Overview of the standard EXAFS data analysis procedure

Introduction

Interpretation of EXAFS data is normally based on the EXAFS equation

$$\chi_i(k) = \sum_j \frac{N_j B_j(k)}{r_j^2} e^{-2r_j/\lambda(k)} e^{-2k^2\sigma_j^2} \sin\left(2kr_j + \delta_j(k)\right), \qquad [1]$$

with appropriate accomodation for forward scattering ("focussing effect") for higher shells, and inclusion of higher cumulants[1,2] as necessary. The aim of data analysis is to determine the distances, coordination numbers, disorder parameters and types of atoms in the various coordination shells for the "unknown" sample. The extent to which this is possible depends on the nature of the system, the quality of standards, the quality of the data.

The scheme of data analysis is very much intertwined with the experimental methods. The standard sequence of operations consists of: normalization of the data to unit edge step; interpolation to k-space; fourier transformation to r-space; isolation of the amplitude and phase of individual coordination shell windowing and inverse transformation; and analysis of the amplitudes and phase using the EXAFS equation by nonlinear least squares fitting, the cumulant expansion/ratio method, and beat analysis. Many variants of these exist, as well as entirely different methods, but here we shall concentrate on the standard methods, which have been carefully thought out and well tested. For additional discussion see the article by D.E. Sayers and B.A. Bunker in reference [1].

Data Reduction

Normalization

In most experiments, the exact concentration of the absorbing atom (e.g. iron) and sample thickness are not precisely known. Also a variety of materials such as mylar windows, and the sample matrix itself, enter into the beam path between the I_0 monitor and the detector. This causes the observed signals to be multiplied by energy dependent absorption factors. In transmission experiments, a logarithm of the ratio of measured currents is taken, so the multiplicative factor turns into an additive background that varies slowly with energy. In fluorescence detection, however, no log is taken and the energy dependent factors are carried through the entire analysis. This causes no problem as long as the effect is the same for standard and unknown, and the energy dependence is very smooth in k-space. The elastically and inelastically scattered x-ray photons from the incident beam contribute an additive background to fluorescence data. Thus, at present, the EXAFS spectra are usually determined only modulo a scale factor and a slowly varying background. In some cases it is desirable to measure the baseline spectrum separately, and then subtract it out. This, of course, will only work if the background is sample-independent, which is not always the case in fluorescence studies on frozen samples. The normalization procedure, by which the data are divided by the size of the edge step, compensates for the uncertainties in the concentration and sample thickness. A common method of determining the edge step is to fit the data within a few hundred ev below the edge, and also above the edge, with low order polynomials (linear or quadratic), and then extrapolate them to the edge.

The difference between the pre-edge and post-edge fits extrapolated to the edge energy is the edge step.

Typical fitting ranges relative to the edge are -200 eV to -20 eV, and $+100$ eV to $+300$ eV, but these depend on the nature of the background. It is generally not appropriate to use an edge feature, such as the highest peak in the spectrum, as a measure of the edge jump, because such XANES features are strongly dependent on the environment of the absorbing atom, and are therefore unreliable indicators of the total amount of the absorbing atom present in the sample. The fitting method described above compensates for trends in the background, and puts little emphasis on the XANES region which is quite variable, and therefore unreliable for normalization purposes. Instead of low order polynomials, any smooth function that fits the data adequately and extrapolates well can be used. One example is the Victoreen function, which empirically fits absorption coefficients well. There is no rigorous justification for use of this function when spurious background (e.g. scattered radiation in fluorescence) is present, however. Usually the pre-edge background is subtracted out at this stage to aid in visual inspection of the data. This is not a crucial or even necessary operation, however, since much more stringent background subtraction is done at a later step.

Conversion to k-space
Sampling

Conversion to k-space should be done before background subtraction so that the background fit does not preferentially follow the data at high energy (which oscillate slowly in energy space). Usually the experimental data are not collected on an even grid in k-space (although this is desirable), and therefore interpolation is performed when the change of variable is made. A uniform grid in k- space is desirable so that standard discrete fourier transform algorithms can be used. The sampling frequency must be high enough that the shortest wavelength in the data is sampled twice in a period, otherwise a rapid oscillation in the data will "alias" (be confused with) an oscillation at half the frequency. Typically a grid in k-space of .05 Å^{-1} is used; this adequately samples the EXAFS from shells out to distances of 30 Å, which are quite negligible. This criterion should be heeded when choosing the sampling in energy space, also.

Choice of E_0

To convert to k-space, the value of E_0, the threshold energy (muffin tin zero level in the theory) must be specified. Fortunately the precise value is immaterial as long as it is within a rydberg or so of the edge, and that it is consistent for standard and unknown. Typically the half-maximum point on the edge is taken, or the bottom of the edge, or the top. Physically somewhere near the bottom of the edge, close to the fermi level, is probably preferred. Only relative shifts in E_0 between standard and unknown are very important: if both are shifted the same amount the answer will be the same to a good approximation. Relative E_0 shifts primarily affect the data at low values of k which are distinguishable from changes in other structural parameters. Thus, ambiguities in absolute E_0 position, and small (≈ 3 eV) differences in relative E_0 position, do not introduce corresponding ambiguities in structure determination by EXAFS. The choice of E_0 does pose significant uncertainties for k-space analysis in the XANES region, however.

Background Subtraction

Atomic Contribution

In the next step, background subtraction is performed, to isolate the oscillatory EX-AFS from the atomic part of the absorption $\mu_0(E)$, and the background absorption or scatter from other elements in the beam path. It is important to note that the quantity defined in equation III is not really what is analyzed. In principle the isolated atom absorption $\mu_0(E)$ (which is generally unknown) is subtracted from the absorption of the atom in condensed matter $\mu(E)$, and the result is divided by $\mu_0(E)$, which is a slow function of energy in the EXAFS region. We can approximate $\mu_0(E)$ by a slowly varying function for the purpose of background subtraction, but the division by $\mu_0(E)$ should not be performed, because experimentally the background contains spurious contributions unrelated to $\mu_0(E)$. The various background curves of different materials (or the same material measured under different conditions) cannot be used to estimate $\mu_0(E)$ for normalization purposes, because $\mu_0(E)$ should be identical (or at least nearly so). On the other hand a straightforward background subtraction is satisfactory, because any errors in background subtraction vary slowly enough in k-space that they don't survive fourier filtering. If the data were divided by $\mu_0(E)$, however, spurious variations in the EXAFS amplitudes would ensue. For this reason one normalizes to the constant edge step; the residual energy dependence in the amplitudes carries through the analysis, but since it is consistent for standard and unknown and is presumed to be a monotonic function of k, it makes no difference in the results. One must be consistent when comparing experiment and theory however. It should be mentioned, however, that theoretical calculations indicate that $\mu_0(E)$ is not strictly monotonic in the EXAFS region, indeed, that the atomic absorption oscillates at a period slow compared to the EXAFS. This would give rise to some amplitude modulation of the EXAFS, which in turn would case sidelobes in the fourier transforms. This may be one reason for low-r sidelobes in many transform spectra; other well-understood causes are described below. These considerations suggest that improvements are possible in the normalization procedure, but this is a topic for future development.

Precautions

Background subtraction is accomplished by ordinary linear least squares fitting, typically using cubic spline functions. The differences between the data and the fit are weighted by an increasing function of k (such as k^3) so that the data at high k (which are of small amplitude) are adequately fit. In background subtraction it is important to remove the background but not alter the data. The first shell EXAFS amplitude will be reduced and distorted if the background subtraction is too severe. If too little background is removed, spurious contributions appear in the fourier transforms, which may (or may not) overlap and interfere with the first shell signal. A variety of smooth functions can be used for background subtraction, as long as the fit is constrained to not oscillate at the same frequency (or higher) as the first shell data. This can (and should) be checked. Plotting the first or second derivative of the background fit on top of the data will indicate if the fit oscillates at the same frequency as the data.

Cubic Splines

The most common functional form used in background subtraction is the least-squares cubic spline. Cubic spline functions are the numerical representation of a "spline" of elastic

material used for drawing smooth curves. The k-space data range is divided into several regions, and separate cubic polynomials are fit to the data over each region. The cubics are constrained so the values, and the first and second derivatives match at the junctures (called "knots"). In some data analysis programs the order of the polynomials (i.e.quadratic, cubic, quartic) and the number of derivatives to match are variable. In other programs one has control over the abscissae of the knots. In most cases adequate background subtraction can be obtained by simply specifying a sufficient number of regions using cubic splines. One normally uses as few regions as necessary to give adequate background subtraction as judged by the fourier transforms. Since an unconstrained cubic can mimic one cycle of a sine wave, one should avoid using as many regions as there are wavelengths of the data. Even this is not a cardinal sin however, because the matching conditions strongly constrain the spline curve, tending to keep it from following the data too closely.

Other Methods

A number of alternative functions can be used for background subtraction. One extremely (dangerously!) flexible function is the "smoothing spline", in which each data point is also a knot. The least squares criterion is used to fit to the data, as described above. However, the smoothing spline possesses a huge number of degrees of freedom, which are restricted by requiring that the mean square curvature of the smoothing spline be equal to a value specified by the user. The data analyst decides exactly how closely the background can fit the data. A smoothing parameter of zero causes the spline to interpolate the data, and a large smoothing factor gives a straight line background. The utmost caution should be used when using smoothing splines to assure that none of the first shell data are inadvertently subtracted out, particularly for noisy data. An automatic background subtraction procedure using iterative fourier transformation has been devised[3] using smoothing splines.

Another standard method of background subtraction is numerical smoothing of the data. The data are convoluted with a smearing function, which rubs out the EXAFS. This function is then subtracted out, and one is left with the EXAFS. This convolution operation should be done in k-space, not E-space, as in the other methods. The smoothing is efficiently accomplished using fast fourier methods, specifically fourier transformation, high-pass filtering, and inverse transformation. Some sort of crude initial background subtraction (such as subtracting a linear trend) is beneficial before this step because of truncation effects, however.

Fourier Transforms and Filtering

At this point, the data consist of sums of damped sine waves corresponding to the different shells of atoms. The EXAFS equation III (extended as necessary) is used as the basis for interpretation. To reduce the number of fitting parameters, it is useful to separate the signals from different shells using fourier filtering methods. These also permit a decomposition of the oscillations into amplitude and phase functions. This step is more difficult than it looks, and great care must be exercised when interpreting fourier transforms. The complications arise primarily from the finite data range in k-space , and non-negligible structure in the backscattering amplitudes and phases. For simplicity the following discussion will be based on fourier integrals rather than discrete fourier transforms; the minor differences for our purposes will be treated as a subtopic below.

Data Range

Typically biological EXAFS data have signal to noise ratios greater than unity up to $k \approx 12\text{Å}^{-1}$, and the region below $k \approx 3\text{Å}^{-1}$ is excluded because of possible complications from multiple scattering etc. Multiple scattering is not a concern for first shell analysis, however, so sometimes lower values of k are included in the transform. Including high-k regions of the spectrum where the signal is swamped by noise mainly adds spurious structure to the transforms and makes them difficult to interpret. Overplotting several independent scans is very helpful (really quite essential) in this regard. Extending the data range to regions in which the signal is negligible doesn't help to resolve transform peaks.

Interpretation of Fourier Transforms

The data are generally weighted with some power of k (typically k^1, k^2, or k^3) to compensate for the decay of the waves, i.e. to make the data more nearly like monochromatic sine waves over the data range If the data are weighted to have a constant amplitude over the k-range, the data outside this range drop abruptly to zero as far the fourier transforms are concerned. Truncating the data set at a zero of the oscillations is of no benefit (contrary to popular belief). However, to reduce spurious "background" peaks at short distances in the transforms, there is some advantage to including an integral number of oscillations within the k-space window, so that the net area is zero. It should be born in mind that if the transform range (or weighting) is different from the background subtraction range (or weighting), low-r transform peaks will generally be present. These are of no consequence as long as their tails do not overlap with the first shell. A complex fourier transform of the data is then performed:

$$\tilde{\chi}_\alpha(r) = \int e^{-ikr} W_\alpha(k)\chi(k)\, dk, \qquad [2]$$

where $W_\alpha(k)$ includes the k^N weighting, and the subscript α indicates a dependence on parameters (such as the width) characterizing the k-space window. The modulus of $\tilde{\chi}(r)$ exhibits peaks that correspond to the various coordination shells. The position of the peak corresponds to the average frequency of the corresponding shell's EXAFS, which is related to the average distance in the shell. The peak height is related to the average amplitude of the (weighted) EXAFS over the data range, and is therefore related to the number of atoms in the shell, the disorder parameter σ^2, the atomic number of the atoms

in the shell, and the k-space window chosen. Thus the modulus of the fourier transform is analogous to a radial distribution function, but it absolutely should be not be called or naively interpreted as "the radial distribution function". The transform depends on many factors including the k-range, and in responsible analysis several transforms with different k weightings and k-ranges are examined until an understanding of the data is acheived. The transform peak position is only a crude measure of the average distance in most cases (it really just measures the average phase slope) and cannot generally be used for accurate distance determination. Moderate to large disorder ($k_{\max}\sigma \approx 1$) in distances can cause significant peak shifts that do not correspond to the average distance. This effect can be dealt with appropriately in k-space using cumulant methods[1] or non-linear least squares fitting (described below). Similarly the peak heights cannot be used to determine coordination numbers except in a very approximate manner, because atomic number, disorder, and other effects are important. The fourier transform is simply an intermediate step in the fourier filtering process; detailed analysis is done in k-space or by suitably careful methods in r-space.

Truncation Effects

Part of the difficulty in interpreting fourier transforms is the truncation ripple, which is a consequence of the sharp truncation of the data. As mentioned above, the data extend over a finite range, and then drop abruptly to zero. This sudden change in amplitude of the data causes "ringing" in r- space, that is, each transform peak acquires a sequence of side lobes which interferes with peaks from other shells and hinders isolation of separate shells. On the other hand, if the data tapered gently to zero at both ends in k-space rather than dropping precipitously, the main part of each peak would be slightly broader, but the side lobes would be suppressed. For example, the fourier transform of a sine wave over an infinite data range is an infinitely narrow and high peak (Dirac's delta function). If this wave were then fourier transformed over a finite data range, the modulus of the transform would be of the form $|\sin(x)/x|$; the delta function would be broadened with a width inversely proportional to the width in k-space, and it also acquires long range oscillatory tails that decay in amplitude only as $1/x$. The full width of the peak to the first minimum is $2\pi/\Delta k$ where Δk is the length of the k range. The height of the peak is proportional to Δk, so the area (height times width) of the peak is independent of the k range. If the sine wave were then multiplied by a bell shaped curve (a gaussian function, say) centered in the middle of the k-range and which dropped to zero at the edges of the data range, the fourier transform peak would be well localized with no ringing; in fact the r-space peak would also be of gaussian shape, with a width inversely proportional to the width in k-space. This inverse relation between r-space and k-space widths is quite general.

k-space Window Taper

Because of these effects, it is frequently useful to taper ("apodize", i.e. put feet on) the window edges in k-space, or multiply by some other function, so that the weighted data rise smoothly from zero on the low k side and fall smoothly to zero on the high k side. Typical window functions consist of uniform weighting over most of the range, with a cosine- squared ("Hanning function") or gaussian taper at the window edges. Frequently it is possible to choose the k weighting so the data naturally fall to small values at the

data boundary, and no additional weighting factor is needed. Normally the standard and unknown are treated in the same manner (that is, the k-space range and taper should be identical), and the weighting factor is carried through subsequent analysis. This is important for cancellation of certain distortions in the filtered data that arise in the fourier filtering process. Sometimes, however, the unknown's EXAFS decays much more strongly than the standard because of disorder effects. In such a case an artificial debye waller factor can (and should) be applied to the standard to make it more similar to the unknown. This artificial weighting simply improves cancellation of errors (see below) in fourier filtering, and is divided out afterwards.

The use of a very broad taper (by applying gaussian window function, for example) is helpful for interpreting transforms, even if it is not used in the filtering process. Such a smooth window almost totally eliminates truncation ripple (at the expense of peak broadening and reduced resolution), and it removes ambiguities about what is "real" and what is artifact. Figures in published manuscripts would be much clearer if such transforms were shown in addition to the ones used in filtering. Initial guesses for windows can be obtained on the basis of such transforms.

Real and Imaginary parts

The transformed data actually consist of a complex function, which has real and imaginary parts, or alternatively a modulus and phase. The modulus is the most frequently used quantity, but the real and imaginary parts are also useful. They exhibit significantly more structure than the modulus does, and they don't suffer from nonlinear interference. The fourier transform is a linear operation, that is, if the k-space data consists of a sum of sine waves, the fourier transform (*not* the modulus, mind you) is also just the sum of the fourier transforms of the different contributions. Calculating the modulus, however, is a nonlinear operation (the modulus of the sum is not the sum of the moduli); this is why adjacent peaks can interfere with (rather than just superimpose on) each other in "transform" (transform modulus) plots. One aspect of intershell interference causes considerable confusion. Often peaks which are not resolved from each other interfere strongly enough that there is a sharp dip in the modulus between the peaks, which may approach zero. Inexperienced data analysts often assume that the peaks are well isolated from each other because the modulus is close to zero there. In fact the opposite is true. A tell-tale sign of such interference is a marked asymmetry in the peak shape; the peak drops off much more rapidly on the "interfering" side than on the other one. If such interfering peaks are separately inverse transformed, the results will be unpredictable, but almost certainly will be seriously in error.

r-window and Inverse Transform

Once a k-range and window function are chosen that give suitably isolated peaks in r-space, the peaks outside the r- space range of interest are set to zero and the data set is inverse-transformed to k-space:

$$\chi_{\alpha\beta}(k) = \int e^{+ikr}\tilde{\chi}_\alpha(r)W_\beta(r)\,dr \qquad [3].$$

$W_\beta(r)$ is the r-space window, which depends on parameters denoted by the subscript α. It is best to include the entire peak in r-space; a narrow window smooths out structure in the

filtered k-space data, which looks nice but reduces the information content. The narrower the window, the more perfectly sine-like the filtered data becomes. If the entire peak (and no others) can be included within the r-space window, fourier filtering distortions are minimal. However, if the shoulders and side-lobes (including truncation ripple) must be excluded because of the proximity of other peaks, some distortions of the data will ensue. To assure that these are the same for standard and unknown, one must choose the same width of the r-space window (the position can differ). It may be desirable to slightly taper the edges of the r-space window, but not to the extent done in k-space , which would unnecessarily distort the data. It is important that a flat topped window be used in r-space; this ensures that the frequency of the filtered data is not sensitive to the precise choice of r-window. This implies, for example, that a gaussian function is not a suitable r-space window, although it is suitable for k-space. The difference is a consequence of the opposite character (sine wave vs sharp peak) of the data in k and r spaces.

Amplitude and Phase

As far as the complex fourier transform is concerned, the sine waves in k-space (which are real quantities) consist of positive frequency parts and negative frequency parts. These correspond to the terms in the the Euler identity $2i \sin(x) = e^{ix} - e^{-ix}$. These negative frequency parts contain redundant information, and are not generally plotted or even discussed in polite company. If the negative frequency components are omitted in the inverse transform, the filtered data set $\hat{\chi}_{\alpha\beta}(k)$ is complex, and single shell amplitude $A(k)$ and phase $\varphi(k)$ functions are readily computed from the real and imaginary parts. The oscillatory form of the filtered data is given by $A \sin(\varphi)$. These functions are unique, in the ideal case in which the various shells, and the positive and negative frequency components of the r-space transform don't interfere with each other. For single shell data, such interference between positive and negative frequency components will be minimal when the separation between the peaks (twice the apparent distance, which is essentially the frequency of oscillation in k-space) is much greater than the intrinsic width of the peak (which is inversely proportional to the range in k-space). Thus, when many oscillations are contained in the k-space filter window, the phase and amplitude are well defined. Often the phase and amplitude determined by fourier filtering depend on how the data are treated. For this reason it is important to use standards that are treated the same way as the unknown, so the systematic effects cancel out.

Cancellation of Window Distortions

The filtering process can be regarded as a convolution in k- space of the weighted, windowed EXAFS data $\chi(k)W_{\alpha}(k)$ with the fourier transform $\hat{\chi}_{\alpha\beta}(k)$ of the r-space window $\tilde{W}_{\beta}(k)$. The width of this smearing function in k-space is inversely proportional to the width of the r-space window, as mentioned above. Therefore the data within this width of the k-space cutoff will be most strongly affected by the finite data range; such distortions, "window effects", can be severe if the data are sharply truncated over a short k-range, and the r-space window is narrow, so the quantity $\Delta k \Delta R/\pi$ becomes small (approaches one or less). Fortunately the window effects cancel out to a good approximation if care is taken. Specifically the k-space window positions and tapers, and the r-space window widths should be the same for standard and unknown. If the amplitude decay is very different for the spectra being compared, the difference should be compensated for before

filtering, and restored afterward. Different compensations may be necessary for different shells, and sometimes a compromise is necessary to obtain adequate isolation. Constant scale factors and constant phase shifts don't affect the window distortions as long as there is no interference from the peaks at negative frequency (negative distance). Furthermore the window effects don't depend on the transform peak position (i.e. the distance), provided the filter window is appropriately positioned on the peak. Thus, as long as the windows are chosen as described above, if the standard and unknown differ only by a constant factor, a constant phase shift, or a shift in distance, the window distortions will exactly cancel out in later stages of analysis. In practice the cancellation is not perfect because of some spectral leakage between shells, but it is usually quite good.

Parameter Estimation

Once the phase and amplitude functions for the appropriate shells in the standards and unknown are determined, structural information can be obtained by various means. If the shell in the unknown consists of only one type of atom, the ratio method is appropriate. If this is not the case, nonlinear least squares fitting is performed using amplitude and phase functions obtained from empirical standards or theoretical work.

Ratio Method

If there is only one type of atom in the shell, analysis is very simple. According to the EXAFS equation, assuming the mean free path term approximately cancels, and transferability holds, the logarithm of the ratio of the amplitude of the unknown to that of the standard is a linear function of k^2. In the simplest form of the ratio method, this quantity is plotted and fit with a line. The intercept gives the ratio of coordination numbers, and the slope gives the difference in mean square disorder. Similarly, the phase difference, when plotted versus k, has zero intercept, and slope equal to twice the difference in average distance for the shell.

In reality, ratio plots are only approximately linear because of noise and other errors. One must exclude the very ends of the data range from the fitting region, because window distortions there do not completely cancel for standard and unknown. It is crucial in the ratio method to choose windows consistently as described above. To reliably determine whether curvature is significant or not, a noise analysis is imperative. It is essential to compare ratios between independent scans and to include error bars for the ratios when fitting is performed.

Also, in practice, the phase intercept is not usually exactly zero because of differences in E_0 between the standards and the unknown. Changing E_0 adds to the phase a term inversely proportional to k. Empirically it is found that reliable distance determinations are obtained when the E_0 of unknown (or standard) is adjusted relative to the standard (or unknown) to give zero intercept as equation III implies.

The method above is valid for small disorder $k\sigma << 1$. If this is not the case, rather than obtaining straight lines, some curvature may be observed that is greater than can be explained by noise in the data. This curvature can be analyzed to determine the first few cumulants[2] of the distribution of atoms in the shell. Because the cumulant expansion is a linear function of the cumulants, the fit is linear, and so there is only one global minimum, a great simplification compared to nonlinear least squares fitting. Suppose, for example, that

9

the standard is radially well-ordered, so its $k_{max}^2\sigma^2$ is small. Then a negative (positive) curvature in the phase difference indicates the distribution of atoms in the unknown is skewed to longer (shorter) distances. Positive (negative) curvature in the log amplitude ratio plot indicates the distribution has more (less) weight in its tails than does a gaussian distribution of the same mean squared width. The cumulant expansion[2] provides a general model-independent framework analyzing the EXAFS of moderately disordered systems.

nonlinear least squares fitting

Once the EXAFS phases and amplitudes have been isolated for the standards, so that, for example, the $B(k)$ and $\delta(k)$ functions for oxygen, nitrogen, and sulfur are known, then, using the EXAFS equation, it is a simple matter to synthesize the EXAFS corresponding to any particular hypothetical structure. This spectrum can be compared to the spectrum of the unknown, using a least-squares or other criterion. Comparison can be made to the amplitude A and phase φ separately, a weighted combination, or to $A\sin(\varphi)$. Fitting to a weighted combination of A and ϕ is flexible; fitting to $A\sin(\varphi)$ tends to give more weight to the phase. The sum of squares function, which measures the goodness of fit, is a function of the hypothetical structural parameters. By minimizing the mean square error (according to any of a number of standard numerical algorithms) the parameter values for the best fit to the data are obtained.

The canonical method of estimating parameters and their error bars (confidence intervals) in nonlinear least squares theory requires one to minimize the mean square error $\epsilon^2(\{\alpha\})$ between the data and the fit:

$$\epsilon^2(\{\alpha\}) = \frac{1}{N_{pts} - N_{par}} \sum_i \frac{(\mathrm{Data}_i - \mathrm{Fit}_i(\{\alpha\}))^2}{s_i^2}, \qquad [3],$$

where $\{\alpha\}$ denotes the set of parameters varied in the fit, and σ_i is the standard deviation of each data point, from noise and whatnot. N_{pts} and N_{par} are respectively the number of independent data points and the number of parameters floated in the fit. Note that ϵ^2 is dimensionless; the noise level represented by s_i^2 in the data has been divided out. Best fit values that are much greater than 1.0 are considered poor fits, and any fit less than about 1.0 is considered acceptable. This function ϵ^2 is called χ^2 in the case of data with normally distributed random errors, but this is generally not the case for EXAFS. The error bars, if sufficiently small,can be estimated by suitable analysis of the curvature (Hessian) matrix $\partial^2\epsilon^2/\partial\alpha_i\partial\alpha_j$ as discussed below.

In this context, two essential points must be kept in mind. The first is that the denominator $N_{pts} - N_{par}$ must be included to properly account for the number of degrees of freedom used in the fit. The second is that number of independent data points is reduced by fourier filtering. In essence, when you do the forward transform, the information content in k-space is spread out over a wide range of r, and when you use an r-space window and inverse transform, only a fraction of the information is retained. The number of independent data points after fourier filtering using windows of width Δk and Δr is $N_{pts} = \Delta k \Delta r/2\pi$. The number of parameters floated in the fit must not exceed this value. Results obtained from overfitting are generally meaningless.

The EXAFS equation is a linear function of the coordination numbers, but it is a nonlinear function of the distance and disorder parameters. Thus the fitting problem is a

nonlinear one, and many local minima in the sum of squares function may exist. For this reason it is important to try many different starting points in the minimization procedure. Usually local minima give poor fits (outside the fluctuations from noise), but not always. Frequently these solutions can be rejected on physical grounds. Otherwise they must be accepted as possible valid structures as far as EXAFS is concerned. The aim of fitting is to identify and describe all of the regions in the parameter space that match the experimental data within the error bars. We are not just interested in the best fit value; a finite domain about that point also gives fits that are consistent with the data, because of uncertainties due to noise.

If the noise level (represented by s_i^2) is known, standard methods can be used to determine the error bars for each parameter. For small errors, the sum of squares is a quadratic function of the parameter deviation about the minimum. The curvature ("hessian") matrix can be inverted and scaled by the noise level to produce an approximate covariance matrix. The square roots of the diagonal elements of the covariance matrix give the parameter error bars. The off-diagonal elements, when divided by the square root of the product of the corresponding diagonal elements, give the correlation matrix, which indicates to what extent variations in one parameter can compensate for variations in another parameter. Such parameter correlation increases the estimated parameter error bars.

If the noise level is not small enough to permit the "parabolic" expansion near the minimum, error bars are obtained by varying one parameter, while optimizing (floating) the others, until the sum of squares function ϵ^2 is increased by unity (not doubled as is often stated). This procedure accounts for correlations (it would not if the other parameters were held fixed at their best-fit values) and is more accurate than the previous method.

The noise level can be estimated in several ways. The simplest method is to analyze separate scans in parallel and compare the results. For biological data, if there are many scans, they can be divided into two equal sized groups, and partial sums, and the total sum of all scans, can be formed and analyzed in parallel along with the standards. This provides a simple and effective means of estimating noise at all stages of analysis, including fourier transforms and nonlinear fitting. There is really no excuse not to do some sort of noise estimate in data analysis.

choosing standards

Clearly standards are of central importance in EXAFS analysis. Basic criteria for a "good" standard in the EXAFS sense are: same atomic number of atom; small disorder ($k\sigma \ll 1$), similar bond lengths (within .1 Å of the unknown, say); similar oxidation states in standard and unknown. It is not necessary for the entire environment of the absorbing atom to be similar in standard and unknown. In mixed ligand (e.g. N,O,S) coordination this would make it difficult to extract phase and amplitude functions from the standard. The single scattering approximation implies that the EXAFS signals from different neighboring atoms are independent. This "superposition" is embodied in the EXAFS equation as a simple linear sum over contributions from different atoms. Thus a good "model compound" for EXAFS purposes may in general be quite different from what is considered a good model compound in chemically sensitive techniques like optical spectroscopy. Thus I will use the term "standard" rather than "model" for this purpose, and reserve the term "model" for hypothetical structures considered in later fitting procedures. It is very often possible, however, to obtain model compounds that are chemically similar to the unknown and good EXAFS standards as well.

The similarity of the edge structure (XANES) is also a useful diagnostic. If the XANES for standard and unknown are similar the standard is probably a good one, but if the XANES is dissimilar the standard is not necessarily bad. The explanation for this is that the XANES can be sensitive to atoms beyond the first shell, which have little bearing on the first shell transferability.

The quality of standards is checked by using several redundant ones and comparing the results. This is a prudent procedure in routine work and should be followed whenever possible. Most workers have evaluated the consistency of results when using different standards and have a good idea of the variation obtained when their own methods of data acquisition and analysis are used. It is probable that much of the variation that has been observed (non-transferability of EXAFS amplitudes) is as much a consequence of the experimental problems (e.g. thickness effects and particle size effects; the latter can be important in fluorescence) and data analysis problems (lack of cancellation of window effects in fourier filtering) as it is a consequence of a "real" limitation of transferability.

theoretical "standards"

Because of the complexities of using empirical standards, some workers find it convenient to use theoretical values that have been tabulated by Teo and Lee[4], or the improved ones of McKale[5] *et al.* One should not have the mistaken impression that these are somehow more "accurate" than experimental amplitudes and phases. The assumption that is implicitly made when using theoretical values is "transferability" between rather crude theoretical models and rather complex experimental phenomena. The isolated atom model used in these calculations cannot compensate for chemical variation between systems. Furthermore, inelastic effects are largely neglected in the theoretical calculation. Such multielectron excitations cause a 30% to 50% reduction in EXAFS amplitude for experiment as compared with theory, which is accounted for by an empirical scale factor. In Teo's and Lee's calculation, the small atom approximation is also made, which requires large artificial changes in E_0 to make experiment agree with theory. McKale's functions solve this problem, however.

Through the use of model compounds, nature does a much better simulation than we can do on a computer. The use of empirical standards has many benefits. It compensates for various instrumental effects such as energy dependent absorption by the sample matrix, energy resolution (if broad), and the energy dependence of I_0 monitor sensitivity; for the breakdown of the small atom approximation and inelastic effects in the EXAFS equation; and distortions and variations that occur in data reduction and analysis.

This does not mean that theoretical calculations are useless (quite the contrary), only that they should be used with discretion. The calculations of Teo and Lee, and the improved spherical wavefunctions of McKale and Knapp are extremely useful for understanding the dependence of backscattering functions on atomic number and energy. Furthermore, it sometimes happens that an appropriate standard cannot be found for a particular system. In this case, theoretical calculations must be used in some way. A good way to do this is to use differences between theoretical functions for different elements to generate corrections to experimental models. For example, if one needed an Fe-S standard, but only had access to an Fe-Cl standard, the Fe-Cl empirical amplitude and phase could be corrected by the theoretical difference between S and Cl. Thus the theory permits the experimenter become an alchemist. To use theoretical functions, it is always necessary to apply some empirical correction factors, but if this is done properly, reliable data analysis is possible, and indeed is routinely accomplished by many groups. It is very important that the theoretical data be treated in the same way as the experimental data however, so that systematic errors that arise in data analysis cancel out. Also the energy dependence of the central atom absorption must be treated in a consistent manner for experiment and theory.

The future looks bright for theoretical calculations of EXAFS and XANES, however. Even now it takes only a few minutes to generate an *ab initio* theoretical EXAFS spectrum on a microcomputer. Because of the difficulty of properly accounting for multiple scattering (e.g. the focussing effect) by traditional methods using empirical standards, the growing sophistication of theoretical calculations, and the diminishing cost of computing power, in coming years it seems likely that EXAFS and XANES data analysis will become more dependent on *ab initio* or semi-empirical modeling and simulation. What will not change is the fact that the methods of data analysis must be firmly grounded in both experiment and theory.

References

[1] *X-ray Absorption* D.C.Koningsberger and R. Prins, eds; Wiley, 1988

[2] G.B. Bunker, Nucl. Inst. Meth. **207**, 437 (1983)

[3] J.W. Cook and D.E. Sayers, J. Appl. Phys. **52**, 5024 (1981)

[4] B.K. Teo and P.A. Lee, J. Am. Chem. Soc. **101**, 2815 (1979)

[5] A.J. McKale and G.S. Knapp, J. Am. Chem. Soc. **110** 3763 (1988)