

OPEN ACCESS

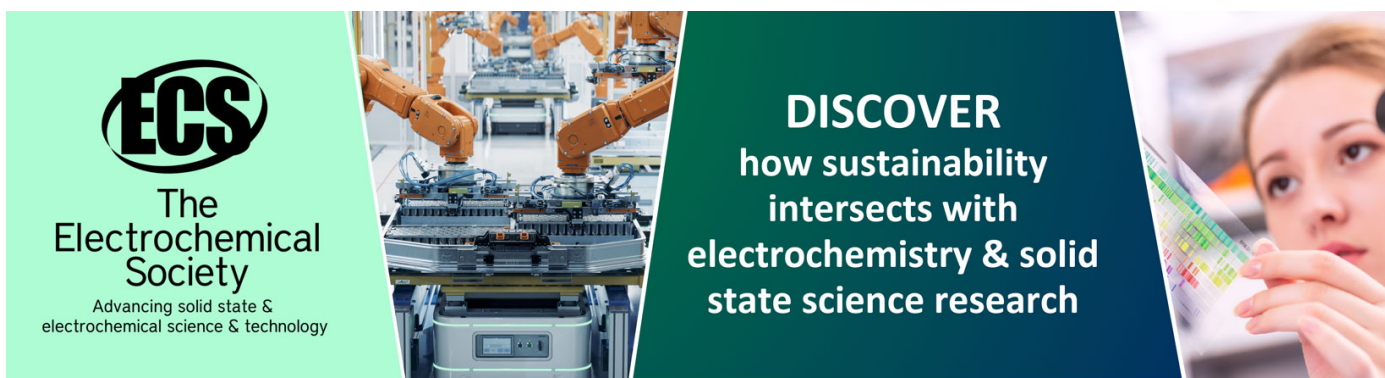
Composing complex EXAFS problems with severe information constraints

To cite this article: B Ravel 2009 *J. Phys.: Conf. Ser.* **190** 012026

View the [article online](#) for updates and enhancements.

You may also like

- [Bond compressibility and bond Grüneisen parameters of CdTe](#)
P Fornasini, R Grisenti, T Irifune et al.
- [Large negative thermal expansion of the Co subnetwork measured by EXAFS in highly disordered Nd_{1-x}Co_x thin films with perpendicular magnetic anisotropy](#)
J Díaz, R Cid, A Hierro et al.
- [Low-dimensional systems investigated by x-ray absorption spectroscopy: a selection of 2D, 1D and 0D cases](#)
Lorenzo Mino, Giovanni Agostini, Elisa Borfecchia et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Composing complex EXAFS problems with severe information constraints

B Ravel

National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899
USA

E-mail: bruce.ravel@nist.gov

Abstract. In recent work, a model for the structural environment of Hg bound to a catalytic DNA sensor was proposed on the basis of EXAFS data analysis. Although severely constrained by limited data quality and scant supporting structural data, a compelling structural model was found which agreed with a similar but less detailed model proposed on the basis on NMR data. I discuss in detail the successes and limitations of the analytical strategy that were implemented in the earlier work. I then speculate on future software requirements needed to make this and similarly complex analytical strategies more available to the wider audience of EXAFS practitioners.

1. Catalytic DNA sensors

Safe use of water that might be contaminated by natural and man-made sources requires monitoring technology that is easily transported to a field setting, that provides immediate feedback, that detects contaminants at or below regulatory limits, and that is selective of the target contaminant. Recently, catalytic DNA sensors meeting all of these criteria have been developed. [1, 2, 3] High sensitivity and selectivity is obtained via an active DNA sequence with high affinity for the target metal. Upon binding, the DNA is cleaved at the active site separating strands bound to a fluorophore and a quencher. Once separated from the quencher, the fluorophore emits a detectable photon.

Earlier one-dimensional ^{15}N nuclear magnetic resonance (NMR) studies [4, 5, 6] of the DNA sequence with an active site targeting Hg was interpreted to place the Hg atom binding site at a thymine–thymine (T–T) mismatch in the DNA duplex. In recent work, Extended X-Ray-Absorption Fine-Structure (EXAFS) was employed [7] to directly study the structural environment of Hg bound to the active site of the Hg-sensitive DNA complex. Using an interface to FEFF [8] and IFEFFIT [9] currently under development (with a working name of DEMETER [10]), the Hg atom was found to be bound to the corner of a six-member ring, such as the pyrimidine residue of the thymine nucleotide (as shown in Fig. 1).

In Ref. [7], we used the near edge spectrum to prove that the Hg is completely bound to the DNA complex. We then used a trial-and-error approach to identifying likely coordination environments for the Hg atom. Having identified probable binding locations between the Hg and the nucleotides, we applied constrained fitting models to the analysis of the EXAFS data to refine details of the local configurational environment about the Hg. To summarize the results of Ref. [7], we found that our data were consistent with the Hg bridging N atoms on the pyrimidine

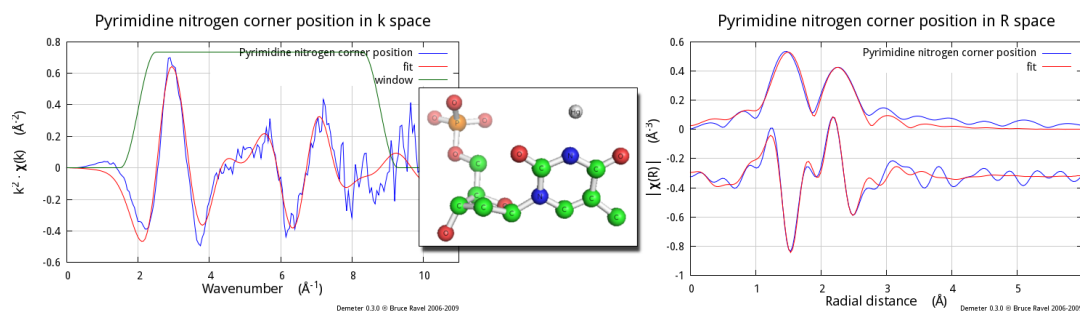


Figure 1. The reported structure in Ref. [7] has the Hg atom bound to the N atom in the pyrimidine residue of thymine. The fit is compared to the data as $\chi(k)$, $|\tilde{\chi}(R)|$, and $\text{Re}[\tilde{\chi}(R)]$.

rings of two thymine atoms, as suggested by the ^{15}N NMR. That result was not, however, unique in the sense that the EXAFS are consistent with having the Hg atom bridge two corner sites of any two six member rings. That is, the EXAFS data by itself cannot confirm that the Hg atom is bridging two thymine nucleotides, specifically.

Despite the inherent ambiguity in the result, a substantive contribution to the understanding of the chemistry of these catalytic DNA sensors was made with EXAFS. The analytical approach was, however, somewhat unusual and employed several elaborate techniques that have been developed around FEFF and IFEFFIT in recent years. In the next section, I will outline in substantial detail how the analysis presented in Ref. [7] was accomplished. In the following section, I will contrast this work with two especially admirable examples of EXAFS data analysis from the literature. I will conclude with a brief discussion of EXAFS analysis developments I would like to see in coming years.

This paper makes explicit reference to FEFF and IFEFFIT. Most of the ideas in this article are readily transferable to any analysis program based on FEFF. I hope that many of these concepts prove useful even to those parts of the EXAFS ecosystem not based on FEFF.

2. Analysis strategy for the catalytic DNA sensors

We approached the EXAFS analysis problem of Hg bound to the DNA complex with few preconceptions. We considered the prospect that the Hg atom could bind to almost any location on any of the four nucleotides. In Fig. 2, several high symmetry locations for decorating the thymine nucleotide with an Hg atom are shown. To simplify the discussion, I will introduce some nomenclature. The balls labeled with yellow question marks all have a carbon or oxygen atom bridging the Hg with the residue. These will be called *ligand positions*. The balls with purple question marks do not have an intervening atom and are called *corner positions*. Not shown in Fig. 2 are other possibilities, such as a position of symmetry above or below a ring or any position of lesser symmetry. The other three nucleotides can be similarly decorated.

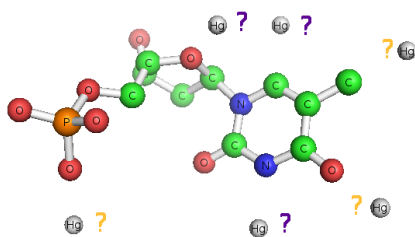


Figure 2. A ball-and-stick figure of the thymine nucleotide showing several positions for the placement of the Hg atom that were considered in the preliminary stage of the EXAFS analysis.

FEFF requires a list of atomic coordinates as the input data for its real-space, multiple-

scattering approach to EXAFS theory. One common way of providing this list is to use the ATOMS [11] program or a similar tool to generate that list from crystallographic data. In this case, ATOMS is of no use as the sample is a molecular complex in solution. Fortunately, the atomic configurations of the nucleotides are well known and easily found with an Internet search engine. The Hg atom can be placed in any location relative to an atomic configurations and Feff can be run using that as its input data.

Examining the Fourier transform shown in Fig. 1, the first peak position indicates a near neighbor distance of just over 2 Å. To make the problem of a more manageable scope, the Hg atoms were placed in positions of local symmetry. For ligand positions, the Hg atom was placed 2.04 Å from its nearest neighbor and at coordinates that made it collinear with the nearest and second nearest neighbors. For corner positions, the Hg atoms was again placed 2.04 Å from its nearest neighbor and at coordinates that placed it in the plane of the nearest and two second nearest neighbors and made it equidistant from the two second neighbors.

A throw-away program was written to read the files containing the nucleotide configurations, solve the coupled equations for Hg atoms placement, and export the decorated nucleotide in the form of a FEFF input file. Without atoms around the periphery of the cluster, the muffin tin radius [8] of the Hg may be poorly constrained. In practice, the short distance of the near neighbor atom sufficiently constrained the Hg atom. The Hg muffin tin radius for a decorated nucleotide was always around 1.38 Å, not too much larger than the 1.12 Å found by FEFF for HgO [12]. This was certainly a source of systematic error in the analysis.

Several of the positions considered provided unsatisfactory structural models when compared to the data, two of which are shown in Fig. 3. A ligand position, as on the left, resulted in a second neighbor distance that was too long to fit the second peak in the Fourier transformed $|\tilde{\chi}(R)|$ data. When a fit was attempted with any ligand position, the result was extremely poor due to this inappropriate path length.

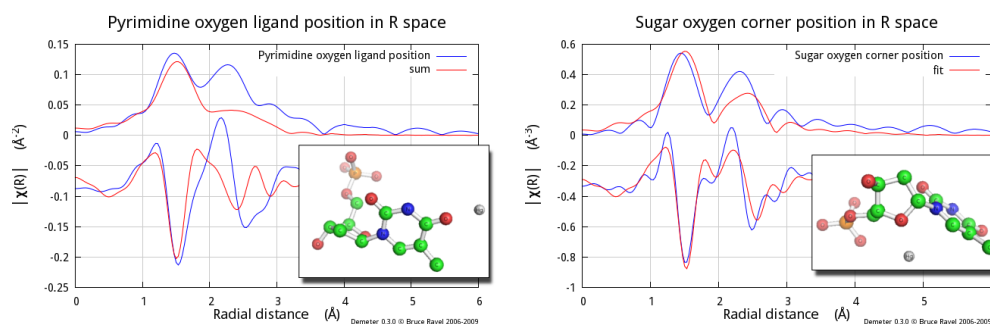


Figure 3. Two unsuccessful fitting models for Hg bound to the catalytic DNA complex. (Left) Summation of relevant paths for Hg in the oxygen ligand position on the thymine pyrimidine residue. (Right) Fit to Hg on a corner position of the thymine sugar residue.

The fit shown on the right side of Fig. 3 had the Hg atom at a corner position on the thymine sugar residue, which was a five-member ring. The details of the parameterization of this fit were quite similar to the more successful fitting model described below. The main failure of the corner position of a five-member ring was that the pentagonal angle was such that the distance to the second neighbor scatterer did not match the position of the second peak in $|\tilde{\chi}(R)|$. This result also precludes corner positions on the five-member sections of the purine residues of adenine and guanine. Positions above and below the rings were also considered, but also proved unsuccessful.

The most successful position for placement of the Hg atom was a corner position of a six-member ring, as shown in Fig. 1. The hexagonal angle of this ring type resulted in a second neighbor distance at the right length to match the second peak in the data. Parameterizing

this fit was made somewhat complicated by the limited information content of the signal. The Fourier transform window was from $2 \text{ \AA}^{-1}$ to $8.8 \text{ \AA}^{-1}$, as shown on the left of Fig. 1, while the fitting range was from $1 \text{ \AA}$ to $3.1 \text{ \AA}$, covering the first three peaks in $|\tilde{\chi}(R)|$. Using the Nyquist criterion [13], one can obtain at most 9 numbers from a fit to these data. The range in R-space included single scattering (SS) paths from the neighboring N, C, and O atoms as well as 7 non-trivial multiple scattering (MS) paths. Careful parameterization of this fitting problem was required to consider so many paths without over-extending the information content in the data.

The first two fitted parameters were the change in distance (ΔR) between the Hg atom and its nearest neighbor (the corner N atom in the case shown in Fig. 1) and the mean square displacement (σ^2) about that distance. Two more fitted parameters were needed for E_0 and for an overall amplitude term which includes both coordination number and S_0^2 . Even before considering the parameterization of the second and third neighbors or any of the MS paths, about half of the available information had been used.

To deal with this information limit, some simplifying assumptions about this structure were required. I asserted that the pyrimidine ring is very stiff compared the bond with the Hg atom. I then treated the ring as completely rigid in this fit while also maintaining the Hg atom at a site of relatively high local symmetry in which the two second neighbor atoms remain equidistant. To implement this restriction, I made use of IFEFFIT's capacity to apply mathematical constraints between variables of the fit. In this case, I related the change in the second neighbor distance to the variable parameter describing the change in the nearest neighbor distance.

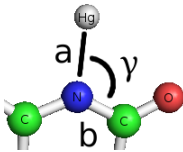


Figure 4. Computing the angle between the Hg–N bond and the adjacent bond on the pyrimidine ring.

The Hg atom and its neighbors are shown in Fig. 4. The distance a was a parameter of the fit. The distance b was fixed to its value from the atomic configuration that was used as the input to FEFF. The angle γ was also fixed and was simply computed from the input to FEFF. Given these assumptions, the distance between the Hg atom and its second neighbor, which is on the ring, was solved by the application of a common trigonometric relationship and without introducing any new variable parameters:

$$D_{(\text{Hg}-2\text{NN})} = \frac{(a - b) \cos(\gamma/2)}{\cos(\theta)} \quad \text{where} \quad \tan(\theta) = \frac{a + b}{a - b} \tan(\gamma/2)$$

Because the ring is assumed to be rigid, the σ^2 for the second neighbor was simply scaled up by the square of the ratio of the second and first neighbor distances. This is certainly a crude approximation, but it allows parameterization of the second neighbor's σ^2 without adding any additional variable parameters and with only a small contribution to the systematic error of the fitting model.

The third neighbor (the Hg–O distance in Fig. 4) could not be sensibly related to existing parameters, so additional ΔR and σ^2 parameters are introduced into the fit, bringing the total number of variable parameters to six. Armed with existing parameters and the assumption that the ring is rigid, ΔR parameters for each of the various MS paths were computed without using any additional variable parameters. Without detailed knowledge of force constants between the atoms, it is impossible to correctly account for MS σ^2 parameters without introducing new variable parameters. I assert that approximating the MS σ^2 parameters using the existing SS σ^2 parameters was adequate and yielded superior results compared to a fit made by introducing new variable parameters for the MS σ^2 s.

In this way, ten SS and MS paths were used in the fit and were parameterized using only six fitting variables. As reported in Ref. [7], all parameters returned sensible values with reasonable uncertainties. The EXAFS data were found to be consistent with the structural model proposed by the authors of the earlier NMR work. Although that was not a unique solution, the EXAFS data certainly demonstrate that the Hg atom is at a corner position of a six-member ring. There are many shortcomings to this analytical approach, even beyond the approximations made for σ^2 parameters. The possibility of scattering from surrounding water molecules was not included, nor was the possibility of MS involving atoms in the two nucleotides and the bridging Hg atom. The assumptions of a rigid ring and Hg placement at a site of relatively high symmetry can both be questioned. Finally, the lack of peripheral atoms constraining the Hg muffin tin radius was a potential fault of the FEFF calculation used as the basis of the fit.

My answer to all these issues is the same. The model presented here is simple enough to use the limited information in the data well while still providing enough structural content to adequately describe the data. Both the parameterization and the details of the FEFF calculation were sufficiently easy both to implement and to automate for testing against many decorations of all four nucleotides. In the end, I am able to assert that the Hg atom is at a six-member ring corner position. This is consistent with the model from the NMR work. Given that a T-T mismatch, as this DNA complex is known to have, would lack a Watson-Crick hydrogen bond between the thymine bases, it is a chemically plausible location for the Hg to bond. By creatively parameterizing the fitting model and invoking physically and chemically plausible constraints on the fit, I was able to make progress on a problem suffering from both limited data quality and an ambiguous physical structure.

3. Other complex analysis problems

There are many examples of EXAFS data analysis in the literature in which the authors made progress in the face of challenges as great as those posed by the Hg/DNA problem. I will briefly discuss two such examples before concluding with speculation on how EXAFS data analysis software might develop to make such sophisticated analysis strategies available to and tractable for a broader audience.

3.1. Uranyl ion adsorbed to biomass

Ref. [14] presents an EXAFS study of the pH dependence of uranyl adsorption to bacterial biomass. As a function of pH, the proportionation of the uranyl ion among phosphoryl, carboxyl, and hydroxyl binding sites were shown to be pH-dependent. These ligand types are shown schematically in Fig. 5. Each ligand type has a characteristic U–O distance as well as characteristic distance to the second nearest neighbor (P for phosphoryl or C for carboxyl). The fit was determined using a model in which the relative amounts of these three ligand types are parameters of the fit. The contributions from the phosphoryl and carboxyl ligands were obtained by running FEFF on common uranyl phosphate and uranyl carbonate crystals, then using only those paths required to describe the structural unit of each ligand type. In short, a subset of a FEFF calculation is used to describe each structural component in the fit.

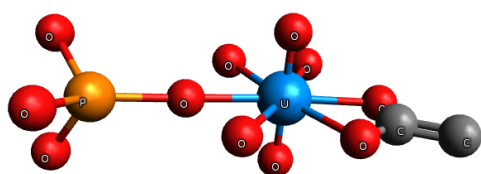


Figure 5. A schematic representation of the different ligation environments of the uranyl ion adsorbed to bacterial biomass. The phosphoryl ligand is monodentate to the uranyl ion, while the carboxyl ligand is bidentate. A hydroxyl ion is indistinguishable in EXAFS from one of the dangling equatorial oxygens.

The theory for each of these structural units resulted from independent FEFF calculations. These were plugged into the fit and parameterized for values of ΔR and σ^2 (making use of sensible physical constraints on those parameters). Only two more variable parameters related to coordination numbers were required to determine proportionation among the three ligand types.

3.2. Anti-site disorder in mixed ferrites

Ref. [15] shows an EXAFS study of a spinel structured ferrite in which the two metal cation sites are populated by three different metals, Mn, Zn, and Fe. Nominally, these cations segregate onto the two different metal sites. However, the free energy barrier to anti-site disorder is very low. In a real sample, either metal site can be populated by any of the three cations. To fully interpret the EXAFS data, the possibility of finding any cation on any site must be considered, both as absorber and as scatterer. Consequently, several FEFF calculations were required for fitting EXAFS data measured with each absorber species. The calculations must be made for the absorber on either cation site and additional calculations might be required to account for the possibility of any of the three scatterers on each neighboring metal site.

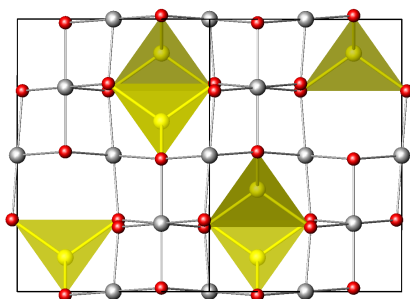


Figure 6. The spinel structure along the $a + b$ direction. The A cations (nominally Mg and Zn) are the yellow balls, the B cations (nominally Fe) are the grey balls, and the O atoms are the red balls. The data analysis in Ref. [15] co-refined EXAFS data from all three cations and considered the possibility of sites swapping between the cation sites.

In Ref. [15], data were measured on all three edges and refined simultaneously. A full complement of FEFF calculations were made for each edge. This large ensemble of data and theory was considered for each to each sample. Along with parameterizing ΔR and σ^2 for each SS and MS path, fitting parameters were used which considered the probability of finding a particular cation on a particular metal site. Because data from all three edges were co-refined, a large number of physically reasonable constraints were applicable and used in the fitting model.

3.3. Implementing complex fitting models

Both of the examples summarized in this section share several salient features with the analysis of the Hg bound to the DNA complex. In any analysis involving FEFF and IFEFFIT, a suitable initial guess for the atomic configuration must somehow be found. In each case, this initial guess was ambiguous. Indeed, determining that configuration was the very purpose of the EXAFS measurement. In each case, one or more FEFF calculations were made and carved up to select those paths required to understand the EXAFS data in question.

The second thing each example shares is the need to cleverly parameterize each path used in the fit, making use of physically reasonable constraints on those parameters. Fortunately, the application of these constraints is a natural part of the use of IFEFFIT. Because any EXAFS analysis problem is an information-limited problem (even the spinel example, with its co-refinement of data from *three* edges), it is simply impractical to expect to independently determine coordination, ΔR , and σ^2 for every SS and MS path used in the fit. These constraints are an essential part of building the fitting model for any EXAFS problem.

There is no fundamental reason that elaborate modeling of the sort discussed here is not more common in the EXAFS literature. The techniques discussed here – intelligent trial and

error in the case of Hg bound to DNA, consideration of structural units as scattering elements in the case of uranyl adsorption, or propagation of the effects of site-swapping through SS and MS paths in the case of the mixed ferrite – could be widely applicable to other problems of EXAFS analysis. And yet, such elaborate modeling remains more exception than rule.

The principle limitation is one of software capabilities. Using FEFF and IFEFFIT or most other software packages to implement any of these clever modeling schemes requires considerable familiarity with the details of the tools. Even armed with an expert's knowledge, implementation of elaborate models remains labor intensive and error prone. Most EXAFS practitioners use EXAFS as one of many tools in pursuit of solutions to scientific problems. Those users rightly depend on a user interface to obviate the very details that must currently be understood in order to implement the more sophisticated data analysis strategies. Until software tools are developed which enable higher level modeling concepts, those sophisticated strategies will remain an overt challenge for a large portion of our community.

While good software tools already exist at the foundation of the EXAFS analysis problem – data processing and the handling of the components of the FEFF calculation – very high level tools are scarce outside of specialized or one-off codes written by and for experts. To my knowledge, there has been scant investigation of how machine learning could be implemented in pursuit of EXAFS analysis solutions.

I hope that this paper might provide a launching point for discussing future trends in EXAFS data analysis.

Acknowledgments

I thank the staff at APS beamline 20BM – M. Balasubramanian, R. Gordon, D. Brewé, and S. Heald – for their support during these measurements; my colleagues from University of Illinois, Urbana-Champaign – S. Slimmer, M. Wang, G.C.L. Wong, and Y. Lu – for introducing me to this fascinating field and for the collaboration which resulted in Ref. [7]; S. Kelly and S. Calvin for years of conversation about EXAFS analysis strategies and for their fine papers cited herein; M. Newville for IFEFFIT which makes all this madness possible; and J.J. Rehr and his students for FEFF, the other thing that makes all this madness possible.

References

- [1] Li J and Lu Y 2000 *J. Am. Chem. Soc.* **122** 10466–10467
- [2] Liu J, Brown A K, Meng X, Cropek D M, Istok J D, Watson D B and Lu Y 2007 *PNAS* **104** 2056–2061
- [3] Liu J and Lu Y 2007 *Angewandte Chemie (International ed. in English)* **46** 7587–7590
- [4] Miyake Y, Togashi H, Tashiro M, Yamaguchi H, Oda S, Kudo M, Tanaka Y, Kondo Y, Sawa R, Fujimoto T, Machinami T and Ono A 2006 *J. Am. Chem. Soc.* **128**
- [5] Tanaka Y, Yamaguchi H, Oda S, Kondo Y, Nomura M, Kojima C and Ono A 2006 *Nucleosides, Nucleotides, and Nucleic Acids* **25** 613–614
- [6] Tanaka Y, Yamaguchi H, Oda S, Kondo Y, Kojima C and Ono A 2007 *J. Am. Chem. Soc.* **129**
- [7] Ravel B, Slimmer S C, Meng X, Wong G C L and Lu Y 2009 *Radiat. Phys. Chem.* DOI: <http://dx.doi.org/10.1016/j.radphyschem.2009.05.024>
- [8] Zabinsky S I, Rehr J J, Ankudinov A, Albers R C and Eller M J 1995 *Phys. Rev. B* **52** 2995–3009
- [9] Newville M 2001 *J. Synchrotron Radiat.* **8** 322–324
- [10] Ravel B Demeter: A comprehensive system for processing and analyzing x-ray absorption spectroscopy data see <http://cars9.uchicago.edu/iffwiki/Demeter>
- [11] Ravel B 2001 *J. Synchrotron Radiat.* **8** 314–316
- [12] Aurivillius K 1964 *Acta Chem. Scand.* **18** 1305–1306
- [13] Brillouin L 1962 *Science and Information Theory* (New York: Academic Press)
- [14] Kelly S, Kemner K M, Fein J B, Fowle D A, Boyanov M I, Bunker B A and Yee N 2002 *Geochem. Cosmo. Acta* **66** 3875–3891
- [15] Calvin S, Carpenter E E, Ravel B, Harris V G and Morrison S A 2002 *Phys. Rev. B* **66** 224405