

A method for normalization of X-ray absorption spectra

Tsu-Chien Weng, Geoffrey S. Waldo and James E. Penner-Hahn

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

A method for normalization of X-ray absorption spectra

Tsu-Chien Weng, Geoffrey S. Waldo[‡] and James E. Penner-Hahn*

Department of Chemistry and Biophysics Research Division, University of Michigan, Ann Arbor, MI 48109-1055, USA. E-mail: jeph@umich.edu

Accurate normalization of X-ray absorption data is essential for quantitative analysis of near-edge features. A method, implemented as the program *MBACK*, to normalize X-ray absorption data to tabulated mass absorption coefficients is described. Comparison of conventional normalization methods with *MBACK* demonstrates that the new normalization method is not sensitive to the shape of the background function, thus allowing accurate comparison of data collected in transmission mode with data collected using fluorescence ion chambers or solid-state fluorescence detectors. The new method is shown to have better reliability and consistency and smaller errors than conventional normalization methods. The sensitivity of the new normalization method is illustrated by analysis of data collected during an equilibrium titration.

Keywords: X-ray absorption spectroscopy; normalization; *MBACK*.

1. Introduction

X-ray absorption spectroscopy (XAS) is often divided into extended X-ray absorption fine structure (EXAFS) and X-ray absorption near-edge structure (XANES) regions. Over the last two decades, EXAFS has come to be recognized as one of the most useful tools for probing the local structure of metal ions in non-crystalline systems. As a consequence, data-reduction strategies for XAS spectra have generally been optimized to obtain EXAFS spectra, and have thus focused on obtaining background-free post-edge data. The XANES region contains important chemical and structural information that complements that obtained from EXAFS; in order to fully utilize the XANES region, where the chemically relevant differences may be quite subtle, it is important to obtain carefully normalized XANES spectra. Many of the data-reduction methods used for XANES data have been adapted from EXAFS data-reduction strategies. In the following, we demonstrate that this can, in some cases, lead to significant errors in normalization. We describe an alternative approach, and show that this yields significantly improved data normalization.

X-ray absorption data reduction typically begins with fitting a first- or second-order polynomial to the pre-edge portion of the data. A first-order polynomial is often appropriate for transmission data, while a higher-order polynomial may be required if there is more curvature to the background, for example in fluorescence data collected using an energy-resolving detector. The pre-edge polynomial is extrapolated throughout the entire energy range of the data to give 'pre-

edge subtracted' data. The pre-edge subtracted data typically are fit with a spline background function of some sort over the EXAFS, *i.e.* the post-edge, region. The appropriate scaling factor is determined by setting an absorbance to 1.0 (Teo, 1986). Often the value of the spline background function at the absorption-edge energy is used to set the scaling factor, although some authors define the maximum in the absorption coefficient near the edge as 1.0.

For EXAFS data, it is most important to obtain background-free post-edge absorbance. For this purpose, the accuracy of the pre-edge function has little consequence, since small errors in the scale factor will result in, at most, a small error in the apparent coordination number. The typical uncertainty in scale factor ($\sim 5\%$) is negligible in comparison with other systematic errors in EXAFS amplitudes.

In contrast, XANES data can be significantly perturbed by this data-reduction process. Inaccurate extrapolation of the pre-edge polynomial can result in the introduction of curvature into the edge region. This can be important when comparing data collected with very different backgrounds, for example, when comparing transmission and fluorescence data. In principle, it should be possible to avoid such problems by careful choice of the pre-edge background function. Thus, the operator can tailor the edge region and the polynomial order until the best possible normalization is obtained. The problem with this approach is that it is subject to errors depending on the skill of the experimenter and that it is thus susceptible to unintentional operator bias, since it lacks an objective definition of the 'best' pre-edge function. The approach that we describe below provides an automatic solution to normalization, with an objectively defined criterion for the 'best' pre-edge function. A second difficulty with the standard normal-

[‡] Current Address: MS-M888, Structural Biology Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

ization procedure is that the details of the post-edge spline function, which typically must be extrapolated to determine the scaling factor, are sensitive to minor variation in polynomial order or the choice of the energies of the spline knots. This can make it difficult to normalize data having very different EXAFS oscillations. For careful comparison of spectra, even minor variations in scaling can be critical.

2. Methodology

2.1. XAS

Spectra were collected at Stanford Synchrotron Radiation Laboratory, on beamline 9-3 using a Si(220) double-crystal monochromator with a focusing mirror for harmonic rejection. Samples of $[(\text{CH}_3)_4\text{N}]_2\text{Zn}(\text{SC}_6\text{H}_5)_4$ were dissolved in freshly distilled DMSO under anaerobic conditions. Samples were prepared and injected into a sample cell in a glove box under nitrogen atmosphere. Zn *K*-edge XAS data were measured at room temperature as fluorescence excitation spectra, using a 30-element Ge solid-state detector equipped with a 3 absorption path-length copper filter and Soller slits focused on the sample. X-ray energies were defined with respect to a zinc foil measured at the same time as the data, with the first inflection point of the zinc foil assigned as 9660.7 eV.

2.2. Conventional normalization

A second-order polynomial was fitted to the raw data in the pre-edge region (9355–9635 eV) and extrapolated through the post-edge region. A three-region cubic spline with k^3 -weighting was fitted to the post-edge region of the pre-edge subtracted data, with spline knots at 9680.0, 9807.6, 9935.1 and 10062.7 eV. The edge jump was defined by extrapolating the spline background to the edge energy (9680 eV) and multiplying the background-subtracted data by a scale factor to give an edge jump of 1.0. This is shown schematically in Figs. 1(a)–1(c). The results shown in this manuscript were obtained using EXAFSPAK (George *et al.*, 2001). However, the results do not depend strongly on the program package that is used. We obtain comparable, and in many cases worse, normalization using other EXAFS analysis packages. Figs. 1(a)–1(c) represent a worst-case scenario with a quadratic pre-edge function chosen to give a flat pre-edge background. Better normalization is possible, but only with subjective intervention by the experimenter.

2.3. MBACK

Many of the problems with conventional normalization methods arise from the need to extrapolate both the pre-edge and post-edge background functions. This can result, for example, in significant curvature in the post-edge region (see Fig. 1b). To avoid the need for extrapolation, we require that a single smooth background function be used over the entire data range. To avoid introducing distortions into the data owing to the XANES structure, we do not fit the background over the edge region, defined for these purpose as extending

from 20 eV below to 80 eV above the edge jump, and shown by the vertical lines in Figs. 1(d) and 1(e).

The background function and a single scale factor are adjusted to give the best fit between the normalized data and the tabulated X-ray absorption cross sections. In order to facilitate comparison between data from different samples, the tabulated cross sections that are used as a reference are those for the absorbance by the appropriate absorption edge. Reference data were taken from tabulated X-ray cross sections (McMaster *et al.*, 1969), as shown in Fig. 1(e). Other reference functions can be used; we chose the McMaster functions because they are readily available as tabulated absorption cross sections. The function that was minimized is given in equation (1),

$$\frac{1}{n_1} \sum_{i=1}^{n_1} [\mu_{\text{tab}}(E_i) + \mu_{\text{back}}(E_i) - s\mu_{\text{raw}}(E_i)]^2 + \frac{1}{N - n_2 + 1} \sum_{i=n_2}^N [\mu_{\text{tab}}(E_i) + \mu_{\text{back}}(E_i) - s\mu_{\text{raw}}(E_i)]^2, \quad (1)$$

where the index i refers to the point number, and n_1 and n_2 are the point numbers corresponding to energies below and above the edge (vertical lines in Figs. 1d and 1e). In equation (1), μ_{tab} is the tabulated absorption coefficient, μ_{back} is the calculated background function, μ_{raw} is the measured absorption coefficient and s is a scale factor. We have found it important to calculate the two halves of equation (1) separately, summing the mean-square deviation below the edge and the mean-square deviation above the edge. Without this, the post-edge region, with many data points, dominates the minimization, thus leading to a significantly worse fit of the background function to the data below the edge. In principle, it would be possible to assign different weights to the two halves of equation (1). However, we have not found this necessary in order to obtain good fits. In order to account for the characteristic rapidly decreasing pre-edge shape that results from residual elastic scatter when using a solid-state fluorescence detector, the MBACK background function includes a complementary error function. This is centered at the energy of the X-ray emission line of the absorbing element (E_{em}), with the spectral width (ξ) and the amplitude (A) treated as variable parameters. The remainder of the background was constructed using a series of Legendre polynomials centered at the energy of the X-ray absorption edge (E_{edge}) up to the m th order (usually $m = 2$ –3), giving a total of $m + 4$ adjustable parameters ($m + 1$ polynomial coefficients, A , ξ and s),

$$\mu_{\text{back}}(E) = \left[\sum_{i=0}^m C_i (E - E_{\text{edge}})^i \right] + A \operatorname{erfc} \left(\frac{E - E_{\text{em}}}{\xi} \right). \quad (2)$$

3. Results

3.1. Reliability

We define reliability as the ability to reliably extract the absorption data regardless of the details of the background

function. In order to evaluate the reliability of *MBACK*, three background functions, associated with different detection methods, were created. These were added to one set of normalized data (Fig. 2, line *a*) in order to mimic the raw data taken with an ion-chamber (monotonically decreasing background, line *b*), a fluorescence ion-chamber (monotonically increasing background, line *c*) or a solid-state Ge detector (exponentially decreasing in pre-edge region and flat or monotonically increasing in post-edge region, line *d*). The conventional normalization method gave normalized data (Figs. 3*a* and 3*b*) that match at the edge position, but which show approximately 5% variation in magnitude of the near-edge region, depending on the background functions. In contrast, the original normalized data were reliably recovered by *MBACK* to within the width of the lines (Figs. 3*c* and 3*d*). This result shows that *MBACK* is insensitive to detection methods, and therefore should facilitate comparison of data sets collected under different conditions.

3.2. Reproducibility

The reliability tests show the ability of *MBACK* to recover a synthetic spectrum after deliberate addition of a background function, but do not address the reproducibility of the method when used with real spectra. In order to address reproducibility, both normalization methods were applied to two replicate sets of data. Data were measured during titration of a 5 mM solution of $[(\text{CH}_3)_4\text{N}]_2\text{Zn}(\text{SC}_6\text{H}_5)_4$ in DMSO (Tobin, 2003). These data arguably provide a more realistic test of *MBACK* than do the synthetic spectra shown in Fig. 2. The dilute Zn complexes have significant noise, and show the small sample-to-sample background variation that is likely to be found in many XANES experiments. Duplicate spectra were measured for two samples, one with and one without added thiolate, as part of a study aimed at characterizing the affinity of the thiolate for the Zn. The conventional normalization method gave apparently reasonable results (Fig. 4*a*). The spectrum-to-spectrum variation is much smaller than that seen in Fig. 3, as expected given the similar background function for the four spectra. However, on closer examination (Fig. 4*a*, inset) it is apparent that the normalized spectra still show $\sim 5\%$ variation in amplitude. In contrast, *MBACK* gives normalized spectra that show negligible variation between spectra for duplicate preparations of a sample (Fig. 4*b*, inset). With this improved reproducibility, it is possible to detect

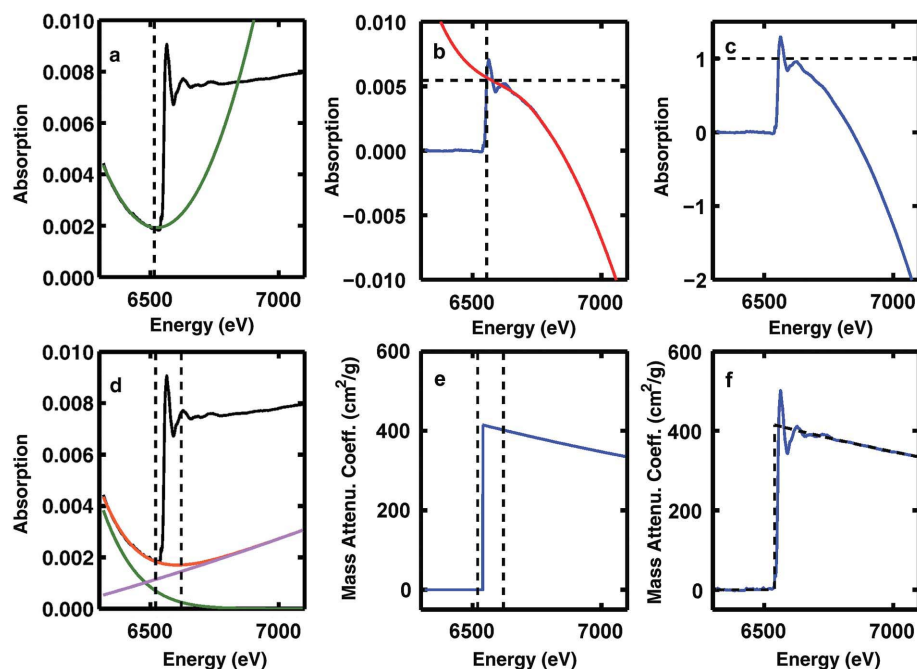


Figure 1

Schematic illustration of conventional normalization (*a–c*) and *MBACK* normalization (*d–f*). (*a*) Measured data (black) are first fitted by a pre-edge polynomial (green) which is then extrapolated over the data range. The vertical line marks the end of the pre-edge fit region. (*b*) A post-edge spline (red) is then fitted to the pre-edge subtracted data (black) and extrapolated to E_0 (vertical line) in order to determine the appropriate scale factor. (*c*) Conventionally normalized data. (*d*) In the *MBACK* procedure, a single background function (red), consisting of a complementary error function (green) plus a polynomial (pink), is used to fit the data both below and above the edge in order to maximize the agreement with the corresponding region of tabulated X-ray cross section, shown in (*e*). The vertical lines in (*d*) and (*e*) mark the edge region that is excluded from the fit. (*f*) *MBACK* normalized spectrum.

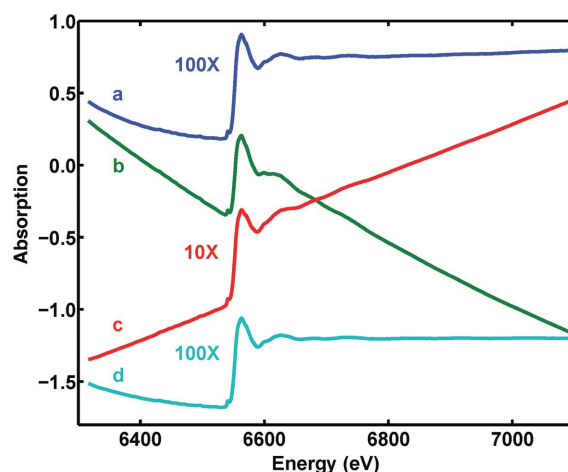


Figure 2

Raw data with synthetic background functions. Different synthetic background functions were added to the original data (*a*, blue) to mimic data collected in transmission mode with ion chambers (*b*, green), in fluorescence mode with a fluorescence ion-chamber detector (*c*, red) or in fluorescence mode with a solid-state fluorescence detector (*d*, cyan). Vertical scales are arbitrary, but have been chosen to correspond to typical measurement conditions.

sample-dependent variations that were obscured by the conventional data analysis. Only with this enhanced reproducibility is it possible to extract the chemically significant variations in these data (Tobin *et al.*, 2005).

3.3. Stability

The normalization in Figs. 2–4 used data that were measured over a wide range, from approximately 300 eV below the edge to approximately 400 eV above the edge. However, XANES spectra are often measured over much more limited energy ranges. In order to evaluate the stability of *MBACK*, defined as the ability to produce the same normalized spectrum regardless of the energy range that was measured, one of the data sets in Fig. 4 was used to check the dependence of the normalization on the energy range of the

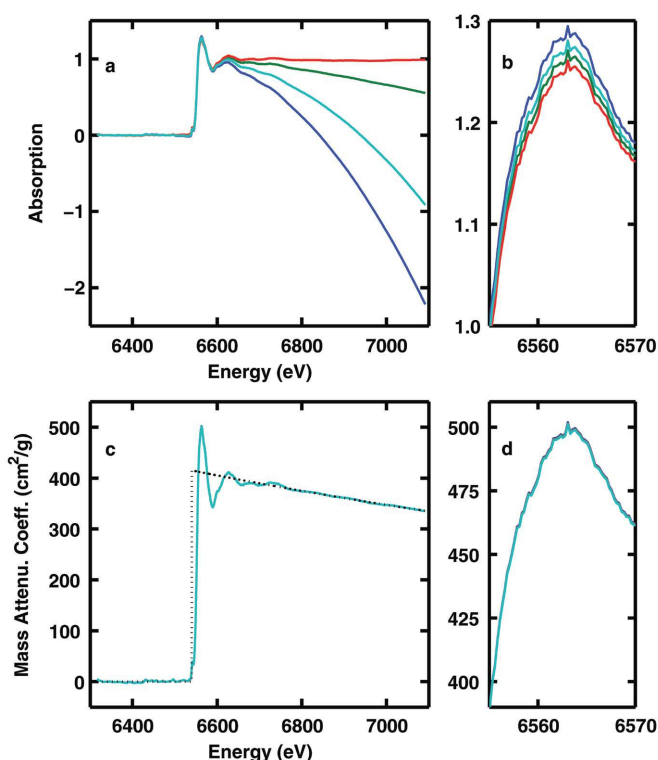


Figure 3
Normalization results for the synthetic spectra in Fig. 2. Conventional method (*a–b*) and *MBACK* (*c–d*). Color code matches that in Fig. 2: reference spectrum (blue), transmission data (green), fluorescence ion-chamber (red), solid-state fluorescence detector (cyan).

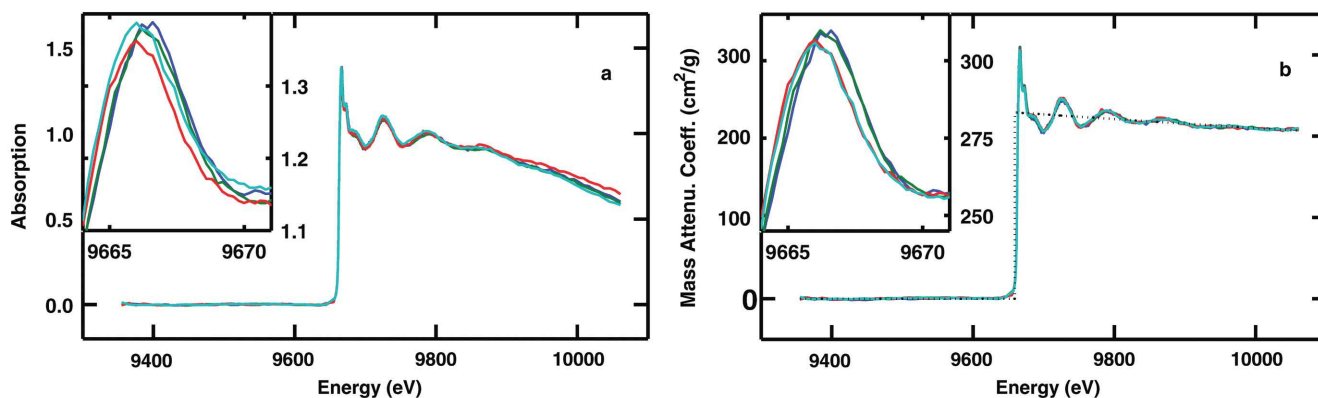


Figure 4
Normalization results of two independently measured data sets by conventional method (*a*) and *MBACK* (*b*). Blue and green lines are the normalized spectra for two independent samples of 5 mM $[(\text{CH}_3)_4\text{N}]_2\text{Zn}(\text{SC}_6\text{H}_5)_4$ in DMSO. Red and cyan lines are the normalized spectra for 5 mM $[(\text{CH}_3)_4\text{N}]_2\text{Zn}(\text{SC}_6\text{H}_5)_4 + 200 \text{ mM } [(\text{CH}_3)_4\text{N}]\text{SC}_6\text{H}_5$ in DMSO. Only with the *MBACK* normalization is the variation between duplicate samples small in comparison with the chemically relevant spectral changes.

data. The data set was truncated at energies as low as 9800 eV. The conventional normalization was extremely sensitive to the end point of the data set (Figs. 5*a* and 5*b*). In contrast, *MBACK* gave identical normalized spectra provided that the end point was ~ 200 eV or more above the edge (Figs. 5*c* and 5*d*).

3.4. Error estimation

If we assume that each pair of replicate samples in Fig. 4 should give identical normalized spectra, we can use these

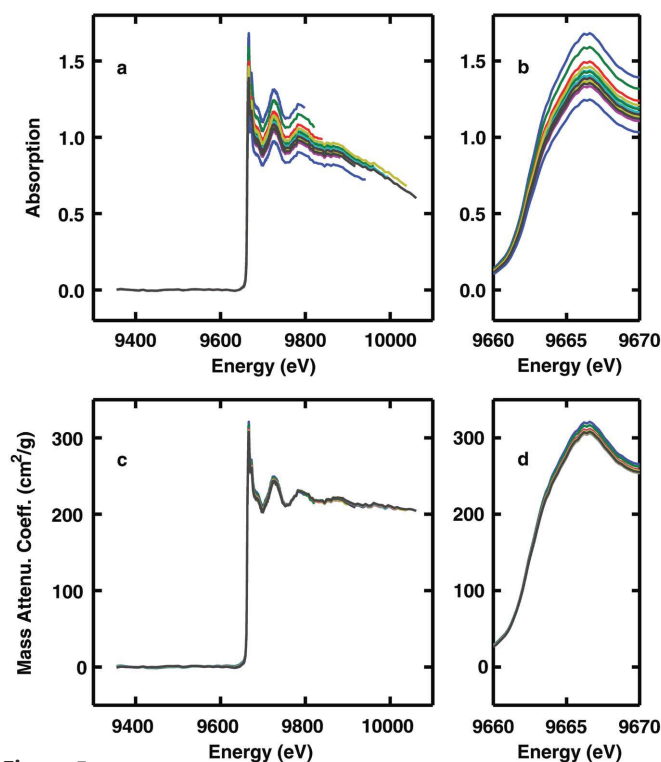


Figure 5
Normalized spectra as a function of data range for conventional normalization (*a–b*) and *MBACK* (*c–d*). The original data file extended to 10060 eV. To test the dependence of the normalization on energy range, the file was truncated at energies ranging from 9800 eV to 10040 eV and then normalized.

spectra to estimate the uncertainty of *MBACK*. Four different pairs of spectra can be used to calculate the difference spectrum for the Zn complex with or without added thiolate. These four difference spectra should, in principle, be identical, and should reflect the spectral changes caused by thiolate addition. The standard deviation of the four difference spectra, σ_{total} , will reflect two sources of uncertainty in the data: statistical error in the measured data, σ_{stat} , and the uncertainty introduced by the normalization method, σ_{norm} . The observed standard deviations over the edge region (9664–9680 eV), normalized to an edge jump of 1.0, are $\sigma_{\text{total}} = 0.0132$ for the conventional normalization method and $\sigma_{\text{total}} = 0.0042$ for *MBACK*. Based on the number of measured fluorescence counts, the expected statistical uncertainty in each spectrum is 0.0025, giving $\sigma_{\text{stat}} = 0.0035$ for the difference spectra. Consequently we can estimate that, at least for these data, the uncertainties introduced by the conventional normalization method and by *MBACK* are $\sigma_{\text{norm}} = 0.0127$ (1.3%) and $\sigma_{\text{norm}} = 0.0023$ (0.2%), respectively, using $\sigma_{\text{total}}^2 = \sigma_{\text{stat}}^2 + \sigma_{\text{norm}}^2$.

4. Discussion

We have shown that the *MBACK* algorithm is reliable and reproducible, and, providing a sufficient range of data is available, is mathematically stable. To the extent that the McMaster tables that are used for reference are accurate, *MBACK* can also provide accurate absorption coefficients. However, for most applications, accuracy is less important than precision. We have shown that *MBACK* provides sufficient precision to distinguish chemically induced spectral changes that are obscured by conventional normalization procedures. Beyond detailed comparison of XANES spectra, *MBACK* should be useful for other analyses that depend on careful edge normalization. One such example is the moment method for determining edge energies (Alp *et al.*, 1989). This approach shows promise for avoiding the sensitivity of energy (defined as the first inflection point) to the shape of the edge, thus providing a more reliable determination of the oxidation state of metalloproteins (DeMarois, 1999). However, the moment method requires integration across the edge, and is thus sensitive to the details of the normalization. To avoid this sensitivity, Iuzzolino *et al.* (1998) have limited integrations to

the immediate vicinity of the edge. While avoiding the sensitivity to normalization, this restores the sensitivity to edge shape, thus undermining many of the advantages of the moment method. *MBACK* should avoid this limitation. Another application where *MBACK* should prove useful is in correcting data for self-absorption (Goulon *et al.*, 1982; Pickering *et al.*, 2001). In order to correct the spectrum of a thick concentrated sample for the effects of self-absorption, it is necessary to find the absorption spectrum that, when subjected to self-absorption, will give the observed spectrum. This is simplified when reliably normalized spectra are available (Waldo, 1992).

We have implemented the *MBACK* algorithm in a *MATLAB* program, which is available on request.

This work was supported in part by the National Institutes of Health grant GM-38047 to JEPH. Zn X-ray absorption spectra were measured at the Stanford Synchrotron Radiation Laboratory, which is supported by the US DOE and the NIH Research Resource program.

References

- Alp, E. E., Goodman, G. L., Soderholm, L., Mini, S. M., Ramanathan, M., Shenoy, G. K. & Bommannavar, A. S. (1989). *J. Phys. Condens. Matter*, **1**, 6463–6468.
- DeMarois, P. S. (1999). PhD thesis, The University of Michigan, USA.
- George, G. N., George, S. J. & Pickering, I. J. (2001). *EXAFSPAK*, <http://ssrl.slac.stanford.edu/exafspak.html>.
- Goulon, J., Goulon-Ginet, C., Cortes, R. & Dubois, J. M. (1982). *J. Phys.* **43**, 539–548.
- Iuzzolino, L., Dittmer, J., Dorner, W., Meyer-Klaucke, W. & Dau, H. (1998). *Biochemistry*, **37**, 17112–17119.
- McMaster, W. H., Del Grande, N. K., Mallett, J. H. & Hubbell, J. H. (1969). *Compilation of X-ray Cross Sections*. Lawrence Radiation Laboratory, Livermore, California, USA.
- Pickering, I. J., George, G. N., Yu, E. Y., Brune, D. C., Tuschak, C., Overmann, J., Beatty, J. T. & Prince, R. C. (2001). *Biochemistry*, **40**, 8138–8145.
- Teo, B. K. (1986) *EXAFS: Basic Principles and Data Analysis*. New York: Springer-Verlag.
- Tobin, D. A. (2003). PhD thesis, The University of Michigan, USA.
- Tobin, D. A., Weng, T.-C., Karabiyik, M. & Penner-Hahn, J. E. (2005). In preparation.
- Waldo, G. S. (1992). PhD thesis, The University of Michigan, USA.