Journal of
Synchrotron Radiation

ISSN 0909-0495

Received 22 June 2011
Accepted 30 December 2011

Stereochemical analysis of ferrocene and the uncertainty of fluorescence XAFS data

Christopher T. Chantler,[a]* Nicholas A. Rae,[a] M. Tauhidul Islam,[a,b] Stephen P. Best,[b] Joey Yeo,[b] Lucas F. Smale,[a] James Hester,[c] Narges Mohammadi[d] and Feng Wang[d]

[a]School of Physics, University of Melbourne, Australia, [b]School of Chemistry, University of Melbourne, Australia, [c]Bragg Institute, ANSTO, Australia, and [d]Faculty of Life and Social Sciences, Swinburne University of Technology, Hawthorn, Australia. E-mail: chantler@unimelb.edu.au

Methods for the quantification of statistically valid measures of the uncertainties associated with X-ray absorption fine structure (XAFS) data obtained from dilute solutions using fluorescence measurements are developed. Experimental data obtained from 10 m*M* solutions of the organometallic compound ferrocene, $Fe(C_5H_5)_2$, are analysed within this framework and, following correction for various electronic and geometrical factors, give robust estimates of the standard errors of the individual measurements. The reliability of the refinement statistics of standard current XAFS structure approaches that do not include propagation of experimental uncertainties to assess subtle structural distortions is assessed in terms of refinements obtained for the staggered and eclipsed conformations of the C_5H_5 rings of ferrocene. Standard approaches (*XFIT*, *IFEFFIT*) give refinement statistics that appear to show strong, but opposite, preferences for the different conformations. Incorporation of experimental uncertainties into an *IFEFFIT*-like analysis yield refinement statistics for the staggered and eclipsed forms of ferrocene which show a far more realistic preference for the eclipsed form which accurately reflects the reliability of the analysis. Moreover, the more strongly founded estimates of the refined parameter uncertainties allow more direct comparison with those obtained by other techniques. These XAFS-based estimates of the bond distances have accuracies comparable with those obtained using single-crystal diffraction techniques and are superior in terms of their use in comparisons of experimental and computed structures.

Keywords: fluorescence; X-ray absorption fine structure; ferrocene; standard error.

1. Introduction

The subtle inter-relationship between the electronic and molecular structure of transition metals and the impact on reactivity underpin their proliferation as the catalysts for biological and industrial processes. For homogeneous catalysis in solution it is vitally important to have available methods that determine, reliably, the structural details needed to understand details of the reaction path. This, to some, is the *raison d'etre* of spectroscopy. Notwithstanding the possibilities presented by optical, EPR (electron paramagnetic resonance) and NMR (nuclear magnetic resonance) spectroscopy for particular systems, X-ray absorption fine structure (XAFS) provides the best general approach for structural characterization at the molecular level for metal species in low concentration. While there is an extensive literature demonstrating the application of XAFS to the identification of the coordinating atoms and extraction of geometrical information from analysis employing multiple scattering formalisms, any assessment of the derived geometry is limited by difficulties in assessing the reliability of the individual experimental observations.

The difficulty in assessing the statistical reliability of the XAFS data is greatly accentuated in cases where fluorescence detection is employed, *i.e.* for the study of samples in dilute solutions. This paper is concerned with quantification of the accuracy and precision of fluorescence XAFS measurements. Ferrocene, $Fe(C_5H_5)_2$ (di-cyclopentadienyl iron), discovered in 1951 (Kealy & Pauson, 1951; Miller *et al.*, 1952), is a prototype of sandwich molecular complexes which epitomizes the different bonding interactions possible between metals and organic molecules and is central to the award of the Nobel prize to Fischer and Wilkinson in 1973 (Fischer & Wilkinson, 1993). The molecule has well defined redox chemistry and is a popular reference material for electrochemical studies conducted in non-aqueous solvents and is an ideal test case for *in situ* electrochemical cells for XAFS and X-ray absorption near-edge structure (XANES) studies.

An interesting aspect of the structural chemistry of ferrocene is the orientation of the two cyclopentadienyl rings. The

initial interpretation of the crystallography suggested a 'staggered' conformation giving a molecule with D_{5d} symmetry, but this was subsequently revised with the 'eclipsed' (D_{5h}) form believed to be more stable by perhaps 4 kJ mol^{-1} (Haaland & Nilsson, 1968; Haaland, 1979). When compared with axial distortions of Jahn–Teller distorted octahedral complexes or square-planar tetrahedral isomerism of four-coordinate complexes, differences in the relative orientations of the cyclopentadiene rings of ferrocene present a serious test of the ability of XAFS methods to reveal subtle differences in the molecular structure. Moreover, since the main differences in structure will be manifested by multiple-scattering interactions, the problem also allows assessment of experimental and theoretical approaches to analysis at the interface of the XANES and XAFS regions.

For absorption spectroscopy, the X-ray extended range technique (XERT) can provide defined accuracies and uncertainties in XAFS experiments (*e.g.* Tran *et al.*, 2003*c*), and investigate experimental systematics including harmonic contamination (de Jonge *et al.*, 2004*b*; Tran *et al.*, 2004). This allows absolute intensity measurements of X-ray attenuation, including measurements in the XAFS region (*e.g.* de Jonge *et al.*, 2005, 2007; Glover *et al.*, 2008; Islam *et al.*, 2010; Rae *et al.*, 2010*a*), and can measure X-ray absorption with an accuracy below 0.05%. It should be possible to obtain reliable data on chemical bonding, thermal effects and anharmonicity, as well as to establish the absolute fluorescent radiation yield. However, the analysis even of accurate absorption data is hampered by difficulties in the propagation of experimental uncertainty, which is exacerbated in the case of fluorescence spectra from dilute samples. Past analyses of XAFS data from ferrocene have found no difference between possible models (Vedrinskii *et al.*, 1991; Ruiz-Lopez *et al.*, 1988; Schnitzler *et al.*, 2006).

Standard methods used by the XAFS community to define coordination environments of an absorbing atom can, for example, distinguish coordination number, changes in absorber–scatter distance or square-planar–tetrahedral isomerism. However, the absence of a statistically rigorous method of assessing the reliability of the individual measurements, particularly in the case of fluorescence experiments, not only has the potential to give misleading results but also limits the possibility of assessing whether the experimental data are of a quality to support even the asking of such questions. We focus on the processing of X-ray absorption spectra (XAS) obtained from frozen solutions of ferrocene data collected at the Australian National Beamline Facility (ANBF) in Tsukuba, KEK, Japan, using a conventional fluorescence geometry. In particular we set out the corrections needed to bring the independent observations from a multi-element detector onto the same absolute intensity scale. This permits a more detailed statistical analysis of the spectra than has previously been reported.

Most XAFS data are represented by a single reduced plot such as Fig. 1. This shows clearly the sharp edge jump after *K*-shell ionization of the central iron atom, together with the interference wave of the resulting photoelectron reflecting

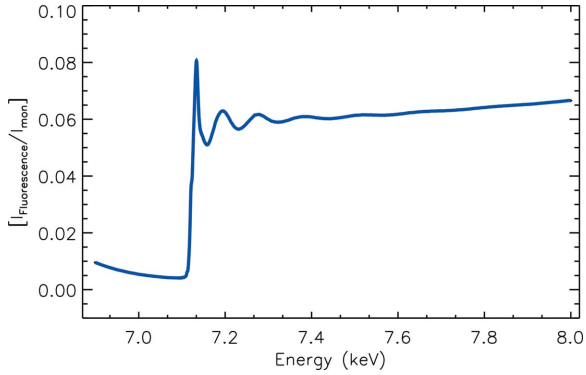

Figure 1
The standard average output using *AVERAGE* for a 10 m*M* solution of ferrocene.

from surrounding electron density of the coordinated ligands. The fluorescence signal arises from the relaxed iron $K\alpha$ $2p$–$1s$ transition lying in energy somewhat below the continuum, and was measured in a 36-channel segmented germanium detector, configured to report the counts in the energy region-of-interest (ROI) corresponding to the iron fluorescence $K\alpha$ region.

Standard analysis in the Australian XAFS research community uses the package *AVERAGE* (Ellis & Freeman, 1995*a*) to reduce the dataset to this plot. Standard analysis may then use *IFEFFIT* (Newville, 2001) or alternatively *XFIT* (Ellis & Freeman, 1995*b*) and *FEFF* (Version 6.1) (Rehr *et al.*, 1991) for example to further reduce this to a χ *versus* k plot, scaled or otherwise, which is then fitted with the usual array of XAFS parameters. In the case of the high-symmetry molecule ferrocene the XAFS may be modelled using five independent variables [E_0, S_0^2, $C(x)$, $C(y)$ and $\sigma^2(C)$] if the scattering from hydrogen is neglected and seven otherwise (if the cyclopentadiene ring is planar).

Inspection of Fig. 1 immediately reveals that the data are subject to systematic errors, especially noting the gradual increase of absorption coefficient with energy above the edge and the peculiar signature below the edge with the qualitative form of an extended quadratic. These distortions of data do not prevent XAFS analysis, as baseline subtraction allows extraction of χ, but these and other effects conspire to obstruct determination of statistically robust estimates of the uncertainty of individual measurements. This, we argue, prevents the use of XAFS methods for the examination of more subtle structural questions such as the staggered or eclipsed structures of ferrocene, or perhaps the percentage of each if there is some thermal disorder. The conformation of the cyclopentadienyl rings may be examined by comparison of the refinement statistics using model structures of the two conformers, or models in which the Fe–C interactions are not equivalent. Assessment of the significance of any discrepancy between the refinement statistics relies on the reliability of primary experimental observations, this most sensibly being statistically based. Clearly it is critical to be able to assess both the accuracy and precision of the data in order to establish whether the question posed is able to be answered with any sort of reliability.

In this paper, we (i) develop a methodology for determining robust standard errors from such typical datasets; (ii) discuss the fluorescence theory for data reduction; (iii) develop techniques for error propagation in least-squares fitting and (iv) illustrate their potential applied to a subtle and long-standing problem of ferrocene.

2. Conformers of ferrocene

The high-symmetry conformations of ferrocene are illustrated in Fig. 2. The iron atom, Fe^{2+}, is centrally located between two parallel cyclopentadienyl rings where the rings are either eclipsed D_{5h} when viewed down the symmetry axis or have a staggered D_{5d} conformation (*i.e.* where one ring is rotated relative to the other by an angle of $2\pi/5$ to yield an inversion centre).[1]

Early determinations of ferrocene assumed that the D_{5d} conformation was the stable form, based in part upon symmetry properties, and it was not until well after the determination of the X-ray structure of ruthenocene, $Ru(C_5H_5)_2$, for which the D_{5h} structure appeared to be the refined structure (Hardgrove & Templeton, 1959; Seiler & Dunitz, 1980), that the question of the stability of ferrocene conformation was raised (Seiler & Dunitz, 1979*a*,*b*). However, for the Fe case, the X-ray structure appeared complicated by disorder (dynamic or static), and the situation remains unclear, even in recent reviews (Yamaguchi *et al.*, 2007). In particular, the early assumption of an ideal staggered conformation was questioned and a necessary disorder of either staggered or eclipsed, or partially staggered, molecules in crystalline form has been developed. The crystallographically refined conformation appears to depend upon crystallization temperature and monoclinic, triclinic or orthorhombic phase. The energy and property differences between the D_{5d} and the D_{5h} conformers of ferrocene appear small (Coriani *et al.*, 2006), which makes it possible that the two conformers may be thermally interchangeable, *i.e.* there is a small rotational energy barrier.

In relation to characterization of molecular structure, X-ray and neutron diffraction reveal accurate mean lattice positions of electron or nuclear density, and hence differences between these mean positions may be obtained for well defined crystals. XAFS has a strong potential advantage over these techniques in that it directly samples the dynamic and static intramolecular disorder and hence the molecular and electronic structure of the complex including the bond lengths. Exploitation of this advantage requires high-quality accurate data.

Important questions that must be addressed with any experimental investigation are 'How accurate is this experimental dataset?' and, perhaps with greater concern, 'How precise is this dataset?' If the dataset is highly accurate, then it can be used for any purpose, including testing of theory,

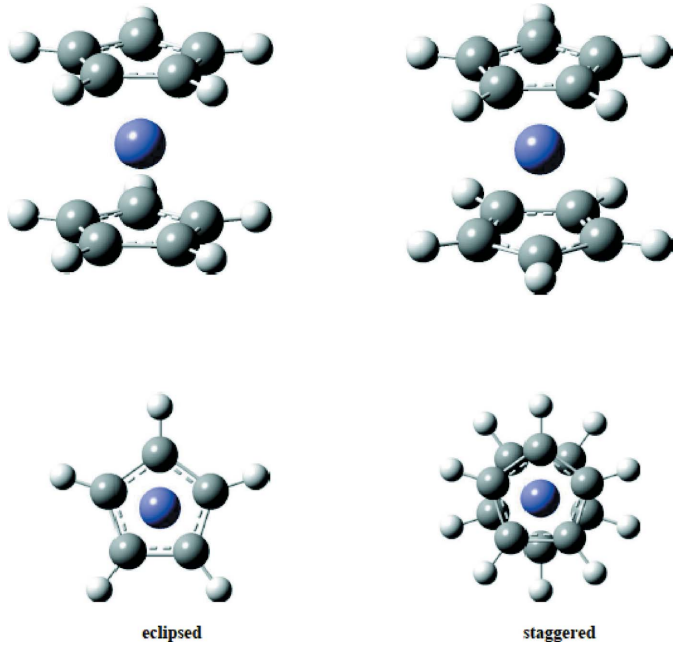

Figure 2
A model of the ferrocene molecule with staggered (D_{5d}) and eclipsed (D_{5h}) symmetries.

investigation of edge-jump ratios, derivation of form factors, and of course investigations of XANES and XAFS. If it is highly precise but of uncertain accuracy, then the first three applications may be invalidated but the dataset may, after normalization, be excellent for XANES or XAFS investigations. A critical question of this investigation is whether accurate XAFS datasets can reliably resolve conformational ambiguity.

3. Brief experimental

In order to avoid microcrystallization at the low temperatures of 10–20 K in the cold cell (cryostage), solutions of 10 m*M* ferrocene were prepared using a mixed solvent of (50%) CH_3CN acetonitrile and (50%) C_3H_7CN butyronitrile. The standard XAFS configuration at ANBF was followed, involving a monitor upstream (ion chamber, 10 cm, 70% He, 30% nitrogen-filled) and the 36-element Ge planar detector (EURISYS EPIX 36-64-7-ER). The detector contains 36 channels with an order of 6 × 6 channels forming a square area of (50 mm × 50 mm) with each pixel capturing an area of 8 mm × 8 mm. The output file runs from channels 0 to 34 (the last pixel is dead), with 0 being the top downstream end and incrementing horizontally. The approximate gap between the active area of each channels is 0.4 mm. So, the separation of pixel centres at the detector surface is 8.4 mm. Three scans of the XAFS region were made to provide more reliable data, so each point in energy has 35 × 3 repetitions (one per active detector channel, one per scan).

Raw signals from the detector channels are fed to digital signal processing (DSP) modules [DXP4C-2X modules from XIA (Warburton *et al.*, 1999)] which calculate the total incoming count rate (ICR) for all energies, the energy binned

[1] An interactive version of Fig. 2 is available as supplementary data for this paper (Reference: WA5021). Services for accessing these data are described at the back of the journal.

count rate (OCR) and the total counts in the energy region of interest.

The angle of the sample in fluorescence (the solution cell) is very close to 45° to the incident beam. The central position of the detector is aligned to be approximately 45° to the solution cell or approximately 90° to the incident beam. In the vertical direction this is well aligned to within 1 mm; in the horizontal direction it is probably accurate to within 5 mm.

4. Intrinsic precision of the dataset

Fig. 3 plots all the scans of the dataset, given by I_{dxp}/I_{mon} where I_{dxp} is the raw count received within the region of interest of a single channel of the detector, while I_{mon} is the raw count of the monitor, the upstream ion chamber. This normalization does not correct for air path, scattering, harmonic or other parameters investigated by XERT; this is a typical conventional dataset. It also does not correct for detector channel offsets or scale errors; both monitor and detector channels typically measure a non-zero count per second as electronic noise. However, this is the standard data which is used for fluorescence studies at ANBF and elsewhere. Channel 35 is dead, and channel 34 (the purple line at the bottom) has virtually zero sensitivity, and is excluded from further data analysis. A second major point is that the different channels have extraordinarily different values for I_{dxp}/I_{mon}. Indeed, we have plotted this to suggest a systematic effect, discussed below.

Fig. 3 looks somewhat confusing, so it may be sensible to compress the dataset from 105 scans (3 scans per channel × 35 channels) into 34 lines, each of which is an average of the corresponding three scans for that channel. Additionally there is a very well known effect of dead-time in the fluorescence detector, caused by the processing time of the channel from the raw data into the region of interest. The dead-time is conveniently determined for the scattering detector because we have one signal (a fast processing signal) prior to measuring the energy (using electronic signal processing and then an analogue-to-digital conversion into energy bins) I_{ICR}, and a similar count of all photons tagged after this energy conversion I_{OCR}. Hence the number of pulses lost owing to dead-time is given by the difference between these two numbers, and the dead-time-corrected spectra are generated by $I_{dxp}/I_{mon} \times (I_{ICR}/I_{OCR})$.

Correction of the results obtained for each detector channel by implementing the dead-time correction and taking the average of the three independent scans gives an improvement in the signal-to-noise ratio, as expected, but with only a small improvement of the variance of the signal for the different detector components (Fig. 4). This set of observations may be used to compute the averaged χ *versus* k spectrum with standard errors giving the explicit absolute (and relative) uncertainty of each point (Fig. 5). However, the relative error thereby obtained is large owing to a systematic variation in the signal intensity which is correlated with the identity of the detector element (Fig. 4). This is a major result, although perhaps not too complex; we have a standard error of our official average result.

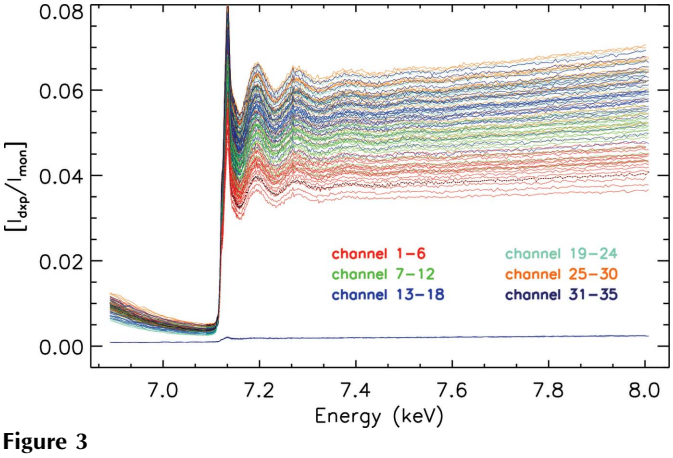

Figure 3
Plot of the full dataset for the raw detector signal within the fluorescence region-of-interest divided by the monitor signal. There are 36 segmented detector channels arranged in a square, and the dataset collected three consecutive scans of the energy range presented. Colour coding is based on vertical columns of detector elements.

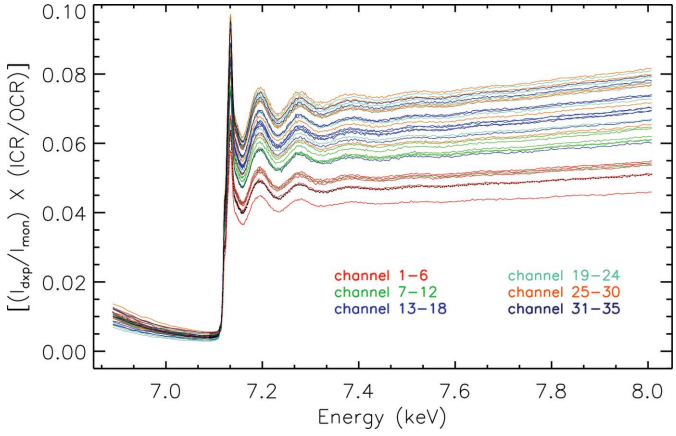

Figure 4
The full usable dataset for the raw detector signal within the fluorescence region-of-interest divided by the monitor signal, normalized by the detector dead-time correction. Each trace represents the unweighted average of the corrected ratios of the three independent scans.

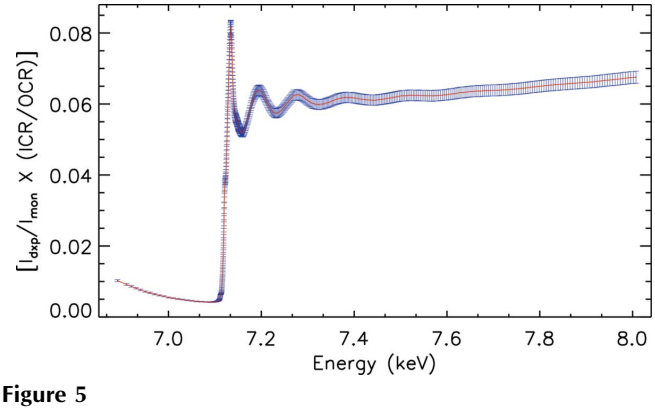

Figure 5
The average signal of the 34 usable detector channels, corrected for detector dead-time, with a standard error representing the precision of determination of the individual data points.

5. Alternative analytic approaches to multi-pixel data

If we investigate further a single scan, with three repeated scans thereof, we can construct a line (just as in Fig. 4) but with a standard error formed from the consistency of the three scans for that channel. If there were electronic drift, a limitation of normalization by the monitor, some scale or amplifier drift with time, or any other functional dependence with time, then we will observe variation between the scans, and the resulting standard error will represent our ability to determine the average for that particular channel. Hence this uncertainty map will provide a measure of our information content for that particular channel. This is provided by way of example, for channel 1, in Fig. 6. The points are relatively noisy, but the standard error is provided by the consistency of the scans, so really is quite noisy. In absolute terms, this standard error is poor below the edge, though with large oscillations, and relatively well defined at higher energies above the edge. The greatest dispersion is in the near-edge region, in part owing to energy drifts and counting statistics. In part, this is a deliberate experimental optimization of the data collection routine; the code counts longer for regions higher in k and hence above the edge, leading to better statistical determination. This variance is not due to dark count correction nor air path correction, but may have causes in alignment, beam motion on sample and ROI settings.

We could therefore process the reduced dead-time-corrected average given by Fig. 5 with its corresponding standard error (a perfectly self-consistent approach); or one could process an individual fluorescence detector channel with its uncertainty as illustrated by Fig. 6. While the latter would be self-consistent, it may be consistent or inconsistent with any other fluorescence detector channel.

If it is supposed that the systematic variation is due to differences in sensitivity of the individual channels owing to pixel thickness, detector pathlength or other scaling or offset parameters, then this may be addressed by normalization of the signal in a self-consistent manner. The amplifier settings have been calibrated with a flat field, and ROIs have been set, so these uncertainties should be small. For the purposes of this argument, Fig. 7 provides the set of all scans and channels

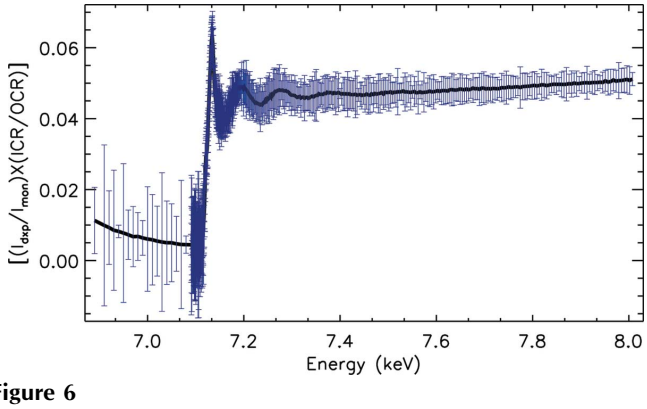

Figure 6
The average and standard error of the dead-time-corrected fluorescence signal normalized to the monitor reading, from three independent scans (detector channel 1).

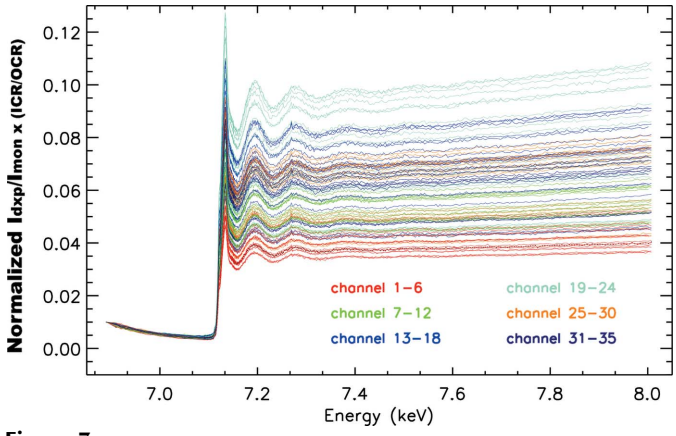

Figure 7
The full usable dataset for the raw detector signal divided by the monitor signal, normalized by the detector dead-time correction, and each scan is normalized to the first pixel average.

normalized to the first pixel of the XAFS scan. This scaling lowers the variance in the pre-edge region but accentuates the variance at and above the edge. Significantly, the dependence of the variance in signal with the position of the detector channel evident in the earlier plots (Figs. 3 and 4) is retained and amplified with this normalization of the data. Therefore the increase in variation is not due to random noise associated with the first data point. Hence, the specific approach of Fig. 7 dramatically increases the standard errors and is *not* justified by the data.

6. Theoretical interpretation of segmented fluorescence data

An alternate explanation of the systematic variation in the signal from the different detector elements may be related to the geometry-dependent efficiency of fluorescence detection. This requires an understanding of the nature of the fluorescence process, and perhaps initially to revisit the absorption process. Absorption is conventionally given by the Beer–Lambert equation,

$$I = I_0 \exp[-(\mu/\rho)(\rho t)], \tag{1}$$

where I_0 is the incident X-ray beam intensity, (μ/ρ) is the X-ray mass absorption coefficient of the material for the energy of the X-ray beam, and t is the thickness of the foil. The beauty of this is that the negative values of the natural logarithms of the measured ratios of I/I_0,

$$-\ln(I/I_0) = [(\mu/\rho)(\rho t)] = \mu t, \tag{2}$$

plotted against t [or (ρt)], fall on a straight line with slope μ [or (μ/ρ)]. Hence the mass absorption coefficient, the photoelectric coefficients, the scattering components and the form factors of the material can be directly evaluated from the logarithm of the normalized ratio. Of course, this requires careful correction for detector efficiencies and air path (Tran *et al.*, 2003*b*), scattering (Tran *et al.*, 2004), harmonics (Tran *et al.*, 2003*a*), detector linearity (Barnea *et al.*, 2011), energy calibration (Rae *et al.*, 2010*b*), thickness calibration (Tran *et*

al., 2004), bandwidth (de Jonge *et al.*, 2004*a*) *etc.*, but yields a highly accurate measurement of the coefficients with the correct scaling and relative amplitudes for processing using, for example, *XERT* for XAFS analysis (Chantler, 2009, 2010).

However, for fluorescence, to first order we have

$$I_{f_0} = f\, I_0\{1 - \exp[-(\mu/\rho)(\rho t_{\rm p})]\}, \tag{3}$$

where I_{f_0} is the total number of fluorescence photons produced, $t_{\rm p}$ is the pathlength through the sample, f is the fluorescence yield for the probability of producing a fluorescent photon (in normal geometries specifically a $K\alpha$ photon if the experiment is around the K-shell, or a characteristic L-shell photon if the experiment is around the L-shell *etc.*) after the process of photoabsorption and photoionization [we should label (μ/ρ) with the subscript 'pe' for the photoelectric effect only, and an asterisk indicating that only the component absorbed in the active centre producing a fluorescent photon is relevant, *i.e.* as $(\mu/\rho)^*_{\rm pe}$]. Again, to first order, these fluorescent photons are emitted isotropically. Some simple analysis suggests an improved functional as

$$I_f = \frac{f\, I_0 \Omega (\mu/\rho)^*_{\rm pe}/4\pi \cos\theta_{\rm inc}}{[(\mu/\rho)/\cos\theta_{\rm inc}] + [(\mu_f/\rho)/(\cos\theta_{\rm out})]} \times \left\{A - \exp\left[-\frac{(\mu/\rho)(\rho t)}{\cos\theta_{\rm inc}} - \frac{(\mu_f/\rho)(\rho t)}{\cos\theta_{\rm out}}\right]\right\}, \tag{4}$$

where t is the depth of penetration into the material, Ω is the solid angle subtended by the detector channel, $\theta_{\rm inc}$ is the angle of incidence of the incident X-rays (relative to the normal), $\theta_{\rm out}$ is the angle of emission of the fluorescent X-rays (relative to the normal), (μ_f/ρ) is the mass attenuation coefficient of the material for the fluorescent photon energy, and the integration constant A may be (loosely) estimated as $A = 1$ (the limit as $t \rightarrow 0\,\mu$m). Some similar derivations have been presented elsewhere (Bunker, 2010; Troger *et al.*, 1992).

Given a realistic geometry, the number of fluorescent X-rays detected (in the relevant region of interest) should include losses due to air path (air), detector windows (w) *etc.* and detector quantum efficiencies overall (ε) as

$$\frac{I_{f_{\rm detected}}}{I_{0_{\rm monitored}}} = \frac{I_f}{I_0}\left[\frac{\varepsilon_{\rm det}(E)}{\varepsilon_{\rm mon}(E)}\right] \exp\left[-\frac{(\mu_f/\rho)_{\rm air}(\rho t_{\rm air})}{\cos\theta_{\rm air}} - \frac{(\mu_f/\rho)_{\rm w}(\rho t_{\rm w})}{\cos\theta_{\rm w}}\right]. \tag{5}$$

$t_{\rm air}/\cos\theta_{\rm air}$ is the pathlength from the sample (surface) to the front face of the detector (window) and $t_{\rm w}/\cos\theta_{\rm w}$ is the pathlength through, for example, a detector window of thickness $t_{\rm w}$. Here we have emphasized the energy dependence of the relative efficiencies of the detectors, as this is quite important. We note that these formulae do not fully incorporate such higher-order effects as bandpass, divergence, polarization and nano-roughness (Glover *et al.*, 2009), but are nonetheless very useful. These latter, finer, effects will be addressed in a subsequent publication.

For normal fluorescence XAFS geometries, the multi-element detector is placed at 90° to the incident beam, with the fluorescent sample, solid or solution, placed at an angle of 45° to the incident beam. A particular detector channel will correspond to an emission angle $\theta_{\rm out}$ which varies depending upon how close the sample stage is to the detector and its orientation *etc.* Similarly, the air path for the fluorescent X-ray to the detector, and the angle for the window attenuation, may then be given by

$$\theta_{\rm w\,h} \simeq \theta_{\rm air\,h} \simeq \theta_{\rm out\,h} - 45^\circ. \tag{6}$$

Some simple observations ensue:

(i) While the equation is a little complex, several of the components are fixed by the geometry. If they are known, then the information content can be recovered effectively.

(ii) Whereas, after correction for systematics, the absorption geometry can yield a straightforward and beautiful relation from the logarithm of the normalized intensity ratio, for a range of physical and theoretical interpretations, this is not the case for the fluorescence equation.

(iii) If L is the distance from the sample surface to the detector, then we may approximate $\Omega \simeq D/L^2$ where D is the area of the individual detector element.

(iv) $\theta_{\rm out}$ varies across the detector and especially between detector channels, so if the detector channel centres are separated by a distance C (known to good accuracy) and some central detector point is indeed at 45° to the sample surface, then the angle of emission in the plane of incidence is $\theta_{\rm out\,h} = 45^\circ + \tan^{-1}(nC/L)$ where n is the number of channel elements from the central point. In general, owing to misalignment, we should generalize this to $\theta_{\rm out\,h} = \theta_0 + \tan^{-1}(nC/L)$. Note, extremely importantly, that, because of the sample self-absorption, different detector channels with different pathlengths will have very strongly different self-absorption correction factors. In principle, the data can be analysed to correct for self-absorption. Channels on the downstream side of the detector have approximately a single angle, and hence a single self-absorption correction; and, for example, those on the other side (upstream) have a much smaller self-absorption correction. While this self-absorption correction is strongly energy-dependent especially owing to $(\mu/\rho)(E)$, it is monotonic.

(v) Hence the pattern of the data expected from different channels is 'exactly' as given in Figs. 3 and 4; and these can be fitted and corrected for self-absorption to provide a more robust dataset with greater information content.

(vi) In many fluorescent geometries, square channel arrays are deliberately quite close to the sample stage to improve scattered fluorescent signals. Then the solid angle to a particular detector channel is important and we must use $\cos\theta_{\rm out} = \cos\theta_{\rm out\,h}\cos\theta_{\rm out\,v}$ where v is the vertical angle, which is zero in the plane of incidence. Then $\cos\theta_{\rm out\,v} = \tan^{-1}(mC/L)$ where m is the number of channel elements from the plane of incidence in the vertical axis.

(vii) In summary, the main parameters are θ_0 and L, allowing potential reduction of the whole equation to a consistent dataset with maximal information content.

(viii) As has been expressed elsewhere (Newville, 2004; Bunker, 2010), there are two particularly useful limits for fluorescence measurements. In the *thin sample limit* where

$(\mu/\rho)\rho t \ll 1$, the $1 - \exp(X)$ term expands by Taylor series expansion, cancelling the denominator (and the self-absorption correction) so that

$$\frac{I_{f_{\rm detected}}}{I_{0_{\rm monitored}}} = \frac{f\, t\, \Omega (\mu/\rho)^*_{\rm pe}}{4\pi \cos\theta_{\rm inc}} \frac{\varepsilon_{\rm det}(E)}{\varepsilon_{\rm mon}(E)} \times \exp\left[-\frac{(\mu_f/\rho)_{\rm air}(\rho t_{\rm air}) + (\mu_f/\rho)_{\rm w}(\rho t_{\rm w})}{\cos(\theta_{\rm out\,h} - 45^\circ)\cos(\theta_{\rm out\,v})}\right], \quad (7)$$

and to first order the observed intensity ratios are proportional to the photoelectric coefficient and the XAFS structure may be cleanly extracted. This *thin sample limit* is invalid whenever a dispersion between detector elements is observed, *i.e.* almost always.

(ix) The second convenient limit is the *thick dilute sample limit* where $(\mu/\rho)\rho t \gg 1$ but $(\mu/\rho)_{\rm pe} \ll (\mu/\rho)$, the exponential goes to zero yielding

$$\frac{I_{f_{\rm detected}}}{I_{0_{\rm monitored}}} = \frac{f\, \Omega (\mu/\rho)^*_{\rm pe}/4\pi \cos\theta_{\rm inc}}{[(\mu/\rho)/\cos\theta_{\rm inc}] + [(\mu_f/\rho)/\cos\theta_{\rm out}]} \frac{\varepsilon_{\rm det}(E)}{\varepsilon_{\rm mon}(E)} \times \exp\left[-\frac{(\mu_f/\rho)_{\rm air}(\rho t_{\rm air}) + (\mu_f/\rho)_{\rm w}(\rho t_{\rm w})}{\cos(\theta_{\rm out\,h} - 45^\circ)\cos(\theta_{\rm out\,v})}\right]. \quad (8)$$

If and only if the energy dependence of the denominator is small (dominated by scattering coefficients or background absorption), then the angular self-absorption can be modelled and the corrected intensity ratio provides the photoelectric absorption coefficients for theoretical modelling using XAFS analysis. However, for most samples, the thin limit is not obeyed (the self-absorption function which we see in Figs. 3 and 4 would be suppressed). Similarly, for most of the X-ray regime $(\mu/\rho)_{\rm pe}$ is dominant and is not dominated by the scattering coefficients. For a typical metallic XAFS investigation, the concentration must be very low for $(\mu/\rho)^*_{\rm pe}$ of the active fluorescent centre in the sample to be dominated by background absorption $(\mu/\rho)_{\rm pe}$. Then, of course, the signal and statistical precision are also very low.

The absorption coefficient for our current investigation of ferrocene at 10 mM is provided in Fig. 8. While the iron photoelectric absorption coefficient is a small fraction of the solvent photoabsorption, it is still 6% of the total above the edge. Hence, as is often stated, this sort of concentration is a compromise between observing a strong fluorescence signal and introducing large distortions owing to self-absorption. Perhaps of equal concern, especially in this experiment, is that the cell is 2 mm deep. Using the total attenuation in the solution cell indicates that we are far from the 'thick' limit. Hence this identifies systematic errors and distortions of the XAFS spectrum, which will affect strong peaks in the oscillations more than minima.

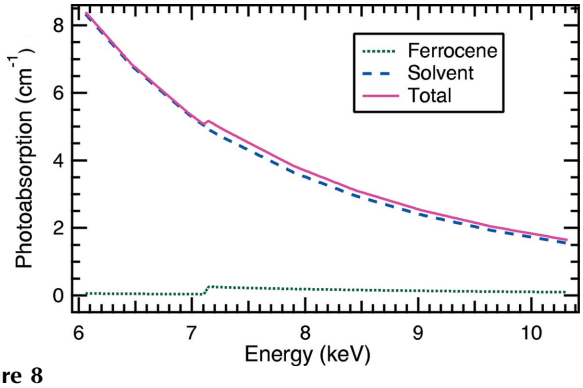

Figure 8
Photoelectric absorption coefficient of the 10 mM iron centre of ferrocene as a component of the total photoabsorption of the solution. Equation (4) will provide more accurate amplitudes for XAFS analysis than equation (8).

Hence, while both extreme limits are used in conventional analysis of fluorescence XAFS, in this case and many others the limiting forms are unrealistic or infeasible, and one should use the full analytic form we have presented. Therefore, in what follows, we use our original equations (4) and (5).

A single horizontal strip of detector pixel outputs is given in Fig. 9, showing the variance observed earlier. The fits of this single strip of channels in the horizontal plane are given in Fig. 10. This is the result from five channels (1, 7, 13, 19, 25). The dispersion is dramatically reduced, thereby proving the accuracy of the model. By also including the vertical angle, we can fit all 34 channels with just two parameters: θ_0 is approximately half a channel width and $L \simeq 50$ mm, and yielding Fig. 11 as the resulting final average of the corrected scans, with much reduced standard error. This analysis has dramatically increased the information content with which to use XAFS analysis to investigate the staggered *versus* eclipsed structures of ferrocene, for example. There is some residual dispersion (of course); part of this will remain intrinsic, and

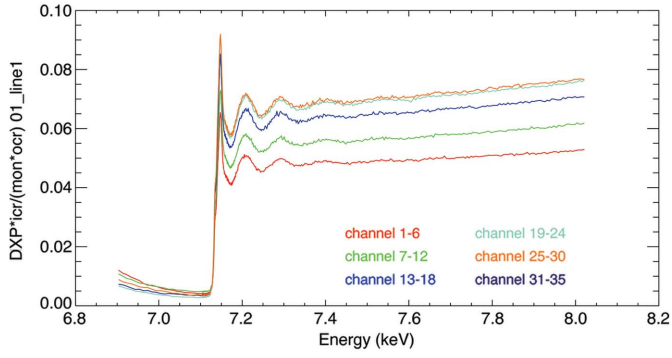

Figure 9
Spectra obtained from a single horizontal line of detector channels.

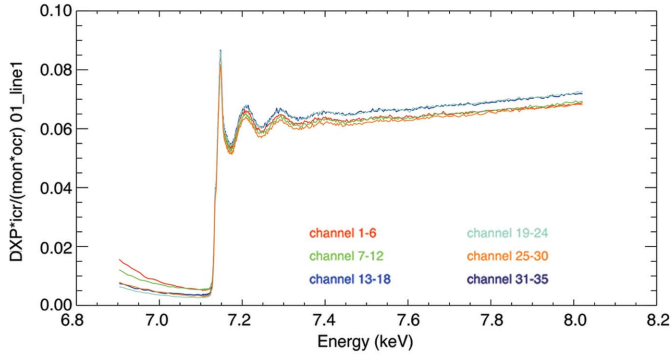

Figure 10
A fitted set of the spectra correcting for self-absorption for a single strip of horizontally arranged detector channels shown in Fig. 9.

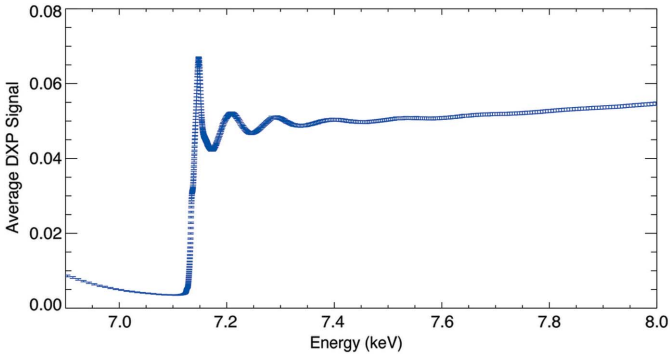

Figure 11
The average signal for the self-absorption fitted spectra with uncertainty. The standard error of the detector channels of the fitted plot is dramatically reduced, and information content of the data is enhanced to possibly define a discriminant for the puzzle of the staggered or eclipsed conformations of ferrocene.

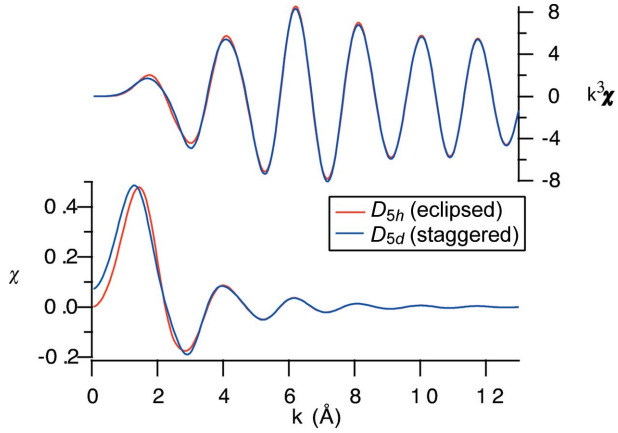

Figure 12
$k^3\chi$ *versus* k plots for ferrocene in staggered (blue) and eclipsed (red) conformations (theory, *FEFF*) (right-handed axis, top plots); and plots of χ *versus* k showing differences up to 10% in χ (lower plots, left-hand axes). These differences are very difficult to measure in standard XAFS analysis.

part may be due to normalization error, pixel response function, dark count correction, efficiency corrections, roughness or non-uniformity. Either way, this final set of standard errors represents explicitly the information content of each measurement.

The job of processing I_0/I for XAFS analysis is a little complex. For absorption measurements, analysis will naturally take the logarithm, while for fluorescence data, under some limits, analysis would extract the information from the ratio directly. Then, a background function, perhaps ideally an atomic or free-atom function, is subtracted and the spectrum above the cut-off energy E_0 is transformed into χ *versus* k plots. Conversely, other theoretical approaches will plot directly against theory in (μ/ρ) *versus* E space, still with a fitting parameter for a starting edge position E_0. Rather than getting embroiled in these details, the standard route for extracting χ from the XAFS spectra will be adopted.

However, the background subtraction algorithm used is the standard *IFEFFIT* spline approach rather than an atomic baseline subtraction; the approach used can certainly distort the XAFS oscillations and features, especially at the far ranges of the dataset. In this case there is distortion at high k ($14.5 < k < 15.0$ Å^{-1}, *i.e.* at the end of the dataset) but the conclusions are robust whether this range is included or excluded from the modelling. Additionally, it is well known that error in the offset energy dramatically distorts χ at low-k (Glover & Chantler, 2007), and is addressed below; but this is separate from the distortion introduced by poor background subtraction, though they can both have a great effect at low k. It is simplest to comment that the low-k background subtraction errors in this analysis are effectively removed by the low-k cut-off to the fitting range, and that low-k errors in structural details are not addressed by the raising or lowering of the heights of particular oscillations.

7. XAFS analytical procedures

The theoretical difference signatures between D_{5d} and D_{5h} conformations, illustrated in Fig. 12, are moderately large up to a k of about 4 Å^{-1} but remain observable up to a k of 8 or 12 Å^{-1} (with a magnitude of 0.01 and rapidly decreasing). These differences are very difficult to measure in standard XAFS analysis because uncertainties are not propagated and correlation amongst fitting parameters are often large. It is also well known that *FEFF* (especially early versions, and other muffin-tin theoretical approaches) has great difficulty in experimental modelling at low k, which is one of the key reasons for standard analysis to use both a window function (*i.e.* to fit over a restricted k-range) and a k^2 or k^3 weighting. Some recent theoretical developments are improving the modelling of low-k, including *FEFF* and *FDM* approaches, but for consistency these are not the subject of the current discussion, and instead current versions for *XFIT* and standard versions for *IFEFFIT* are compared on a common footing.

We have investigated three approaches to modelling the experimental χ *versus* k spectrum: (i) *XFIT* using the corrected averaged plot but no propagation of uncertainty; that is, having uniform error bars throughout k, or uniform error bars for data weighted by k^2 or k^3; (ii) *IFEFFIT* using the corrected averaged plot but no propagation of uncertainty; (iii) an *IFEFFIT*-like approach developed by our group (Smale *et al.*, 2006; Glover *et al.*, 2010) but incorporating, propagating and fitting the derived error bars as explicit in Fig. 11.

The *IFEFFIT*-like approach follows a traditional least-squares fitting procedure of $\chi(k)$, but with uncertainties in $\chi(k)$ at each k-point provided by the propagation of experimental uncertainty derived from the experimental variance and in principle from any systematic uncertainties, point-by-point. Hence any fitting of $\chi(k)$ *versus* k, or $k^2\chi(k)$ *versus* k, or $k^3\chi(k)$ *versus* k, for example, will return uncertainties in fitted parameters given by the least-squares method but with accurate amplitudes based upon the standard errors of the source data.

At least in the standard implementation of *XFIT*, the fit appears to follow an R-factor minimization rather than a least-

squares technique or a χ^2 minimization. *IFEFFIT* follows a conventional unweighted χ^2 minimization. The *IFEFFIT*-like routine follows an explicitly weighted χ^2 minimization. It is also known that χ^2 values output by *IFEFFIT* (or *XFIT*) are not reliable measures of goodness-of-fit. Hence, general advice is to consider the relative χ^2_r of disparate models and to determine the best fit from the smallest value. We can use this approach in all three methods.

8. XAFS results: the conformation of ferrocene

We now focus on the question of the stereochemistry of ferrocene molecules. Clearly, the results depend upon k-range. We have investigated $0 < k < 15$, $1 < k < 15$, $2 < k < 15$, $3.2 < k < 15$, $3.6 < k < 14.5$, and other ranges (k is implicitly and always given in units of Å^{-1} as standard throughout the rest of this paper). In a range of detailed investigations we have completed a first fit, recomputed *FEFF* paths with the new spacing, completed a second fit, recomputed the *FEFF* paths with the new atom sites, and fitted for the translation of the cyclopentadienyl ring above the iron atom along the fivefold axis. The relative χ^2 values obtained remained robust in these later cycles of refinement. That is, they generally improved by small amounts or were stable to within uncertainty. We emphasize that each comparison between fitting packages was made using exactly the same k-range, k-weighting, model and constraints.

FEFF involves computation of independent photoelectron wave paths up to some maximum pathlength $r_{\max}$. We investigated the modelling for $r_{\max} = 4, 5, 6$ Å. In general, consistent (robust) conclusions were obtained for the choice of conformation within these series; but χ^2_r was clearly improved with $r_{\max} = 5$ Å compared with $r_{\max} = 4$ Å. We view this as an endorsement at this level of the *FEFF* code, and an endorsement of the significance of outer paths in the molecular system, even though the individual amplitudes are fairly weak. Hence further discussion will correspond to the $r_{\max} = 5$ Å option only.

XFIT analysis returned claimed χ^2_r values typically around 4.3, 0.078, 0.005 and 0.0003. Despite the unphysical χ^2_r values, all modelling (different k ranges and other variations) suggested that the 'eclipsed' conformation represented the better fit. Surprisingly, the relative χ^2_r values were typically a factor of 2.5 or 8 lower for the eclipsed model than for the staggered model, despite the theoretical *FEFF* differences being really quite small and the number of degrees of freedom being identical. Irrespective of the cause of this, we must be aware that the lack of error bar propagation will question this conclusion.

IFEFFIT analysis returned claimed χ^2_r values typically around 530, 105, 45, heavily dependent upon the choice of k-weighting, as expected. Generally, the conventional implementation without propagation of errors for the individual measurements, with variable windowed k-range, suggested that the staggered conformation represented the correct structure. Equally surprisingly, the significance of the result appeared generally highly conclusive, with the staggered conformation yielding a χ^2_r a factor of three or two better than that for the eclipsed conformation. However, there were exceptions, including using k^0-weighting or the $0 < k < 15$ range. Most analysts would conclude that this is normal and expected, and that neither k^0-weighting nor using $0 < k < 15$ yields a robust or reliable result for *FEFF*. Analysis of XAFS normally assumes a weighting of k^2 or k^3 and a lower window cut-off of approximately $k = 3$ precisely because the user is investigating the higher oscillation periodicity and because the theoretical formalism is designed for this range, and not for the low-k range. The evidence of our analysis supports the conventional need for a window or k-range when modelling XAFS, and a lower cut-off of order $k = 3$ (and in the case of *IFEFFIT* and *XFIT* that k^2 or k^3 weighting should be used), since we observe that the fitting parameters tend to become unreliable; the residuals remain large, especially in the high-k range, S_0^2 tends to drop to 0.7, for example, and occasionally σ^2, the thermal parameter, becomes negative. With this caveat, then the conclusion from *IFEFFIT* on this moderately accurate dataset would be conclusively in favour of the staggered conformation.

The key point here is that different fitting methods can yield dramatically different conclusions under these circumstances, especially when the experimental uncertainty is not propagated, and when an individual χ^2_r output cannot be directly investigated but can only be compared in a relative manner.

Our new *IFEFFIT*-like analysis propagates the experimental uncertainties, though as mentioned above can explicitly distort them to mimic the k^n weighting approaches. In general, these k^2 and k^3 weighting options emphasize the high-k oscillations of XAFS at the expense of the natural error bars and the first one or two oscillations. Once again, a robust conclusion was obtained. χ^2_r values only have proper meaning for the k^0-weighting option, *i.e.* where the uncertainties are explicitly propagated without distortion. In this case we find naturally that χ^2_r is smaller when a smaller k-range is fitted, and values vary typically from $\chi^2_r = 4$ ($0 < k < 15$, $r_{\max} = 4$) to $\chi^2_r = 1.7$ ($0 < k < 15$, $r_{\max} = 5$) or $\chi^2_r = 2.1$ ($1 < k < 15$, $r_{\max} = 4$) to $\chi^2_r = 1.02$ ($1 < k < 15$, $r_{\max} = 5$), or χ^2_r values of 0.4 (for $2 < k < 15$, $r_{\max} = 5$). A dramatic success of our approach to report is that we have achieved χ^2_r values of order unity.

However, these k ranges are not particularly robust nor reliable. The pattern of XAFS oscillations is poorly followed and the discrepancies at low k must force us to question the reliability of the fit and the *FEFF* theory to cover this range. In some cases, S_0^2 dropped to below 0.7, and σ^2 became negative. Hence, while the formalism fitted the model, it could hardly be claimed that the parameters retained their original meaning. *IFEFFIT* and similar programs permit constraints and restraints. It is generally reasonable to require $0.85 < S_0^2 < 1.1$ and $\sigma^2 > 0$. When this is not observed in a free fit, it can simply mean that the parameters are not independent but are correlated, and hence ill-defined as specific physical values.

Inspection of the residuals, or the experiment and fitted profiles themselves, shows that the theoretical (fitted) location of the first peak at $k \simeq 1.5$ is far from the experimental value of $k \simeq 1.8$; similarly, a pronounced asymmetry on the low-k side

of the second extrema (trough) at $k \simeq 3$ is quite poorly modelled in all fits. This is a strong justification for focusing on the k-range where the model function appears to match the experimental data and yield a reliable result. Additionally, the experimental range for $14.5 < k < 15.0$, the last part of the experimental data, appears distorted, quite possibly by the (standard) background-subtraction algorithm. Hence, the range of reliable data where by-eye agreement between experiment and fitted theory is obtained suggests a useful maximal fitting range of $3.6 < k < 14.5$ for this dataset. Explicit uncertainties computed above $k = 14.5$ also argue for this upper limit.

None of these uncertainty-propagating *IFEFFIT*-like analyses produce factors of two or eight in relative χ_r^2 between the fits of the different conformations. This is another strong endorsement of the procedure. In general, the preferred model has an absolute χ_r^2 (on this basis) reduced by approximately 10–15% compared with the less optimized model. This fit does not attempt to model $0 < k < 3.6$, but the qualitative structure in this region is clearly not modelled by the fit or theory. In observing the basic theoretical differences illustrated by Fig. 12, the differences in relative χ^2 seem much more reasonable, or even remarkable given the small signatures of difference and the relatively constrained k-range. Such a discrepancy is in principle sufficient to draw a conclusion, though such a conclusion would be fraught if uncertainties were not propagated, as we have just discussed.

Narrower k ranges yielded $\chi_r^2 = 0.09$ (for $3.6 < k < 14.5$, $r_{\max} = 5$). This suggests that our derived experimental uncertainties are modest overestimates by an average factor of perhaps three. While the determination of individual point uncertainty in this approach was independent, there was clearly some correlated uncertainty (between channels and points in k) which would lead to overestimated input standard errors, and hence one might expect χ_r^2 values somewhat less than unity when the theoretical model is valid. This is indeed what is observed. Future work will seek to improve the determination of fluorescence and other experimental uncertainty to yield an optimum dataset for analysis.

The final results are plotted in Figs 13–17. Table 1 presents the parameters fitted and their uncertainties for the optimized model discussed. Both χ *versus* k and $k^2\chi$ are presented to permit comparison with typical standard fits, and to allow the central region of XAFS oscillations to be inspected more clearly.

For $3.6 < k < 14.5$, $r_{\max} = 5$, the eclipsed conformation provides an improved fit by 11% compared with the staggered model, with no constraints or restraints. The deviation above $k = 14.5$ (not fitted in this or any model) is most likely due to the standard background subtraction being inadequate near the maximum of the dataset. When the weighting is distorted by k^2 or k^3 weighting, the same result is obtained, with typically 3% discrepancy. Interestingly, the same conclusion regarding conformation is obtained for the $2 < k < 14.5$, $r_{\max} = 5$, range, although a constraint of S_0^2 is needed to prevent the physical parameter from becoming less meaningful. The preference of conformation does not depend upon this constraint, but the meaning of the parameters does.

Table 1
Fitted parameters for the minimum χ_r^2 models for each conformer.

	Conformation	
	Eclipsed	Staggered
Fitted parameters		
χ_r^2	0.089	0.099
ΔE_0 offset (eV)	−1.72 ± 0.94	−2.21 ± 1.04
$1 + \alpha$ scaling of lattice	1.0036 ± 0.0037	1.0021 ± 0.0041
σ^2 thermal parameter	0.0049 ± 0.0013	0.0048 ± 0.0014
S_0^2 amplitude reduction	1.069 ± 0.086	1.063 ± 0.093
Fixed values (Å)		
Fe(x, y, z)	0, 0, 0	0, 0, 0
C1(x, y, z)	1.6555, 1.2007, 0.0000	1.6555, 1.2007, 0.0000
C2(x, y, z)	−1.6555, 1.2007, 0.0000	−1.6555, −1.2007, 0.0000
C3(x, y, z)	1.6555, −0.9714, 0.7058	−1.6555, 0.9714, 0.7058
C4(x, y, z)	−1.6555, −0.9714, 0.7058	1.6555, −0.9714, 0.7058
C5(x, y, z)	1.6555, −0.9714, −0.7058	−1.6555, 0.9714, −0.7058
C6(x, y, z)	−1.6555, −0.9714, −0.7058	1.6555, −0.9714, −0.7058
C7(x, y, z)	1.6555, 0.3710, 1.1420	1.6555, 0.3710, 1.1420
C8(x, y, z)	−1.6555, 0.3710, 1.1420	−1.6555, −0.3710, 1.1420
C9(x, y, z)	1.6555, 0.3710, −1.1420	1.6555, 0.3710, −1.1420
C10(x, y, z)	−1.6555, 0.3710, −1.1420	−1.6555, −0.3710, −1.1420
Derived parameters including α scale uncertainty (Å)		
Fe−C_5†	1.6555 (1.0036 ± 0.0037)	1.6555 (1.0021 ± 0.0041)
Fe−C1	2.045 (1.0036 ± 0.0037)	2.045 (1.0021 ± 0.0041)
C−C	1.4116 (1.0036 ± 0.0037)	1.4116 (1.0021 ± 0.0041)
Fe−C_5	1.6615 ± 0.0061	1.6590 ± 0.0069
Fe−C1	2.0524 ± 0.0076	2.0493 ± 0.0084
C−C	1.4167 ± 0.0052	1.4146 ± 0.0058

† C_5 is the centroid of the C_5 ring.

The figures emphasize the difficulty of drawing conclusions on *XAFS* interpretation by eye, or by any χ^2 measure which does not represent the information content of the dataset, *i.e.* the experimental uncertainties. The standard plots are represented by Figs 14 and 16 but these do not show the significance of the reduction in χ_r^2 owing to the improved fitting of the

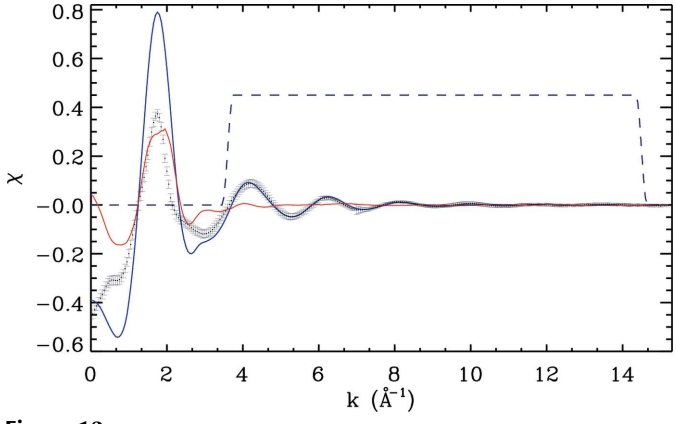

Figure 13
Fitted χ *versus* k plots for ferrocene in staggered conformation are shown. Blue line: theory, *FEFF*. The window function for $3.6 < k < 14.5$ is shown by the dashed line. Diamonds: experiment with propagated uncertainty as standard error (error bars). Red line: residual light. Discrepancies in the medium and high k region are not visible.

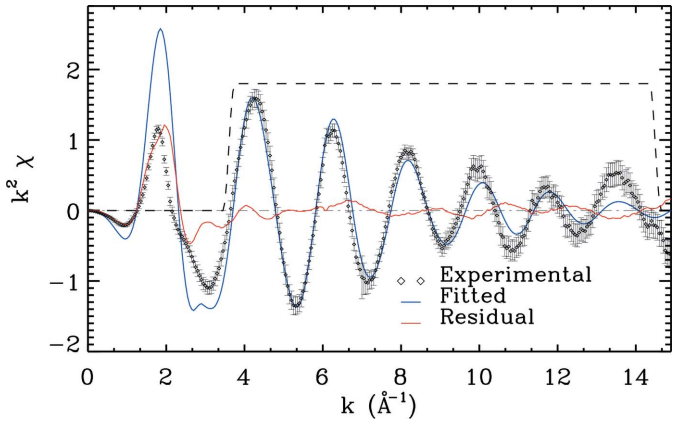

Figure 14
Fitted $k^2\chi$ *versus* k plots for ferrocene in staggered conformation. Legend as per Fig. 13. The uncertainties at higher k define the information content of the dataset. While the periodicity and general fit is very good, amplitude and attenuation discrepancies are visible.

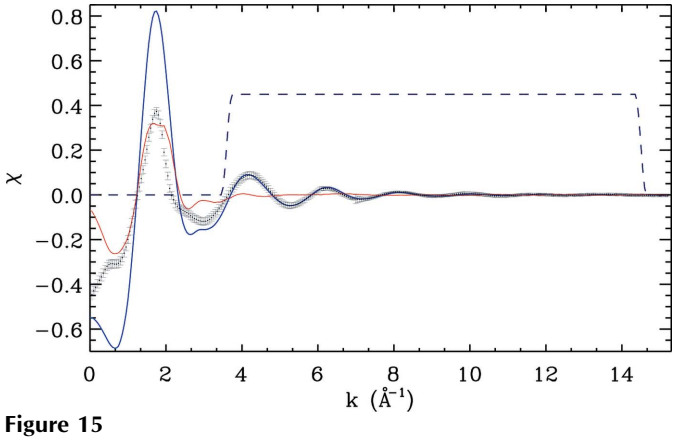

Figure 15
Fitted χ *versus* k plots for ferrocene in eclipsed conformation are shown. Blue line: theory, *FEFF*. The window function for $3.6 < k < 14.5$ is shown by the dashed line. Diamonds: experiment with propagated uncertainty as standard error (error bars). Red line: residual light. By eye, improvements of 11% in χ^2_r are not visible.

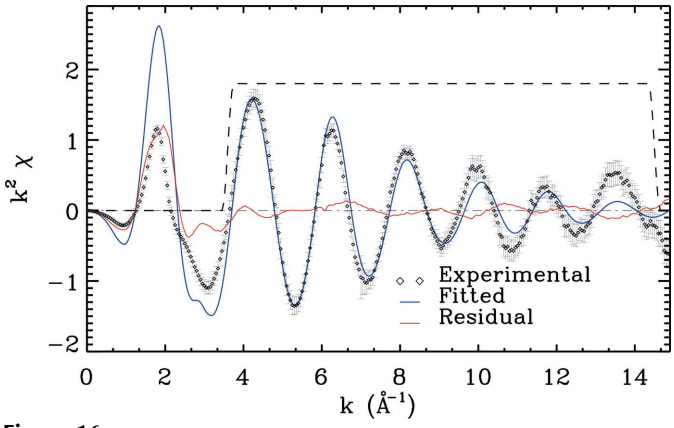

Figure 16
Fitted $k^2\chi$ *versus* k plots for ferrocene in eclipsed conformation. Legend as per Fig. 15. Here the uncertainties at higher k give the information content of the dataset. While the periodicity and general fit is very good, amplitude and attenuation discrepancies remain visible.

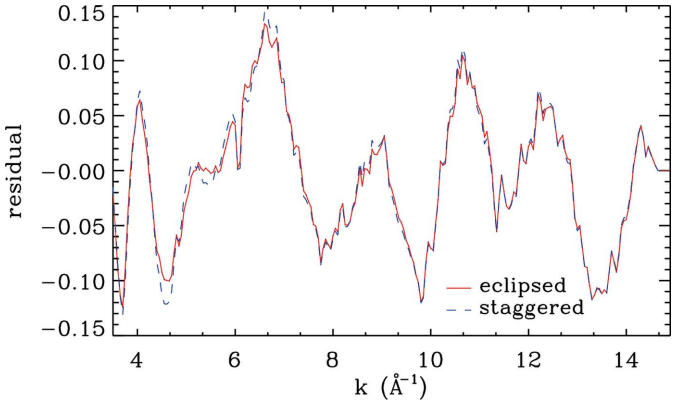

Figure 17
Residual plots for the eclipsed (red line) *versus* staggered (dashed line) conformations. The dominant signature, as theoretically expected, lies in the lower k range, and repeatedly suggests the eclipsed conformation as a better representation of the data compared with the staggered conformation. This is in agreement with the broad brushstroke of crystallographic determinations.

features at $k \simeq 4.3$ and $k \simeq 5.3$. The differences are subtle, but become visible in the residual Fig. 17. The disorder observed at room temperature for crystalline ferrocene is interestingly not apparent in these plots; that is, the information content does not appear blurred beyond a plausible thermal parameter. The residuals display strong deviation at low k, which is the cause for the standard low-k cut-off, and which we would argue is due to the theory presented needing improvement in the future. Within the range of k fitted in any of the schemes, there is a common pattern of residual or 'mis-fit' apparent in Fig. 17. The shared oscillation of both conformers with k seems significantly larger than the difference between them (11%). It would be helpful to better understand the remaining residual. Since the residual has an oscillatory component, it would be interesting to test whether this might be due to multiple scattering, or a longer distance, perhaps to the solvent ion. Fourier transforms of the fits are not presented here for two reasons: one is the difficulty of interpreting, by eye, fits without error propagation (notoriously difficult for Fourier transforms, as known from crystallography) and the dangers of inconsistent transforms especially due to E_0 or k offsets of the origin; and secondly because the *raison d'etre* of this manuscript has been the introduction of error propagation in the analysis of fluorescence spectra using the conventional approaches of *XFIT* and *IFEFFIT*. However, we can confirm that Fourier transform analysis, combined with judicious filtering, can demonstrate that much of the structure of the residual between experiment and either modelling appears in unphysical low-R space far from any bonding region.

In principle, Table 1 permits conclusions to be drawn about bond distances, thermal parameters, amplitude reduction coefficients and energy offsets, as summarized in Table 2. We note that we have made modest investigation of additional parameter modelling (*e.g.* by adding a lateral expansion of the C_5-rings) and that this yielded no significant reduction of χ^2_r. The simplest statement of the final analysis is that the structure is converged and that remaining variation in α (scale of molecular structure) and S_0^2 (scale of backscattering/oscillation amplitude) are consistent with no change within uncertainty;

Table 2
Comparison of experimental bond lengths (Å) and theoretical predictions.

Bond *T* Lattice	XAFS eclipsed 10 K –	e-scattering†	Neutron‡ 173 K	X-ray diffraction§ 98 K Orthorhombic	XRD¶ 101 K Triclinic	XRD†† 173 K Monoclinic	MP2‡‡	CCSD/T‡‡	This study§§ Theory
Fe—C1	2.0524 ± 0.0076	2.064 ± 0.003	± 0.003–0.005	2.056, 2.059 ± 0.005	2.046, 2.052 ± 0.007	2.033–	1.910	2.056	2.065
Range		–	[2.005–2.050]	[2.051–2.062]	[2.041–2.052]	[2.017–2.048]			
C—C	1.4167 ± 0.0052	1.440 ± 0.002	± 0.005–0.009	1.429, 1.431 ± 0.006	1.426, 1.433 ± 0.007	1.395–	1.441	1.433	1.428
Range		–	[1.349–1.468]	[1.421–1.437]	[1.423–1.429]	[1.346–1.441]			
Fe—C_5	1.6615 ± 0.0061	1.660 ± 0.003	–	1.658 ± 0.006	1.646 ± 0.007	1.651–	1.464	1.655	1.670

† Haaland & Nilsson (1968). ‡ Takusagawa & Koetzle (1979). § Seiler & Dunitz (1982). ¶ Seiler & Dunitz (1979*b*). †† Seiler & Dunitz (1979*a*). ‡‡ Coriani *et al.* (2006). §§ B3LYP/m6-31G model.

that there is a reasonable thermal parameter reflecting actual ellipsoids or structural disorder; and that the δE_0 offset is weakly determined; however, these will be explored in a later investigation.

At this point, although intriguing, the fitting uncertainties speak for themselves. While the fitting uncertainties are quite competitive with techniques of electron scattering, neutron diffraction and X-ray crystallography, we must remind ourselves that each is asking a different question of a different sample. The relative consistency in this sense is then remarkable.

Past analyses of XAFS data from ferrocene (Vedrinskii *et al.*, 1991; Shuvayev *et al.*, 1985) noted the importance of the spherical wave approach and the muffin-tin potentials (both used in this analysis) but showed no sensitivity to conformation. They claimed that discrepancies in the residuals were due to multiple-scattering processes discussed by Ruiz-Lopez *et al.* (1988), but noted significant discrepancy from analytical models of two other groups (Teo & Lee, 1979; McKale *et al.*, 1988). Perhaps the most detailed analysis of the XAFS of ferrocene was completed in 1988 (Ruiz-Lopez *et al.*, 1988), with an extensive discussion of the development of theoretical calculations and XAFS data extending to k = 16.5 Å^{-1} compared with our k = 15.5 Å^{-1} and an earlier k = 12 Å^{-1} (Cramer *et al.*, 1976). They noted that no useful comparison could be made with the earlier dataset owing to this range over k-space. Perhaps a more poignant reason is that the early work (Cramer *et al.*, 1976) used analytic methods appropriate for the time but which would, even in 1988, have been regarded as superceded. We note that their Fig. 5 displays significant oscillations in the Fourier transform which appear more as artefacts than as bonding orbitals.

The 1988 study (Ruiz-Lopez *et al.*, 1988) concluded that double and triple scattering paths were explicitly necessary and observed in the structure; however, they did not observe any differences between the conformations which would for example have been revealed by double scattering paths, as has been done in this study. Further, full XANES analysis on the conformers was found to yield a null result; that is, there was no distinction between conformers. This conclusion may be attributed to the short range of XANES investigated, but it lies in conflict with our theoretical results for the XAFS structure. One of the difficulties of comparison of this analysis lies in their predominant use of the Fourier transform (R-space) presentation and fitting, with filtering or otherwise, which of course can be quite sensitive to the origin for the transform. Additionally, a recent paper (Schnitzler *et al.*, 2006) discusses ferrocene XAFS and presents near-edge XANES, and a Fourier transform thereof, but without further analysis nor discussion.

The bond lengths Fe—C and C—C, as well as the distance between Fe and the centre of cyclopentadienyl (C_5H_5) rings, Fe—C_5 in the eclipsed ferrocene (D_{5h}) listed in Table 2, agree well between the experiments and available theory. For example, our experimental Fe—C bond length is given by 2.0524 ± 0.0076 Å (T = 10 K), which is in an excellent agreement with 2.064 ± 0.003 Å of Haaland & Nilsson (1968) using electron scattering techniques, 2.056 or 2.059 ± 0.005 Å (T = 98 K) in an orthorhombic symmetry (Seiler & Dunitz, 1982) and 2.046 or 2.052 ± 0.007 Å (T = 101 K) in a triclinic symmetry (Seiler & Dunitz, 1979*b*), together with other earlier X-ray determinations such as the presumed monoclinic determination at 2.033 Å (T = 173 K) (Seiler & Dunitz, 1979*a*). The three results from the same research group suggest that the Fe—C bond slightly decreases as the temperature increases.

We have computed potentials and bond lengths within the density functional theory (DFT) based B3LYP/m6-31G(d) formalism (Mitin *et al.*, 2003). Given that the measured Fe—C bond length varies from 2.005 Å to 2.064 Å, the present theoretical value of 2.065 Å using the DFT-based B3LYP/m6-31G(d) is well within the experimental range and uncertainties. In addition, this Fe—C bond length is significantly longer than the Fe—C length of diatomic iron carbide FeC($X^3\Delta$) at 1.581 Å calculated by recent multi-reference variational *ab initio* models (Tzeli & Mavridis, 2010). The measured C—C bond length of ferrocene is in the range 1.395–1.440 Å as given in Table 2. The present measurement gives this bond length as 1.4167 Å, which is supported by our DFT-based theory of 1.428 Å. The weighted mean C—C bond length for cyclopentane is given by 1.483 ± 0.009 Å (Song *et al.*, 2004). The results for the C—C bond length of ferrocene indicate that the cyclopentane rings in ferrocene exhibit character between C—C and C═C bonding, leaning more towards the C═C side than cyclopentane. Finally, the distance between Fe and the centre of the pentagonal carbon ring is given by 1.6615 Å from the present measurement, which agrees well with 1.670 Å from our theoretical calculations. Of the three available

quantum mechanical models, that is, MP2, CCSD(T) and DFT-B3LYP, the present DFT model can provide excellent results with respect to accuracy and computational costs.

Neutron diffraction and X-ray crystallographic determinations are of a crystal lattice, with different determined space groups in different phases at different temperatures; however, the input standard deviations show that the technique of fluorescence XAFS can be competitive even for biometallic systems, if and only if the intrinsic experimental uncertainties are propagated. Additionally, the method of XAFS deliberately measures dynamic (and static) bond lengths, rather than differences of mean lattice parameters. The implications of this investigation particularly include an understanding of the strength of π bonding (multi-centre bonding) in organometallics; and key implications about crystallographic past determinations and space groups. There are numerous other implications of a biochemical nature, but we will address these in further investigations.

Can this conclusion regarding conformation be dominated by systematic uncertainties or random noise in the dataset of individual point errors, or incomplete convergence or the inadequacy of the full model used? Yes, of course. We have shown that a very small signature which is really quite important can be investigated and a tentative conclusion can be made on the basis of critical error analysis, self-absorption correction, uncertainty propagation, and a meticulous investigation of standard *XAFS* fitting. The conclusion has withstood this analysis, and yet the future must be much brighter still.

Theory has much to say on the crystallographic determinations and on the ferrocene solution modelling. According to accurate quantum mechanical calculations (Coriani *et al.*, 2006) as well as the present theoretical calculation using DFT models (Xu *et al.*, 2003), the staggered conformation appears not to be the energetically most stable form. However, this molecular (DFT) simulation may or may not include crystal packing energies and the requirements of particular space groups. Of course, if a crystal is formed in, for example, orthorhombic symmetry, there will be requirements which may constrain the lattice to be, for example, perfectly staggered or eclipsed. However, in solution there may be significant potential contributed by the surrounding solvent molecules which are certainly not modelled in the current XAFS investigation, and are certainly not present in the crystallographic analyses. In other words, the current best evidence for the conformation of ferrocene in solution is provided by the data of this paper.

The energy difference between the staggered and eclipsed structures is very small, approximately 2 kJ mol^{-1} based on the present DFT-based B3LYP/m6-31G calculations, or $\sim$4 kJ mol^{-1} (Haaland & Nilsson, 1968; Haaland, 1979). This energy is strongly dependent on the model employed and ranges from 0.42 kJ mol^{-1} using HF to 4.8 kJ mol^{-1} using CCSD(T) (Coriani *et al.*, 2006). The present DFT-based models give 3.2 kJ mol^{-1} using the PBE0/6-31++G(d,p) model and 2.4 kJ mol^{-1} using the B3LYP/m6-31G(d) model. Despite quite different ΔE values between the ferrocene conformers obtained from various quantum mechanical models, the eclipsed (D_{5h}) conformer is consistently found to be energetically more stable that the staggered (D_{5d}) conformer. At room temperature (300 K), if the energy difference between D_{5d} and D_{5h} is given by an experimental value of 3.8 kJ mol^{-1}, the ratio of D_{5d} and D_{5h} may be approximately 1:5. If this energy difference is 0.42 kJ mol^{-1} as obtained using the HF model, the ratio will become almost 1:1. In either case the thermal ellipsoids would be quite dominant. These questions cannot be explored with the current dataset in isolation. However, this is a ripe area for future research.

9. Conclusions

Fluorescence data and absorption data can be collected to provide independent uncertainty estimates for the whole range of XAFS. This approach will automatically produce weights which can and should be used in XAFS or XANES analysis. The theoretical model of the fluorescence signal will not normally follow the ideal limits. Accurate and valuable data can be obtained from XAFS investigations on concentrated, crystalline or dilute systems, including for standards such as, in this case, ferrocene, and explicitly including disordered systems. Subtle questions which previously could not be contemplated using XAFS or fluorescence detection can certainly be investigated, especially including subtle conformation alternatives in dilute non-crystalline systems. Reflecting upon the divergence of apparent conclusion from different methods without error propagation and hence without a statistically valid χ^2_r test of significance or information content, there is an urgent need to propagate experimental errors and determine accurate fitted χ^2_r to confirm the significance of observations. Noting the subtle distinction between the conformers of ferrocene, we are looking towards incisive theoretical analysis of XANES data, as this may be insightful. Ideally the two should complement one another.

Further work will investigate alternative and advanced theoretical approaches to this type of problem, especially including agreement over a more extended range of k. Further work will also investigate more optimized fluorescence data collection, and improved datasets. However, this work has observed the signature of conformation changes in dilute disordered ferrocene solution, in the central XAFS region, and has observed that for this particular system the eclipsed D_{5h} conformation is indicated by the data and the theory modelled. Parameters have been derived with uncertainties which are competitive, and with parameters which are loosely consistent with low-temperature crystallographic determinations of pure ferrocene phases. Latest theory in particular supports the bond lengths, more so than earlier theory or experimental data.

This work was supported by the Australian Synchrotron Research Program which is funded by the Commonwealth of Australia under the Major National Research Facilities Program and by a number of grants of the Australia Research Council. We are grateful for the assistance of Dr Mun-Hon

Cheah at the Australian National Beamline Facility, at beamline 20B of the Photon Factory in Tsukuba, Japan. NM and FW acknowledge Victoria Partnership for Advanced Computing for their supercomputing support.

References

Barnea, Z., Chantler, C. T., Glover, J. L., Grigg, M. W., Islam, M. T., de Jonge, M. D., Rae, N. A. & Tran, C. Q. (2011). *J. Appl. Cryst.* **44**, 281–286.

Bunker, G. (2010). *Introduction to XAFS: A Practical Guide to X-ray Absorption Fine Structure Spectroscopy*, pp. 92–95. Cambridge University Press.

Chantler, C. T. (2009). *Eur. Phys. J. ST*, **169**, 147–153.

Chantler, C. T. (2010). *Radiat. Phys. Chem.* **79**, 117–123.

Coriani, S., Haaland, A., Helgaker, T. & Jorgensen, P. (2006). *Chem. Phys. Chem.* **7**, 245–249.

Cramer, S. P., Eccles, T. K., Kutzler, F., Hodgson, K. O. & Donaich, S. (1976). *J. Am. Chem. Soc.* **98**, 8059–8069.

Ellis, P. J. & Freeman, H. C. (1995*a*). *XFIT*, http://www.esrf.eu/computing/scientific/exafs/xfit.html.

Ellis, P. J. & Freeman, H. C. (1995*b*). *J. Synchrotron Rad.* **2**, 190–195.

Fischer, E. & Wilkinson, G. (1993). In *Nobel Lectures, Chemistry 1971–1980*, edited by S. Forsen. Singapore: World Scientific.

Glover, J. L. & Chantler, C. T. (2007). *Meas. Sci. Tech.* **18**, 2916–2920.

Glover, J. L., Chantler, C. T., Barnea, Z., Rae, N. A. & Tran, C. Q. (2010). *J. Phys. B*, **43**, 085001.

Glover, J. L., Chantler, C. T., Barnea, Z., Rae, N. A., Tran, C. Q., Creagh, D. C., Paterson, D. & Dhal, B. B. (2008). *Phys. Rev. A*, **78**, 052902.

Glover, J. L., Chantler, C. T. & de Jonge, M. D. (2009). *Phys. Lett. A*, **373**, 1177–1180.

Haaland, A. (1979). *Acc. Chem. Res.* **12**, 415–422.

Haaland, A. & Nilsson, J. (1968). *Acta Chem. Scand.* **104**, 2653–2670.

Hardgrove, G. L. & Templeton, D. H. (1959). *Acta Cryst.* **12**, 28–32.

Islam, M. T., Rae, N. A., Glover, J. L., Barnea, Z., de Jonge, M. D., Tran, C. Q., Wang, J. & Chantler, C. T. (2010). *Phys. Rev. A*, **81**, 022903.

Jonge, M. D. de, Barnea, Z. & Chantler, C. T. (2004*a*). *Phys. Rev. A*, **69**, 022717.

Jonge, M. D. de, Barnea, Z., Tran, C. Q. & Chantler, C. T. (2004*b*). *Meas. Sci. Technol.* **15**, 1811–1822.

Jonge, M. D. de, Tran, C. Q., Chantler, C. T., Barnea, Z., Dhal, B. B., Cookson, D. J., Lee, W.-K. & Mashayekhi, A. (2005). *Phys. Rev. A*, **71**, 032702.

Jonge, M. D. de, Tran, C. Q., Chantler, C. T., Barnea, Z., Dhal, B. B., Paterson, D., Kanter, E. P., Southworth, S. H., Young, L., Beno, M. A., Linton, J. A. & Jennings, G. (2007). *Phys. Rev. A*, **75**, 032702.

Kealy, T. & Pauson, P. (1951). *Nature (London)*, **168**, 1039–1040.

McKale, A. G., Veal, B. W., Paulikas, A. P., Chan, S.-K. & Knapp, G. S. (1988). *J. Am. Chem. Soc.* **110**, 3763–3768.

Miller, S., Tebboth, J. & Tremaine, J. F. (1952). *J. Chem. Soc.* pp. 632–635.

Mitin, A. V., Baker, J. & Pulay, P. (2003). *J. Chem. Phys.* **118**, 7775–7783.

Newville, M. (2001). *J. Synchrotron Rad.* **8**, 322–324.

Newville, M. (2004). *Fundamentals of XAFS*, pp. 23–24. Chicago: CARS, University of Chicago.

Rae, N. A., Chantler, C. T., Barnea, Z., de Jonge, M. D., Tran, C. Q. & Hester, J. R. (2010*a*). *Phys. Rev. A*, **81**, 022904.

Rae, N. A., Islam, M. T., Chantler, C. T. & de Jonge, M. D. (2010*b*). *Nucl. Instrum. Methods Phys. Res. A*, **619**, 147–149.

Rehr, J., de Leon, J. M., Zabinsky, S. I. & Albers, R. C. (1991). *J. Am. Chem. Soc.* **113**, 5135–5140.

Ruiz-Lopez, M. F., Loos, M., Goulon, J., Benfatto, M. & Natoli, C. R. (1988). *Chem. Phys.* **121**, 419–437.

Schnitzler, M. C., Mangrich, A. S., Macedo, W. A. A., Ardisson, J. D. & Zarbin, A. J. G. (2006). *Inorg. Chem.* **45**, 10642–10650.

Seiler, P. & Dunitz, J. D. (1979*a*). *Acta Cryst.* B**35**, 2020–2032.

Seiler, P. & Dunitz, J. D. (1979*b*). *Acta Cryst.* B**35**, 1068–1074.

Seiler, P. & Dunitz, J. D. (1980). *Acta Cryst.* B**36**, 2946–2950.

Seiler, P. & Dunitz, J. D. (1982). *Acta Cryst.* B**38**, 1741–1745.

Shuvayev, A. T., Helmer, B. Y., Lyubeznova, T. A., Kraizman, V. L., Mirmilshtein, A. S., Kvacheva, L. D., Novikov, Y. N. & Volpin, M. E. (1985). *Izv. Akad. Nauk SSR*, **49**, 1471–1475.

Smale, L. F., Chantler, C. T., de Jonge, M. D., Barnea, Z. & Tran, C. Q. (2006). *Radiat. Phys. Chem.* **75**, 1559–1563.

Song, Q. B., Hu, Y. H., Shen, T. H. & Jin, Z. M. (2004). *Acta Cryst.* E**60**, o539–o541.

Takusagawa, F. & Koetzle, T. F. (1979). *Acta Cryst.* B**35**, 1074–1081.

Teo, B. K. & Lee, P. A. (1979). *J. Am. Chem. Soc.* **101**, 2815–2832.

Tran, C. Q., Barnea, Z., de Jonge, M. D., Dhal, B. B., Paterson, D., Cookson, D. & Chantler, C. T. (2003*a*). *X-ray Spectrom.* **32**, 69–74.

Tran, C. Q., Chantler, C. T. & Barnea, Z. (2003*b*). *Phys. Rev. Lett.* **90**, 257401.

Tran, C. Q., Chantler, C. T., Barnea, Z. & de Jonge, M. D. (2004). *Rev. Sci. Instrum.* **75**, 2943–2949.

Tran, C. Q., Chantler, C. T., Barnea, Z., Paterson, D. & Cookson, D. J. (2003*c*). *Phys. Rev A*, **67**, 42716.

Tran, C. Q., de Jonge, M. D., Barnea, Z. & Chantler, C. T. (2004). *J. Phys. B*, **37**, 3163–3176.

Troger, L., Arvanitis, D., Baberschke, K., Michaelis, H., Grimm, U. & Zschech, E. (1992). *Phys. Rev. B*, **46**, 3283–3289.

Tzeli, D. & Mavridis, A. (2010). *J. Chem. Phys.* **132**, 194312.

Vedrinskii, R. V., Bugaev, L. A. & Airapetian, V. M. (1991). *J. Phys. B*, **24**, 1967–1975.

Warburton, W. K., Hubbard, B. & Zhou, C. (1999). *X-ray Instrumentation Associates*, http://www.xia.com/Papers/Automat.pdf.

Xu, Z. F., Xie, Y., Feng, W. L. & Schaefer, H. F. (2003). *J. Phys. Chem.* A**107**, 2716.

Yamaguchi, Y., Ding, W., Sanderson, C. T., Borden, M. L., Morgan, M. J. & Kutal, C. (2007). *Coord. Chem. Rev.* **251**, 515–524.