12-2007

Zinc cysteine active sites of metalloproteins: A density functional theory and x-ray absorption fine structure study

Nicholas Dimakis

Mohammed Junaid Farooqi

Emily Sofia Garza

Grant Bunker

Zinc cysteine active sites of metalloproteins: A density functional theory and x-ray absorption fine structure study

Cite as: J. Chem. Phys. **128**, 115104 (2008); https://doi.org/10.1063/1.2835601
Submitted: 06 July 2007 . Accepted: 26 December 2007 . Published Online: 18 March 2008

Nicholas Dimakis, Mohammed Junaid Farooqi, Emily Sofia Garza, and Grant Bunker

Zinc cysteine active sites of metalloproteins: A density functional theory and x-ray absorption fine structure study

Nicholas Dimakis,[1,a)] Mohammed Junaid Farooqi,[1] Emily Sofia Garza,[1] and Grant Bunker[2]
[1]*Department of Physics and Geology, University of Texas-Pan American, Edinburg, Texas 78539, USA*
[2]*Department of Biological, Physical and Chemical Sciences, Illinois Institute of Technology, Chicago, Illinois 60616, USA*

(Received 6 July 2007; accepted 26 December 2007; published online 18 March 2008)

Density functional theory (DFT) and x-ray absorption fine structure (XAFS) spectroscopy are complementary tools for the biophysical study of active sites in metalloproteins. DFT is used to compute XAFS multiple scattering Debye Waller factors, which are then employed in genetic algorithm-based fitting process to obtain a global fit to the XAFS in the space of fitting parameters. Zn-Cys sites, which serve important functions as transcriptional switches in Zn finger proteins and matrix metalloproteinases, previously have proven intractable by this method; here these limitations are removed. In this work we evaluate optimal DFT nonlocal functionals and basis sets for determining optimal geometries and vibrational densities of states of mixed ligation $Zn(His)_{4-n}(Cys)_n$ sites. Theoretical results are compared to experimental XAFS measurements and Raman spectra from the literature and tabulated for use. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2835601]

I. INTRODUCTION

The active sites of metalloproteins typically consist of a central metal ion that is coordinated to various amino acid residues. X-ray diffraction (XRD)[1] on protein crystals is the primary technique for obtaining structural information on proteins, but its utility is dependent on the availability and the quality of the protein crystals. X-ray absorption fine structure (XAFS) is a useful complementary probe that can be applied equally well to crystalline and noncrystalline samples such as protein solutions.[2–4] Conventional spectroscopic techniques are of limited utility in the important case of Zn^{+2} ions, owing to the fact that Zn is a spectroscopically silent metal ion because of its filled $3d$ orbital (d^{10}).

Zinc active sites typically are approximately tetrahedral complexes with histidines (His), cysteines (Cys) and carboxylate (e.g., aspartic and glutamic acid) amino acid ligands coordinated to the metal ion. Highly accurate information about the local structure around the zinc ion can be obtained by XAFS. However, due to destructive interference between the XAFS backscattering contributions of sulfur and the carbon, oxygen, or nitrogen atoms of $Zn(His)_3(Cys)$ and $ZnCys_4$, owing to the approximate π relative phase difference between them, their contributions cannot be readily distinguished because of parameter correlation.[5,6] This leads to the presence of multiple solutions that are consistent within the experimental XAFS uncertainties. Clark-Baldwin *et al.*[7] explained the difficulty of correctly modeling zinc sites that contain cysteine amino acid residues by performing an XAFS study of $Zn(His)_{4-n}(Cys)_n$ complexes (zinc finger proteins). Their study showed that a high degree of parameter correlation between the energy edge-shift ΔE_0, the R_i ith metal-first neighbor scattering atom distances, the amplitude reduction factor S_0^2, and the mean square variation σ_j^2 of a jth half-scattering path during data fitting may lead to inaccurate structures, even if multiple scattering (MS) is minimal (e.g., $ZnCys_4$ case).

MS is present on XAFS spectra of matrix metalloproteinases (MMPs).[8] MMPs play an important role on cell behavior; developing MMP inhibitors could help fighting fatal diseases such as cancer. MMP structure is of $Zn(His)_3X$, X being an amino acid including Cys residues, water, or other molecule. It has been reported that XAFS was unable to provide any useful structural information on higher shells of human gelatinase B MMP.[9] MS XAFS amplitudes are affected primarily through path degeneracies and the presence of additional σ_j^2 parameters for a jth scattering path. These parameters appear in the XAFS equation via terms of exponential form $e^{-2k^2\sigma_j^2}$ called Debye–Waller factors (DWFs); $\hbar k$ is the photoelectron momentum. For a low symmetry structure, such as an active site of a metalloprotein, the number of DWFs in the XAFS often exceeds the number of parameters experimental XAFS data can support ($2\Delta k\Delta R/\pi+2 \simeq 20-30$).[10] In this case DWFs must be calculated before the least squares fitting is performed, essentially eliminating them from the set of fitting parameters.

For a scattering path, the σ^2 of an XAFS DWF is proportional to the summation over the normal modes $\rho(n) \times(\hbar/2\omega_n)\times\coth(\hbar\omega_n/2K_BT)$, $\rho(n)$ is the projected vibrational density of states (pVDOS) over the normal mode frequencies ω_n, and T is the sample temperature. Single scattering (SS) and MS XAFS DWFs have been obtained previously for active sites of metalloproteins, using force-field methods,[11–14] semiempirical,[15] and density functional theory[16,17] (DFT) to calculate either the interatomic spring constants or the phonon normal mode spectrum (eigenfrequencies and eigenvectors) of a given sample. Dimakis and

[a)]Electronic mails: dimakis@utpa.edu and ndimakis@gmail.com.

Bunker[16,17] expressed σ_j^2 for zinc coordinated to Cys, His, and carboxylate amino acid residues as precomputed functions of the nearest atom distance (and bond angle where applicable) and the sample temperature. The modeled σ_j^2 (i.e., those generated from the parametrizations) were validated mostly by direct calculations of the phonon normal mode spectrum of large DFT-optimized[18,20] structures that served as a reference and allowed path-by-path comparison of the σ_j^2s. The validity of using large DFT-optimized structures as a reference was based on the demonstrated reliability of the DFT to provide accurate structural properties when using appropriate nonlocal spin density corrections with an adequate wave function basis set. However, it has been reported that DFT on metalloprotein sites in which zinc is coordinated to Cys residues overestimates the Zn–S bond by $\simeq$0.1 Å (Refs. 21 and 22) (depending on the nonlocal functional and basis set pair used) when compared to the value of 2.34(2) Å for Zn–S as averaged over 22 $ZnCys_4$ active sites from the Protein Data Bank; this first shell Zn–S distance overestimation would affect the hypothetical active site's local vibrational frequency spectrum and thus its SS and MS DWFs.

The goals of this paper are to (1) generate accurate polynomial parametrizations of the distance and temperature dependence of important SS and MS path σ^2's for Zn-Cys conformations by performing a cluster study on $Zn(His)_{4-n}(Cys)_n$ compounds optimized under various nonlocal DFT functionals and basis sets, and (2) validate these results on experimental XAFS spectra and identify how erroneous DWF estimation may lead to incorrect structures. In order to minimize parameter correlation during least squares fitting of XAFS experimental spectra with simulated spectra of a hypothetical structure, a technique[23] based on the global fitting genetic algorithm "differential evolution"[24] (DE) is used as an initial guess for the structure under question, in which the following quantity is minimized (Ref. 7):

$$\epsilon^2 = \frac{1}{N}\sum_i^N (k_i^3(\chi(k)_i^{\text{th}} - \chi(k)_i^{\text{exp}}))^2 \quad (1)$$

of XAFS experimental and hypothetical structure spectra over the available k spectra range with noise. DE proposed structures can then be in turn used as an input to the standard least squares for further refinement, with DWFs being eliminated from the parameter set.

II. METHOD

A. DFT calculations

Zinc coordinated to Cys amino acids in the form of $Zn(His)_{4-n}(Cys)_n$ $n=1,2,3,4$ and HS-Zn-Cys active sites (Fig. 1) were constructed; their optimal geometry and phonon normal mode spectrum were calculated using the DFT commercial packages DGAUSS (Ref. 25) and JAGUAR.[26] DGAUSS 5.0 utilizes a traditional DFT approach scaling as $\simeq N^4$ for a nonlocal spin density or hybrid functional, N being the dimension of the basis set of the compound, whereas JAGUAR 6.0 under the pseudospectral method[27] scales roughly as $N^{1.5}$ by calculating most of the fundamental CPU time-consuming integrals over a grid rather than the space defined by the basis set functions.

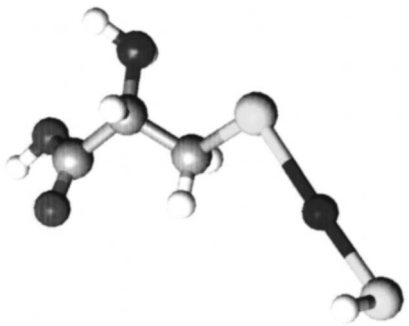

FIG. 1. Zn-Cys model for DWF calculation. $C^{(i)}$, $i=1,2$ are the first and second carbon neighbor atoms from the Zn^{+2} ion.

Under DGAUSS the generalized gradient approximation B88PW91 functional was employed, which consists of the Slater local functional[28] with the Becke 1988 nonlocal gradient correction[29] and the Perdew–Wang 1991 local and nonlocal functionals[30,31] as electron correlation functional. The double zeta DZVP2 (Ref. 32) basis set that includes polarization functions for all atoms is an improved basis set compared to the DZVP (Ref. 32) that has been reported earlier (Refs. 16 and 17) and did not include polarization functions on hydrogen atoms.

JAGUAR has the option of employing the hybrid functionals of B3LYP (Ref. 33) and X3LYP (Refs. 34 and 35) that uses the Vosko–Wilk–Nusair local functional[36] and Lee–Yang–Parr local and nonlocal functionals[37] for electron correlation. X3LYP is an extension to the B3LYP functional and includes the Perdew–Wang 1991 gradient correction exchange functional with exchange parametrized to fit a Gaussian exchange density. Hybrid functionals are paired with the LACVP (Ref. 38) basis set that uses effective core potentials for the metal Zn ion, whereas all remaining atoms are treated with the 6-31G basis set.[39–42] A calculation that includes the all-electron 6-31 (TM) basis set[43] to all atoms is included. Polarization and diffusion functions for all atoms have been added to LACVP and 6-31(TM) basis sets denoted by "**"[44] and "++,"[45] respectively, with the exception of the zinc ion that polarization functions were not available. DWF calculations for the B3LYP/LACVP* method ("*" adds polarization to nonhydrogen atoms) have been included; for this particular functional/basis set combination calculated normal mode frequencies are scaled by 0.9614.[46] No frequency scaling factor was employed on any other method.

The σ_j^2 of an arbitrary scattering path is expanded as a series of the distance ΔR and sample temperature as

$$\begin{aligned}\sigma_j^2(\Delta R,T) &\simeq \sum_i D_{ij}(T)\Delta R^i \\ &\simeq \sigma_{j0}^2 + \sum_{k=1}^{3} D_{j0k}T^k + \Delta R\sum_{k=0}^{3} D_{jik}T^k \\ &\quad + \sum_{i=2}^{4}\Delta R^i\sum_{k=0}^{2} D_{jik}T^k, \end{aligned} \quad (2)$$

where D_{jik} are ΔR and temperature independent factors de-

TABLE I. Calculated polynomial coefficients D_{jik} for j-SS and MS σ^2s ($\times 10^{-3}$ Å^2) for (HS)-Zn-Cys modeled Zn-Cys structures. The model equilibrium distance $R^0_{\mathrm{Zn-S}}$ is 2.188 Å and ΔR is in the interval of 2.188–2.6 Å.

Term		–S	$-C^{(1)}$	$-C^{(2)}$	$-S-C^{(1)}$	†	$-S-C^{(2)}$	†	$-C^{(1)}-C^{(2)}$	†	$-S-C^{(1)}-C^{(2)}$
σ_0^2		2.016	4.447	4.200	3.245	4.410	3.738	4.473	4.226	4.880	3.811
	$\times 10^{-3}$	−0.69	18.91	24.83	4.35	−1.27	3.75	−0.78	17.98	19.78	5.54
	$\times 10^{-5}$	1.89	10.69	10.00	4.03	2.47	6.76	7.22	10.05	11.44	4.42
	$\times 10^{-8}$	−1.45	−12.68	−11.83	−4.19	−1.54	−7.59	−7.79	−11.93	−13.64	−4.64
ΔR		6.190	0.975	1.515	3.106	5.364	2.886	3.571	1.328	0.516	3.374
	$\times 10^{-3}$	−108.13	−103.18	−350.36	−81.69	−95.28	−127.21	−91.91	−160.35	−104.97	−139.05
	$\times 10^{-4}$	2.36	0.38	0.74	1.44	2.21	1.54	1.74	0.68	0.30	1.58
	$\times 10^{-7}$	−2.75	−0.44	−0.91	−1.75	−2.62	−1.95	−2.17	−0.84	−0.35	−1.95
ΔR^2		−25.82	−2.65	−3.78	−1.16	−22.32	−9.82	−14.13	−3.04	−1.27	−12.27
	$\times 10^{-2}$	143.60	117.94	290.48	117.00	132.62	151.14	131.83	160.97	118.66	157.98
	$\times 10^{-4}$	−3.37	−0.31	−0.54	−1.60	−3.00	−1.46	−2.03	−0.43	−0.14	−1.71
ΔR^3		97.53	12.96	16.70	43.57	86.03	37.21	56.99	13.41	8.79	46.01
		−7.03	−5.17	−10.03	−5.70	−6.60	−6.69	−6.11	−6.46	−5.15	−6.91
	$\times 10^{-3}$	1.30	0.17	0.25	0.60	1.17	0.52	0.83	0.20	0.12	0.64
ΔR^4		−71.44	2.52	0.94	−22.96	−61.61	−16.00	−33.34	3.25	6.31	−24.43
		11.61	8.35	13.26	9.52	11.02	10.61	10.80	9.75	8.31	10.84
	$\times 10^{-4}$	−7.22	1.42	1.12	−1.66	−6.26	−1.12	−3.20	1.42	1.88	−1.90

termined by least squares fitting, $D_{ij}(T) \simeq \Sigma_k D_{jik} T^k$. The term σ^2_{j0} is distance independent factor and depends only on the type of the atoms involved in the scattering path and their position during the initial geometry optimization, thus contributing to a "static" DWF. Coefficients D_{jik} are given in Table I.

B. XAFS fitting

Modeled DWFs are tested using a two-stage fitting of experimental XAFS spectra of $Zn(His)_{4-n}(Cys)_n$ $n=2,3,4$ active sites: an initial guess of a hypothetical structure is obtained using the automatic algorithm of Dimakis and Bunker (Ref. 23) that couples the DE algorithm with the concurrent elimination of the MS DWFs from the fitting parameter list. The DE fit is performed over the actual $\chi(k)$ spectra with noise. This technique is employed as follows: the process starts generating an initial "population" by randomly selecting parameter vectors (analogous to chromosomes) $(\Delta E_0, R_i, S_0^2, \Delta\sigma^2)$. A new parameter vector is generated through a nonuniform crossover operation. The new vector will replace the old one in the population if the corresponding error given by Eq. (1) is smaller; the DE algorithm is terminated when the parameters $(\Delta E_0, R_i, S_0^2, \Delta\sigma^2)$ of the DE population are within an acceptable small range (e.g., $\Delta R_i = \pm 0.005$ Å). Hypothetical structures obtained from the DE algorithm are then used as input guesses for the IFEFFIT (Ref. 47) program that employs conventional least squares fitting methods using Levenberg–Marquardt minimization, with the SS and MS DWF being kept constant throughout the fit. Elimination of the floating DWF parameters and fitting with groups of atoms is the primary means of reducing the dimensionality of the parameter space. XAFS spectra of hypothetical structures that used in the DE and IFEFFIT programs are calculated via FEFF8.[48]

III. RESULTS AND DISCUSSION

A. DFT calculations

1. DWF calculations and geometry optimizations for $ZnCys_4$ clusters

DFT optimized geometries of $ZnCys_4$ tetrahedral clusters and SS/MS σ^2s directly calculated from their phonon spectrum ("reference" σ^2) are used to select the DFT functional/basis set that best describes XAFS DWFs for Zn-Cys compounds. It is observed (Table II) that reference σ^2s at 150 K are in agreement with the corresponding σ^2s calculated from the phonon spectrum of the model (HS)-Zn-Cys structures. In this case the first shell Zn-Cys distance of the model matches the distance of the reference structure. The use of effective core LACVP basis set caused a higher overestimation of the Zn–S bond ($\simeq$0.15 Å) with respect to the 2.34 Å average distance as being reported by XRD measurements Ref. 22. When the all-electron basis set of DZVP2 and 6-31G(TM) was used, DFT accuracy was improved, leading to a smaller Zn–S distance overestimations of about 0.08 and 0.06 Å, respectively. The B88PW91/DZVP2 method provided the closest geometry to the corresponding XRD-recorded structure than other DFT methods employed here concomitant with the agreement between reference and modeled obtained σ^2s for the Zn-Cys SS/MS paths examined.

2. Projected vibrational density of states for $ZnCys_4$ clusters

The pVDOS ($\rho(n)$ versus ω_n) spectrum reveals the vibrational contribution to the XAFS DWFs of a scattering path independently of the sample's temperature. The pVDOS of the SS Zn–S and $Zn-C^{(1)}$ paths, and the double scattering (DS) $Zn-S-C^{(1)}$ path is plotted on Fig. 2. The σ^2 value of

TABLE II. Zn–S distances and SS/MS σ^2 values (at 150 K) for DFT optimized $ZnCys_4$ clusters under various functionals and basis sets. Standard deviations are shown in parenthesis.

Functional Basis set	R_{Zn-S} (Å)	σ^2 ($\times 10^{-3}$ Å) Zn–S Ref.	Zn–S Mod.	Zn–$C^{(1)}$ Ref.	Zn–$C^{(1)}$ Mod.	Zn–S–$C^{(1)}$ Ref.	Zn–S–$C^{(1)}$ Mod.	Zn–$C^{(1)}$–$C^{(2)}$ Ref.	Zn–$C^{(1)}$–$C^{(2)}$ Mod.
B88PW91									
DZVP2	2.395(2)	3.62(3)	3.750	9.84(20)	11.15	5.57(3)	5.94	9.45(4)	9.82
	2.453(10)	4.71(30)	4.810	12.35(100)	10.83	6.94(50)	6.51	13.4(35)	0.12
B3LYP									
LACVP*	2.48(1)	5.00(42)	5.392	15.90(350)	13.35	8.13(120)	7.87	27.5(102)	19.56
	2.51	5.633	6.145	13.71	13.70	7.74	8.44	17.06	19.13
B3LYP									
LACVP**++	2.45	4.348	4.994	11.29	12.58	6.21	7.31	12.82	11.65
	2.485(3)	4.925(80)	5.000	14.74(170)	13.24	6.86(42)	7.49	15.16(180)	12.01
	2.522	5.641	5.767	15.74	13.31	7.18	8.00	12.53	13.38
X3LYP									
LACVP**++	2.45	4.297	4.373	16.46	13.22	7.30	7.02	13.69	11.35
	2.482(5)	4.938(60)	4.934	14.22(2)	12.49	7.51(8)	7.25	17.54(50)	11.55
	2.518	5.474	5.676	14.83	13.04	7.50	7.87	17.33	11.18
X3LYP									
6-31G(TM)**++	2.391(9)	3.395(70)	3.527	9.23(45)	8.11	5.23(1)	5.17	12.26(114)	7.75
	2.428(6)	3.926(7)	3.958	12.45(205)	8.08	6.10(32)	5.43	13.55(112)	7.81

the SS Zn–S path only depends on the following three normal modes of the $ZnCys_4$ vibrational spectrum: the low frequency $C^{(1)}$–$C^{(2)}$ bending plus N torsion, the Zn–S stretching, and the S–$C^{(1)}$–$C^{(2)}$–C_α–N deformation mode, C_α being the α-carbon atom. Normal mode frequency contributions to the σ^2_{Zn-S} can be seen on Table III. The Zn–S stretching mode is the dominant vibration for this path; it contributes about 61%–78% on σ^2_{Zn-S} with the low frequency mode to follow (≃12%–33%), depending on the method used to calculate the phonon spectrum. The deformation mode has a minimal impact on the σ^2_{Zn-S} (≃6%–9%). The pVDOS spectrum of the SS Zn–$C^{(1)}$ is more complex; the $\sigma^2_{Zn-C^{(1)}}$ highly depends on two bending modes, the Zn–S–$C^{(1)}$ and $C^{(1)}$–$C^{(2)}$–C_α (≃50%–56%) with the remaining contribution originating from the low frequency libration modes ($\omega_n < 100$ cm^{-1}) and other vibrations, as shown on Fig. 2 (middle graph). The high sensitivity of the pVDOS of the SS Zn–$C^{(1)}$ path on these low frequency bending modes is responsible for $\sigma^2_{Zn-C^{(1)}} > \sigma^2_{Zn-S}$ (at the same sample temperature). The pVDOS of the DS Zn–S–$C^{(1)}$ path mostly depends on three vibrations that are already included on the SS Zn–$C^{(1)}$ path pVDOS spectrum: the two bending modes Zn–S–$C^{(1)}$ and $C^{(1)}$–$C^{(2)}$–C_α, and the S–C stretch (Fig. 2 lower graph). Therefore accurate prediction of the $\sigma^2_{Zn-C^{(1)}}$ factors leads to accurate prediction of the $\sigma^2_{Zn-S-C^{(1)}}$. The $\sigma^2_{Zn-C^{(1)}} < \sigma^2_{Zn-S-C^{(1)}}$ is due to the absence of the libration modes on the DS Zn–S–$C^{(1)}$ pVDOS spectrum.

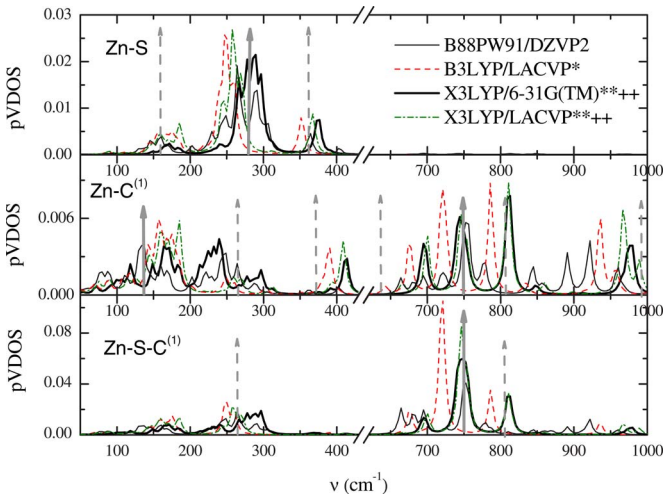

FIG. 2. (Color online) pVDOS spectrum for the SS Zn–S path (upper graph), Zn–$C^{(1)}$ (middle graph), and DS Zn–S–$C^{(1)}$ (bottom graph) of the DFT optimized $ZnCys_4$ clusters under various DFT functionals and basis set combinations. Vertical lines represent normal mode frequencies as reported in the literature. Solid lines refer to Zn-Cys complexes (Zn–S stretch, upper graph, Ref. 49; Zn–S–C bend, left solid line middle graph, Ref. 53; C–S stretch, right solid line middle graph, Ref. 55), whereas dashed lines refer to Cys vibrations alone.

The Zn–S stretching mode has been experimentally recorded at 282 cm^{-1} by Vargek *et al.*[49] on COM39 (Ref. 50) compound using resonance Raman spectroscopy. The other two modes are internal Cys modes and do not involve the Zn^{+2} ion. Thus vibrational modes from Cys and Cys-containing compounds may serve, to a good approximation, as a reference for comparison with corresponding DFT calculated vibrations of the $ZnCys_4$ cluster. The low frequency mode is difficult to measure by Raman spectroscopy. Therefore, for this particular mode, our DFT-calculated frequencies are compared with the corresponding value calculated by B3LYP/6-31G(p,d) on 4-fluorobenzidine L-Cys as

TABLE III. Centroids ($\bar{\nu}$) of DFT calculated normal mode frequency bands for $ZnCys_4$ clusters for various functional/basis set combinations and their apparent contribution to the σ^2 for the SS Zn–S and Zn–$C^{(1)}$ paths.

		Functional and basis set								
		B88PW91 DZVP2		B3LYP LACVP*		X3LYP LACVP**++		X3LYP 6-31G(TM)**++		
Path	Mode	$\bar{\nu}$(cm^{-1})	%(σ^2)	$\bar{\nu}$(cm^{-1})	%(σ^2)	$\bar{\nu}$(cm^{-1})	%(σ^2)	$\bar{\nu}$(cm^{-1})	%(σ^2)	Expt.
Zn–S	$\delta(C^{(1)}-C^{(2)}-N)$	156	24	166	33	170	32	170	12	157[a]
	ν(Zn–S)	271	71	251	61	260	62	281	78	282[b]
	$\delta(S-C^{(1)}-C^{(2)}-C_\alpha-N)$	370	6	353	8	367	8	373	9	358[c]
Zn–$C^{(1)}$	δ(Zn–S–$C^{(1)}$)	146	40	163	50	167	47	159	36	138[d]
	$\delta(C^{(1)}-C^{(2)}-C_\alpha)$	245	16	262	4	264	3	251	19	268[c]
	$\delta(C^{(1)}-C^{(2)}-C_\alpha-N)$	387	4	388	5	407	5	405	4	370[c]
	$\nu(C^{(2)}-C_\alpha)$	696	2	678	9	701	6	696	6	635[c]
	ν(S–C)	752	10	721	7	749	5	745	7	750[e]
	$\nu(C^{(1)}-C^{(2)}-C_\alpha-N)$	844	1	793	5	811	6	810	5	871[c]
	$\nu(C^{(1)}-C^{(2)}-N)$	923	9	941	6	970	7	994	6	999[c]

[a]Reference 51.
[b]Reference 49.
[c]Reference 52.
[d]Reference 53.
[e]Reference 55.

reported by Ye *et al.*[51] The frequency value serving as reference for the low contribution deformation mode has been recorded by Raman spectroscopy on L-Cys as reported by Pawlukojc *et al.*[52] The eigenfrequencies of these two vibrational modes are well estimated by the B88PW91/DZVP2 and X3LYP/6-31G(TM)**++ methods, e.g., for the bending mode $e_{\mathrm{err}}^{\mathrm{B88PW91}} \equiv \nu_{\mathrm{ref}}^{\mathrm{B88PW91}} - \nu_{\mathrm{calc}}^{\mathrm{B88PW91}} = 1$ cm^{-1} versus $e_{\mathrm{err}}^{\mathrm{X3LYP}} = -13$ cm^{-1}. This implies that the above methods provide accurate $\sigma^2_{\mathrm{Zn-S}}$ values for any sample temperature, i.e., at 150 K $|\sigma^2 - \sigma^2_{\mathrm{ref}}| \simeq 5\times10^{-5}$ Å^2 under B88PW91/DZVP2, where σ^2_{ref} is the corresponding value obtained by substituting the DFT calculated spectrum with reference frequency values as provided by Table III. Therefore these two methods provide accurate XAFS DWFS for $ZnCys_4$ complexes at the overestimated Zn–S distances, as stated in Sec. III A 1.

The Zn–S–$C^{(1)}$ bending mode that highly contributes to the $\sigma^2_{\mathrm{Zn-C^{(1)}}}$ has been experimentally recorded at 138 cm^{-1} by Elgren and Wilcox[53] using low frequency resonance Raman spectroscopy on metallothionein. The B88PW91/DZVP2 predicts this mode at higher accuracy that other methods tested here. This method is also accurate on predicting the S–C stretching mode, however, frequency values for the C–C stretch and $C^{(1)}-C^{(2)}-C_\alpha$ bend are about 8%–9% off with respect to the reference values. On the contrary these modes are more accurately predicted by other methods, e.g., the former by the X3LYP/6-31G(TM)**++ and the latter by the B3LYP/LACVP**++. For this path either B88PW91/DZVP2 or X3LYP/6-31G(TM)**++ will reproduce $\sigma^2_{\mathrm{Zn-C^{(1)}}}$ at an improved accuracy with respect to DFT nonlocal functionals paired with effective core LACVP basis sets due to their more accurate prediction on the low frequency bending modes; the former method is chosen as the optimal DFT functional/basis for describing Zn-Cys XAFS DWFs due to its higher accuracy on describing $ZnCys_4$ optimized geometries with respect to the later method, as shown on Sec. III A 1.

3. Zn–S bond overestimation versus Cys coordination number on $Zn(His)_{4-n}(Cys)_n$ clusters

Overestimates of the Zn–S distance by DFT occur for any zinc site in which a Cys residue is present within the metal nearest coordination sphere. This is observed for Zn–S distances of DFT geometrically optimized $Zn(His)_{4-n}(Cys)_n$, $n=1-4$ heterogeneous clusters under the B88PW91/DZVP2 method. Zn–S distance is observed at 2.27 versus 2.22(2) Å for $Zn(His)_3(Cys)$, 2.31 versus 2.28(2) Å for $Zn(His)_2(Cys)_2$, 2.355 versus 2.31(6) Å for $Zn(His)(Cys)_3$, and 2.42 versus 2.36(6) Å for $ZnCys_4$, for DFT optimized structures and XRD results from the Cambridge Structural Database, respectively, as reported by Simonson and Calimet.[54] The average distance $R_{\mathrm{Zn-S}}$ for the highly symmetric site of $ZnCys_4$ reported by Simonson and Calimet of 2.36(6) Å is in agreement-within error ranges provided—with the value of 2.34(2) Å given by Dudev and Lim[21] that was used in our DWF analysis. Therefore, the Zn–S distance overestimation by DFT is systematic and at approximately 0.06(2) Å irrespective of the cysteine coordination number (Fig. 3). The $R^0_{\mathrm{Zn-S}}$ equilibrium distance for modeled Zn-Cys DWF is upshifted by +0.1 Å that is within the upper limits of the error reported above.

B. XAFS measurements

1. Comparison of modeled Zn-Cys DWF with experimental XAFS

The $\sigma^2_{\mathrm{Zn-S}}$ (SS) values of $Zn(His)_{4-n}(Cys)_n$, $n=1-3$ active sites are obtained by fitting experimental XAFS with simulated spectra; the latter are generated by the FEFF8 pro-

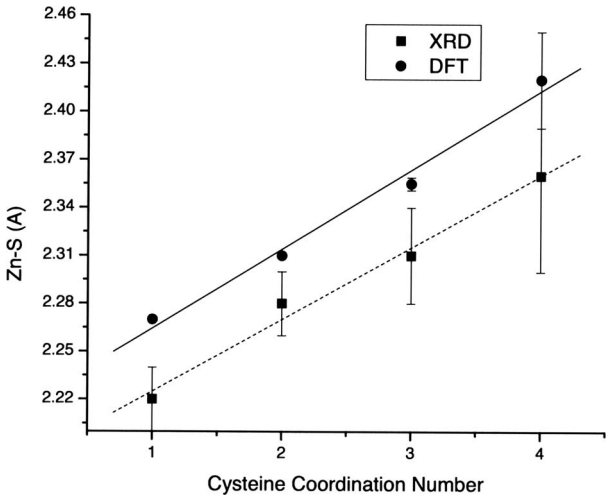

FIG. 3. Zn–S distance vs Cys coordination number for tetrahedral $Zn(His)_{4-n}(Cys)_n$ structures $n=1,2,3,4$ as deducted from experimental XRD spectra and DFT B88PW91/DZVP2 optimized geometries.

gram (Table IV, col. 4). These values must be McMaster-corrected[56] (McM) by adding a small negative constant $\Delta\sigma^2$ to their original "raw" values, i.e., $\sigma^2_{\mathrm{McM}}=\sigma^2_{\mathrm{raw}}+\Delta\sigma^2$, $\Delta\sigma^2=-3.35\times10^{-4}$ Å^2 for the Zn ion (Table IV, col. 5). This procedure is justified as follows: any sufficiently slowly varying multiplicative factor $f(E)$ (such as the McMaster correction) appears as a change in σ^2 (or more generally, a correction to the even order cumulants). This follows from the fact that the data are normalized and $E-E_0$ is quadratic in k. The extended x-ray absorption fine structure data are normalized so that the value immediately above the edge (energy E^+) is set to 1.0. We assume the energy variation of $f(E)$ is weak enough that we can make the approximation $E^+\approx E_0$. In this case the effective correction factor is $f(E)/f(E_0)\approx\exp(f'(E_0)/f(E_0)(E-E_0))$ $=\exp(f'(E_0)/f(E_0)\hbar^2k^2/2m)=\exp(-2k^2\Delta\sigma^2)$ expanding to first order in $E-E_0$. Thus the correction appears as a small (usually negative since μ_0 decreases with increasing energy) increment to σ^2. Expanding to higher order introduces small increments to the cumulants of fourth and higher orders. Omission of this correction will result in a larger relative error for the low temperature XAFS measurements. It is observed from Table IV that modeled σ^2s are (1) within the error ranges of the McM corrected DWFs as obtained by fitting experimental XAFS spectra with spectra of hypothetical structure, irrespective of the sample's temperature, Cys coordination number, and $R_{\mathrm{Zn-S}}$ average distance, and (2) XAFS $\sigma^2_{\mathrm{Zn-S}}$ as reported in the literature are not always correlated with $R_{\mathrm{Zn-S}}$, indicating miscalculations of DWFs due to either high parameter correlation, absence of MS in the fitting, and steric hindrance effects that increase structural disorder.

2. Using DE and DWF elimination to improve XAFS structural ability on Zn-Cys active sites

The DE algorithm (Ref. 23) is employed to obtain a hypothetical structure that will serve as an initial guess for the IFEFFIT program. DWFs for Zn-Cys structures are calculated using the model polynomial expressions reported on Sec. II A and Table I for SS/MS paths; DWFs for Zn-His structures are obtained from expressions of Ref. 17. To avoid any bias the DE fit is performed over the actual noisy $\chi(k)$ XAFS spectra in the range $k=2-12$ Å^{-1} for the mixed ligation $Zn(His)_{4-n}(Cys)_n$, $n=2,3$ zinc sites, and $k=2-12.5$ Å^{-1} for the $ZnCys_4$ site. The slightly increased range on the $ZnCys_4$ k high end is due to improved quality of the $ZnCys_4$ XAFS experimental spectra. Structural information and XAFS fitting parameters obtained from the DE algorithm can be seen on Table V, cols. 2–4. Our DE-based method performs a four-component fit; it is observed that on the highly symmetric $ZnCys_4$ site Zn–S distances are not all equal, in agreement with results from the corresponding DFT optimized structure calculations. The error on the overall $\sigma^2_{\mathrm{err}}=\sigma^2_{\mathrm{DFT}}-\sigma^2_{\mathrm{exp}}$ obtained by the DE algorithm is at the order of 10^{-5} Å^2, and thus is negligible, e.g., for the SS Zn–S path of the $Zn(His)_2(Cys)_2$ active site, $\sigma^2_{\mathrm{DFT}}=2.695\times10^{-3}$ Å^2 (overestimated by $\simeq5\%$ with respect to the value reported by Ref. 7, Table IV), $\sigma^2_{\mathrm{err}}=1.65\times10^{-5}$ Å^2. For the other two samples, $ZnCys_4$ and $ZnHisCys_3$, the average $\sigma^2_{\mathrm{DFT}}\simeq3.32\times10^{-3}$ Å^2 is underestimated by $\simeq7\%$ with respect to corresponding Ref. 7 values.

The hypothetical structure from the DE algorithm output in turn serves as an input for IFEFFIT, an XAFS fitting programs based on the traditional least squares fit to test the validity of the DE output, and examine the effect of noise contained in the original XAFS experimental spectra on the DE fit. IFEFFIT performs a least squares fit on back-Fourier transformed XAFS experimental spectra with $R=0.5-4.5$ Å

TABLE IV. XAFS SS $\sigma^2_{\mathrm{Zn-S}}$ as directly obtained by fitting experimental XAFS with simulated FEFF spectra (Raw), McMaster corrected (McM), and DFT modeled values at various temperatures T and Cys coordination numbers N_{Cys}.

			σ^2 ($\times10^{-3}$ Å) Experimental			
T(K)	$R_{\mathrm{Zn-S}}$ (Å)	N_{Cys}	Raw	McM	Modeled	Ref
20	2.28	1	2.5(4.5)	2.165(4.5)	2.870	55 and 57
40	2.30	4	3.9	3.565	3.279	7
	2.28	3	3.9	3.565	3.030	
	2.26	2	2.9	2.565	2.790	
77	2.33	4	4.83(1.21)	4.495(1.21)	3.386	52 and 58
289	2.31	4	6.53	6.195	6.095	51 and 59

TABLE V. DE and IFEFFIT output fitting parameters for $Zn(His)_n(Cys)_{4-n}$ $n=0-2$ active sites. DWFs were not obtained by fitting in either case. Unless specified S_0^2, ΔE are same for either ligand. All measurements are at 40 K.

	Parameters							
	DE				IFEFFIT			
Sample	R_{Cys}	R_{His}	S_0^2	ΔE_0	R_{Cys}	R_{His}	S_0^2	ΔE_0
$ZnCys_4$	2.315		0.84	10.47	2.319		0.89	6.00
	2.330				2.334			
	2.340				2.344			
	2.375				2.379			
$Zn(His)(Cys)_3$		2.089	0.84	10.24		2.093	0.98	4.87[a]
	2.330				2.334			7.90[b]
	2.350				2.354			
	2.350				2.354			
$Zn(His)_2(Cys)_2$		2.002	0.93	11.06		2.012	1.10[a]	7.78[a]
		2.064				2.074		
	2.298				2.308		0.93[b]	5.03[b]
	2.316				2.326			

[a]His ligands.
[b]Cys ligands.

to account for inclusion of MS. The fitting parameter list consists of ΔE_0, ΔR, and S_0^2. DWFs are not allowed to vary during fitting to minimize parameter correlation. IFEFFIT high quality fits and corresponding Fourier transforms can be seen on Fig. 4. For $ZnCys_4$ and $Zn(His)(Cys)_3$ structures $\Delta R'_{Zn-S} \equiv R^{DE}_{Zn-S} - R^{fit}_{Zn-S} \simeq 0.004$ Å, whereas for $Zn(His)_2(Cys)_2$ this error is increased to $\Delta R'_{Zn-S} \simeq 0.01$ Å; this is due to the quality of the original $\chi(k)$ XAFS spectra. On the latter case distinct ΔE_0s for Cys and His residues have been employed. The small error on the first shell distances R_{Zn-S}, R_{Zn-N} verifies that the initial DE guess provide accurate enough structural information for XAFS analysis.

Our experimental XAFS spectra data are the spectra used by Clark-Baldwin *et al.*[7] The structural parameters of the zinc sites reported there are at the SS limit with E_0 kept fixed during the fitting procedure. It is observed (Tables IV and V) that the Zn–S and Zn–N average distances reported at Ref. 7 differ in some cases as much as 0.06 Å, e.g., on $ZnHisCys_3$ average $R_{Zn-S} \simeq 2.34$ Å by DE/IFEFFIT versus 2.28 by Ref. 7. This is mainly due to the high correlation between the first shell distances and the overall σ^2 parameter; a small deviation on the σ^2 may lead to large discrepancies on the first shell distances. In our DE/FEFFIT procedures SS and MS are not obtained by fitting; they are expressed as a function

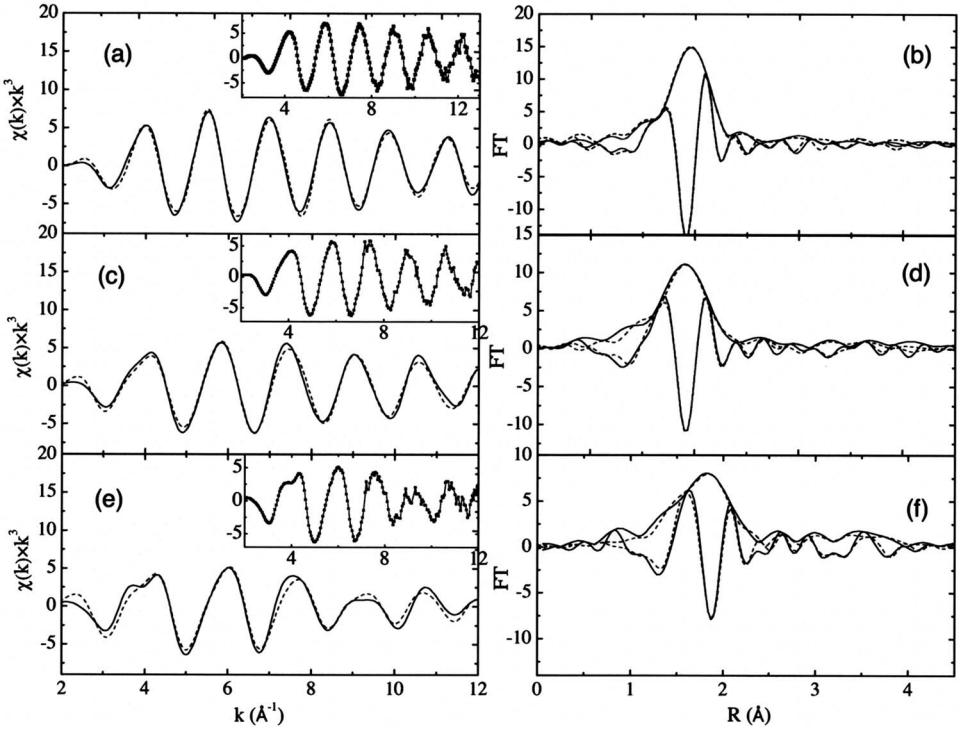

FIG. 4. Least squares fit of hypothetical $Zn(His)_{4-n}(Cys)_n$, $n=4$ [(a) and (b)], $n=3$ [(c) and (d)] and $n=2$ [(e) and (f)] structures with respect to filtered experimental XAFS spectra. Original $\chi(k)$ XAFS spectra can be seen at the inserts. (b), (d), and (f) are Fourier transforms (magnitude and imaginary part) of (a), (c), and (e) filtered $\chi(k)$ spectra. Fit is over filtered k-range with $\Delta k=2-12.5$ Å^{-1} for $ZnCys_4$, $\Delta k=2-12$ Å^{-1} for the mixed ligation complexes due to noise at high k range; $\Delta R=0.5-4.5$ Å for all cases to include MS. SS and MS σ^2s were kept fixed during the fitting procedure to the values obtained by the DE algorithm, whereas ΔE_0, S_0^2, and ΔR were allowed to vary. In all cases the $\Delta R^{error} < 0.01$ Å.

of the first shell distance and eliminated from fitting, thus minimizing correlation with the first shell distance parameters.

IV. CONCLUSION

XAFS conventional analysis ability to determine the structures of the active sites of Zn-Cys sites is seriously compromised by high fitting parameter correlation and the presence of MS. MS introduces a larger number of parameters into the XAFS equation than typically can be supported by the spectral information content. In particular, Debye–Waller factors (DWFs) must be specified for each path when fitting. In our approach the size of the parameter space under full MS treatment is dramatically reduced by calculating the DWFs by DFT, and expressing them as precomputed functions of the interatomic distances. A physically constrained global best fit is found through the use of the DE minimization algorithm.

DWFs on Zn-Cys sites are highly correlated on the first shell distances (Ref. 7), and erroneous DWF prediction leads to inaccurate structure determination. In this work we address this problem for Zn-Cys coordination. It is observed that for Zn-Cys active sites, DFT under the nonlocal spin density approximation, overestimates the Zn–S distances when compared with corresponding measurements from XRD; this effect is more severe when an effective basis set is employed. However, the pVDOS spectrum of the $ZnCys_4$ clusters reveals that vibrational frequencies of importance to SS and MS Zn-Cys paths are accurately predicted (compared to experimentally recorded Raman spectra) when an all-electron basis set is used; the use of this type of basis set also reduces the Zn–S distance overestimation error reported by DFT when compared to corresponding optimal geometries obtained with an effective core basis set. The choice of the nonlocal DFT functional plays a minor role on the accuracy of the structural and vibrational properties of Zn-Cys active sites. The B88PW91/DZVP2 is used to express SS/MS σ^2s of zinc coordinating to Cys residues; similar values could be obtained by the use of X3LYP/6-31G(TM)****+.

DWFs obtained in this work were tested against experimental XAFS spectra of $Zn(His)_{4-n}(Cys)_n$, $n=2-4$ active sites using a two-step process: (1) computed SS/MS $\sigma^2(R_{Zn-S})$ are coupled with DE and a hypothetical structure is obtained; (2) this structure was refined using the IFEFFIT program while constraining the DWFs to the predetermined values during the fitting procedure. Structural information obtained by IFEFFIT fits the XAFS data within the uncertainties, verifying the accuracy of our calculated DWFs, and the final distances agree with the structure determined by the DE algorithm. First shell distances R_{Zn-S} obtained using this method are within the corresponding XRD recorded values; however, R_{Zn-S} differs up to 0.006 Å with corresponding values reported by XAFS. Our approach can be potentially applied to other metal-Cys complexes.

ACKNOWLEDGMENTS

We would like to thank Professor James Penner-Hahn for providing us with the experimental XAFS spectra.

[1] J. Drenth, *Principles of Protein X-Ray Crystallography* (Springer-Verlag, New York, 1999).
[2] E. A. Stern, in *X-Ray Absorption*, edited by D. C. Koningsberger and R. Prins (Willey, New York, 1988), Chap. 1.
[3] P. A. Lee and J. B. Pendry, Phys. Rev. B **11**, 2467 (1975).
[4] S. S. Hasnain and K. O. Hodgson, J. Synchrotron Radiat. **6**, 852 (1999).
[5] J. Guo, S. Wang, J. Dong, H. Qiu, R. A. Scott, and D. P. J. Giedroc, J. Am. Chem. Soc. **117**, 9437 (1995).
[6] S. R. Hubbard, W. R. Bishop, P. Kirschmeier, S. J. George, S. P. Cramer, and W. A. Hendrickson, Science **254**, 1776 (1991).
[7] K. Clark-Baldwin, D. L. Tierney, N. Govindaswamy, E. S. Gruff, C. Kim, J. Berg, S. A. Koch, and J. E. Penner-Hahn, J. Am. Chem. Soc. **120**, 8401 (1998).
[8] H. Birkedal-Hansen, Curr. Opin. Cell Biol. **7**, 728 (1995).
[9] O. Kleifeld, P. E. Van den Steen, A. Frenkeli, F. Cheng, H. Liang Jiang, G. Opdenakker, and I. Sagi, J. Biol. Chem. **275**, 34335 (2000).
[10] E. A. Stern, Phys. Rev. B **48**, 9825 (1993).
[11] P. W. Loeffen and R. F. Pettifer, Phys. Rev. Lett. **76**, 636 (1996).
[12] P. W. Loeffen, P. F. Pettifer, and J. Tomkinson, Chem. Phys. **208**, 403 (1996).
[13] A. V. Poiarkova and J. J. Rehr, Phys. Rev. B **59**, 948 (1999).
[14] A. V. Poiarkova, "X-ray Absorption Fine Structure Debye-Waller Factors," Ph.D. Thesis, University of Washington, 1999.
[15] N. Dimakis and G. Bunker, J. Synchrotron Radiat. **6**, 266 (1999).
[16] N. Dimakis and G. Bunker, Phys. Rev. B **65**, 201103(R) (2002).
[17] N. Dimakis and G. Bunker, Phys. Rev. B **70**, 195114 (2004).
[18] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).
[19] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
[20] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, New York, 1989).
[21] T. Dudev and C. Lim, J. Phys. Chem. B **105**, 10709 (2001).
[22] M. Elstner, Q. Cui, P. Munih, E. Kaxiras, T. Frauenheim, and M. Karplus, J. Comput. Chem. **24**, 561 (2002).
[23] N. Dimakis and G. Bunker, Biophys. J. **91**, L87 (2006).
[24] R. Storn and K. Price, J. Global Optim. **11**, 341 (1997).
[25] J. Andzelm and E. Wimmer, Physica B **172**, 307 (1991).
[26] JAGUAR, Version 6.0, Schrodinger Inc., New York, N. Y. (2005).
[27] R. A. Friesner, Chem. Phys. Lett. **116**, 39 (1985).
[28] J. C. Slater, *Quantum Theory of Molecules and Solids* (McGraw-Hill, New York, 1974), Vol. 4.
[29] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[30] J. P. Perdew, in *Electronic Structure Theory of Solids*, edited by P. Ziesche and H. Eschrig (Akademie Verlag, Berlin, 1991).
[31] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).
[32] N. Godbout, D. R. Salahub, J. Andzelm, and E. Wimmer, Can. J. Chem. **70**, 60 (1992).
[33] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[34] X. Xu and W. A. Goddard, Proc. Natl. Acad. Sci. U.S.A. **101**, 2673 (2004).
[35] X. Xu, Q. Zhang, R. Muller, and W. A. Goddard, J. Chem. Phys. **122**, 14105 (2005).
[36] S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).
[37] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[38] P. J. Hay and W. R. Wadt, J. Chem. Phys. **82**, 299 (1985).
[39] R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).
[40] W. J. Hehre and J. A. Pople, J. Chem. Phys. **56**, 4233 (1972).
[41] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, and J. A. Pople, J. Chem. Phys. **77**, 3654 (1982).
[42] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).
[43] V. A. Rassolov, J. A. Pople, M. A. Ratner, and T. L. J. Windus, J. Chem. Phys. **109**, 1223 (1998).
[44] M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).
[45] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer, J. Comput. Chem. **4**, 294 (1983).
[46] A. P. Scott and L. Radom, J. Phys. Chem. **100**, 16502 (1996).
[47] M. Newville, J. Synchrotron Radiat. **8**, 322 (2001).
[48] A. L. Ankudinov, B. Ravel, J. J. Rehr, and S. D. Conradson, Phys. Rev. B **58**, 7565 (1998).
[49] M. Vargek, X. Zhao, Z. Lai, G. L. McLendon, and T. G. Spiro, Inorg. Chem. **38**, 1372 (1999).

[50] R. T. Witkowski, S. Hattman, L. Newman, K. Clark, D. L. Tierney, J. Penner-Hahn, and G. McLendon, J. Mol. Biol. **247**, 753 (1995).
[51] Y. Ye, M. Ruan, Y. Song, Y.-Y. Li, and W. Xie, Spectrochim. Acta, Part A **68**, 85 (2007).
[52] A. Pawlukojc, J. Laciejewicz, A. J. Ramirez-Cuesta, and J. Nowicka-Scheibe, Spectrochim. Acta, Part A **61**, 2474 (2005).
[53] T. E. Elgren and D. E. Wilcox, Biochem. Biophys. Res. Commun. **163**,1093 (1989).
[54] T. Simonson and N. Calimet, Proteins **49**, 37 (2002).
[55] T. Miura, T. Satoh, and H. Takeuchi, Biochim. Biophys. Acta **1384**, 171 (1998).
[56] N. Dimakis and G. Bunker, Phys. Rev. B **58**, 2467 (1998).
[57] S. Mangani, W. Meyer-Klaucke, A. J. G. Moir, M. Ranieri-Raggi, D. Martini, and A. Raggi, J. Biol. Chem. **278**, 3176 (2003).
[58] R. Bogumil, P. Faller, P. A. Binz, M. Vasak, J. M. Charnock, and C. D. Garner, Eur. J. Biochem. **255**, 172 (1998).
[59] J. Sun, J. Zhang, F. Wu, C. Xu, S. Li, W. Zhao, Z. Wu, J. Wu, C.-Z. Zhou, and Y. Shi, Biochemistry **44**, 8801 (2005).