zn-N path 2.59×10^{-3} Å^2 whereas the corresponding value from temperature-dependent experimental EXAFS data is $(2.5\pm0.2)\times10^{-3}$ Å^2 (Ref. 7) and the calculated value using a force field model (FFM) is $(2.60\pm0.01)\times10^{-3}$ Å^2 (Ref. 8) (tetrafluoroborate compound). Our model with only one imidazole ring for the same path and at the same temperature is 2.77×10^{-3} Å^2 for the first shell distance at 1.98 Å. At room temperature, $T=300$ K and for the same path, our model gives 4.57×10^{-3} Å^2 vs 4.29×10^{-3} Å^2 (Refs. 7, 8) for the FFT which agrees with the FSDFT-calculated factor (4.27×10^{-3} Å^2). Also using our model for the perchlorate compound at $T=120$ K we get 2.96×10^{-3} Å^2, which is very close to the value $2.70\pm7\times10^{-3}$ Å^2 determined by experimental EXAFS.[4] This 6.5–7 % overestimation is expected because the Zn-$N^{(1)}$ bond will appear to be more rigid in an environment of higher imidazole coordination number than in a single one. Experimental XAFS data beyond the first shell cannot provide individual DWF's for individual paths, but rather measures these factors as a whole. Therefore we can compare individual factors only vs calculated factors from trusted reference models such as the FSDFT and FFM calculations. For DS Zn-$N^{(1)}$-$N^{(2)}$ path at 20 K and tetrafluoroborate compound our model gives 3.19×10^{-3} Å^2 vs $2.83-3.17\times10^{-3}$ Å^2 (Refs. 5, 7–9) (FFM) and 3.17×10^{-3} Å^2 for FSDFT. For the same path at 300 K, our model gives 5.31×10^{-3} Å^2 vs 4.93×10^{-3} Å^2 (Refs. 8, 9) (FFM) and $5.58\pm0.08\times10^{-3}$ Å^2 for FSDFT. (Factors reported under FFM models depend on number of parameters used and accuracy of force-field constants to describe the dynamics of the structure under investigation. It is generally known that fully *ab initio* calculations are far more accurate than FFM methods.) Due to different dynamics among the one-ring and four ring case some paths will have their DWF overestimated or underestimated and others will remain almost neutral. Paths such as Zn-$N^{(1)}$-$N^{(1)}$, triangle double-scattering path Zn-$N^{(1)}$-$C^{(1),(2)}$ have their factors to be very close to these predicted by our model. (In XAFS $C^{(1)}$ cannot be distinguished from $C^{(2)}$ so we use the average MSV factor of both paths.) Paths that have their factors overestimated are SS Zn-$C^{(3)}$, double-scattering Zn-$N^{(1)}$-$C^{(3)}$, Zn-$C^{(1)}$-$C^{(3)}$, and triple three atom Zn-$C^{(2)}$-$C^{(3)}$-$N^{(2)}$. Paths that they have their factors underestimated are SS Zn-$N^{(2)}$, SS Zn-$C^{(1),(2)}$, DS Zn-$N^{(2)}$-$C^{(2)}$, and TS Zn-$C^{(2)}$-$N^{(2)}$-$N^{(1)}$. If necessary for better agreement we can systematically correct this extra amount from these important paths. Figure 2 shows Fourier

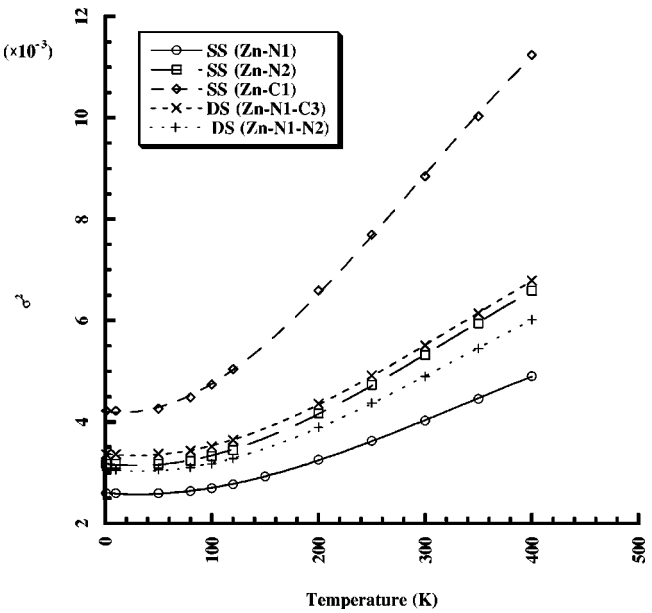

FIG. 2. Various DFT-calculated σ^2 SS and MS vs temperature T for radial distance 1.956 Å. Lines represent third order polynomial fits for the T dependence.

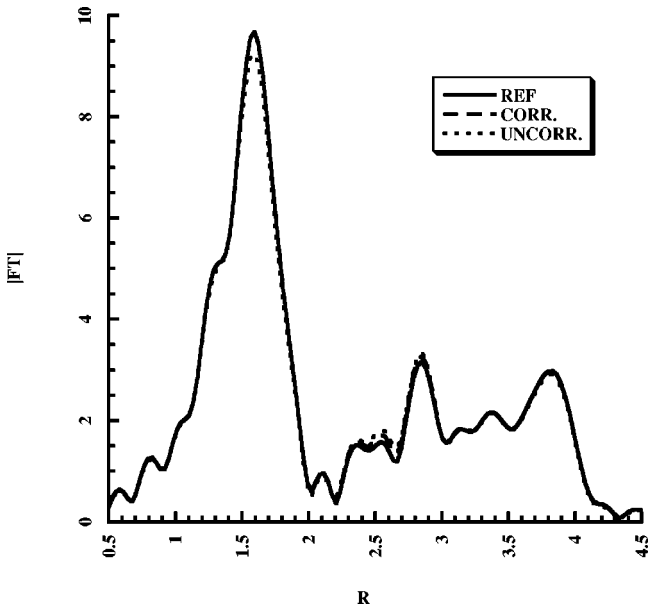

FIG. 3. Fourier transform of $\chi(k)$ for Zn-tetraimidazole at 150 K. The k range was taken from 2 to 18 Å^{-1}.

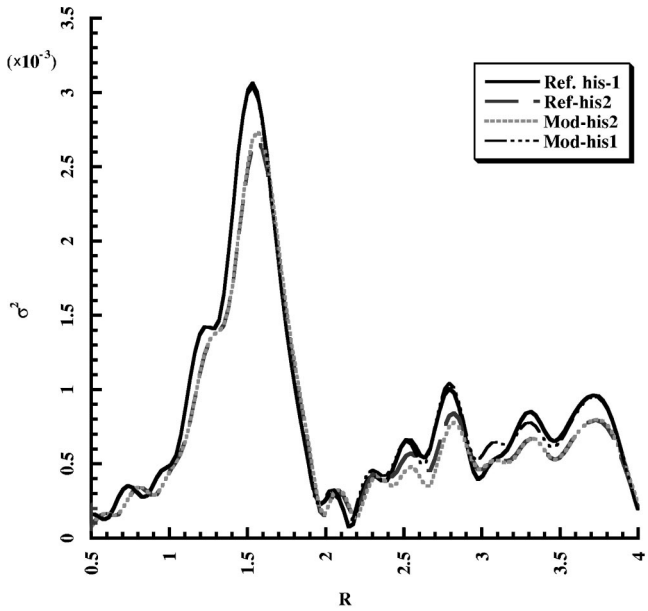

FIG. 4. Fourier transform of $\chi(k)$ for Zn-histidine-1 and Zn-histidine-2 structures at 150 K.

transformed $\chi(k)$ data at 150 K for Zn-tetraimidazole tetrafluoroborate using both corrected and uncorrected model-calculated factors. Agreement with the reference case is more than expected (Figs. 3, 4).

We now test our model for ''unknown'' cases. By the term ''unknown'' we mean structures that consist of a known metal and other compounds about which we have little information. Zn-histidine will be used in this test case. Histidines contain imidazole residues but are not simple imidazole rings, as in the first test case. In histidines the $C^{(3)}$ atom binds to a ''tail'' that consits of a carbon atom, amine, and carboxylate groups. In proteins the metal can bind to histidine nitrogen atoms in two ways, designated histidine 1 or 2 (his-1 and his-2). When Zn binds to his-1 the first shell radial distance is 1.906 Å, and as his-2 is 1.964 Å as has been calculated by using DFT under LSD approximation. Using the above proposed model and without any correction we input calculated DWF's into FEFF and compare the Fourier transforms. For histidine-2 only the ring atoms will be present in the XAFS spectrum but for histidine-1 also some of the ''tail'' atoms will be visible. In this case two carbon atoms respectively located at 3.52 and 3.23 Å away from the Zn absorbing atom will contribute via SS and double scattering to XAFS spectrum. Also the amine group at 3.65 Å will contribute to XAFS spectrum similarly to the carbon. The SS MSV is on the order of $1-3\times10^{2}$ Å^2, whereas the weak contributions from DS paths are at 5×10^{3} Å^2. Therefore contributions from these paths are minimal. In case of ''unknown'' samples, full DFT is not feasible since there are endless combinations (ligand type, coordination number) of test structures.

Analytical formulas that describe SS and MS paths for the Zn atom and imidazole group have been derived using *ab initio* DFT calculations of normal mode phonon eigenfrequencies and eigenvectors of small metal-group structure. This small model was proven to be sufficient to describe far more complicated structures. Application of the above model to other metals and groups is underway.

[1] E. A. Stern, in *X-Ray Absorption*, edited by D. C. Koningsberger and R. Prins (Wiley, New York, 1988), Chap. 1.
[2] P.A. Lee and J.B. Pendry, Phys. Rev. B **11**, 2467 (1975).
[3] P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864 (1964).
[4] W. Kohn and L.J. Sham, Phys. Rev. **140**, 1133 (1965).
[5] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
[6] E.A. Stern, Phys. Rev. B **48**, 9825 (1993).
[7] N. Dimakis and G. Bunker, Phys. Rev. B **58**, 2467 (1998).
[8] N. Dimakis and G. Bunker, J. Synchrotron Radiat. **6**, 266 (1999).
[9] A.V. Poiarkova and J.J. Rehr, Phys. Rev. B **59**, 948 (1999).
[10] P.W. Loeffen and P.F. Pettifer, Phys. Rev. Lett. **76**, 636 (1995).
[11] P.W. Loeffen, P.F. Pettifer and J. Tomkinson, Chem. Phys. **208**, 403 (1996).
[12] M. L. Hanham, P. F. Pettifer (unpublished).
[13] J.J. Rehr, J. De Leon, S.I. Zabinsky, and R.C. Albers, Phys. Rev. Lett. **69**, 3397 (1992).
[14] N. Dimakis and G. Bunker, J. Synchrotron Radiat. **8**, 297 (2001).
[15] G. Bunker, E.A. Stern, R. Blankenship, and W.W. Parson, Biophys. J. **37**, 539 (1982).
[16] UNICHEM/DGAUSS, Ver. 5.0 Oxford Molecular Group (unpublished).
[17] S.H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[18] N. Godbout, D.R. Salahub, J. Andzelm, and E. Wimmer, Can. J. Chem. **70**, 560 (1992).
[19] E.A. Stern, M. Newville, B. Ravel, Y. Yakoby, and D. Haskel, Physica B **208&209**, 117 (1995).