Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 15 September 2009
Accepted 16 December 2009

Analysis of time-resolved X-ray scattering data from solution-state systems

Kristoffer Haldrup, Morten Christensen and Martin Meedom Nielsen*

Niels Bohr Institute, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark. Correspondence e-mail: martin.meedom.nielsen@nbi.ku.dk

As ultrafast time-resolved studies of liquid systems with the laser pump/X-ray scattering probe method have come of age over the past decade, several groups have developed methods for the analysis of such X-ray scattering data. The present article describes a method developed primarily with a focus on determining structural parameters in the excited states of medium-sized molecules (∼30 atoms) in solution. The general methodology is set in a maximum-likelihood framework and is introduced through the analysis of the photoactive platinum compound PtPOP, in particular the structure of its lowest triplet excited state ($^3A_{2u}$). Emphasis is put on structure determination in terms of model comparisons and on the information content of difference scattering signals as well as the related experimental variables. Several suggestions for improving the accuracy of these types of measurements are presented.

1. Introduction

Chemical and biological reactions in solution are fundamental to many important questions in science and technology, and provide the very foundations of life. While methods primarily based on optical spectroscopy have been used to study the time–energy landscape of such processes down to the femtosecond regime (*e.g.* Zewail, 2000; Kukura *et al.*, 2005; Zheng *et al.*, 2006), direct information on the accompanying structural changes has been sparse. However, recent advances in experiment and analysis methodology (Wulff *et al.*, 2002; Coppens *et al.*, 2005; Nozawa *et al.*, 2007; Ejdrup *et al.*, 2009) using pulsed X-rays to probe the structure of optically excited molecular states (Collet *et al.*, 2003; Techert & Zachariasse, 2003; Schotte *et al.*, 2003; Christensen *et al.*, 2009; van der Veen *et al.*, 2009) are opening up the field of time-resolved structural analysis of reactions in solution (Ihee *et al.*, 2005; Kong *et al.*, 2008; Cammarata *et al.*, 2008; Haldrup *et al.*, 2009) as discussed in two recent reviews (Ihee, 2009; Kim *et al.*, 2009*a*,*b*).

With respect to structural analysis, a very appealing aspect of X-ray scattering is the ability to calculate the scattering signal directly from the distribution of electron density which, in the cases relevant to liquid X-ray scattering, can be accurately described by the spatial arrangement of atoms and the atomic form factors. Compared with typical spectroscopic methods, there is no need to evaluate the detailed electronic potentials for ground or excited states of the molecules in question. This simplifies the interpretation of the measured results, and allows cases where such potentials may not be known *a priori* with sufficient accuracy.

X-ray scattering contains information about all the atoms in the sample that are within the X-ray beam path. While this ensures that the measured signal represents the ensemble properties of the sample, it also provides a significant challenge in extracting the scattering signal from a fraction of the sample, *e.g.* the excited-state solute, which, depending on the excitation fraction, may be only a small part of the signal from the solvent and unexcited solute.

Studying chemical reactions in solutions allows tracking of their structural implications in the environment relevant to realistic chemical and biological processes, without any restrictions imposed on molecular mobility by, for example, a crystalline environment. However, the lack of crystalline order in the sample means that the X-ray scattering signal only contains information about the orientation average of the molecules, and thus only reflects the pair correlation function of the electronic charge distribution in the sample. This effect reduces the signatures in the X-ray scattering signal of, for example, changing atomic distances and bond angles.

The two points mentioned above imply that, in order to obtain information on the excited-state structure of the solute, the changes to the pair correlation function must be extracted from a diffuse signal on a large background due to unexcited sample. In this respect, the technical challenges are similar to anomalous scattering (Haubold *et al.*, 1994). Further challenges arise as it is necessary to extract the signal corresponding to the excited-state solute from the signal corresponding to changes in the structure of the solvent resulting from the laser excitation of the solute. In practice, the scattering signal is composed of three terms: (1) the solute-only term arising from the internal structure of the solute, (2) the solute–solvent cross term (or the solvation cage term) reflecting the organization of solvent molecules around the solute, and (3) the solvent-only term reflecting the bulk solvent structure, all of which may also display time-resolved effects, in the case of (3) caused by temperature and density

changes due to heat imparted directly or indirectly by the laser pump pulse (Cammarata *et al.*, 2006).

Consequently, the utmost care must be taken in the data treatment to ensure that the subtle signals are retained despite noisy measurement conditions, and that quantitative measures for the accuracy of the corresponding structural parameters can be provided with statistical significance. The present paper demonstrates how these considerations have been implemented in a robust scheme for data treatment and fitting of key structural parameters of excited-state molecules. This will be discussed in terms of analysis of time-resolved X-ray scattering data obtained at ID09B at the European Synchrotron Radiation Facility (ESRF) during investigations of the excited-state structure of the photoactive diplatinum compound tetrakis-μ-pyrophosphitodiplatinate(II) (PtPOP) in aqueous solution (Rice & Gray, 1983; Novozhilova *et al.*, 2002; Christensen *et al.*, 2009; van der Veen *et al.*, 2009).

2. Data reduction

For the experiments described in this work, the acquired raw data are two-dimensional scattering patterns recorded on a position-sensitive detector. In addition to the desired information regarding the scattering properties of the sample, these images often also contain single- or few-pixel outliers due to, for example, stray radiation. The signal is further convoluted with effects from polarization of the X-ray beam, geometric effects from absorption in the sample and in the active detector medium, as well as intrinsic detector noise. These effects are taken into account by software developed for each particular beamline set-up. In the cases where the solution-state system is disordered and non-aligned, the scattering will be isotropic and the two-dimensional scattering pattern can be converted to a one-dimensional $S(2\theta)$ curve by azimuthal integration of the image [see, for example, the discussion given in Kim *et al.* (2009*b*)]. The starting point for the analysis described below is taken to be $S(2\theta)$ curves produced by software developed and available at beamline ID09B. No loss of generality of the present work is expected due to this, as all data processing up to this point is quite straightforward and mechanical in nature.

2.1. Difference signals, $\Delta S(2\theta)$

For pump–probe experiments, the changes to the sample system induced by the pump pulse are highlighted by taking the difference scattering signal, $\Delta S(2\theta) = S_{\rm pumpON} - S_{\rm pumpOFF}$, which contains all information regarding the structural changes (Bratos *et al.*, 2002; Henriksen & Møller, 2008). However, as the ratio $\Delta S/S$ is often significantly less than 1%, proper scaling and outlier removal are crucial issues and will be discussed before turning to how structural information is derived from $\Delta S(2\theta)$.

All scattering curves $S(2\theta)$ are scaled to the same 'absolute' value by requiring a suitable region (Kim *et al.*, 2009*b*; Cammarata *et al.*, 2006) in the high-angle part of the scattering curve to match the calculated coherent (Als-Nielsen & McMorrow, 2001) and incoherent (Hajdu, 1972) scattering from a single 'liquid unit cell' reflecting the stoichiometry of the solution. For a 12 m*M* solution of PtPOP in water, the single liquid unit cell consists of one PtPOP molecule (two Pt, eight P, 20 O, H atoms omitted), 4625 water molecules and four K atoms, the latter being the counter-ions in the PtPOP crystals dissolved in water for the experiments described by Christensen *et al.* (2009) and Haldrup *et al.* (2009). In the present case, scaling was performed in a 4° interval around an isosbestic point at $2\theta = 47°$, where the isosbestic point had been determined through a pre-analysis utilizing a 30–55° interval; in cases where no isosbestic points can be established, scaling intervals wide enough to contain a full oscillation of the difference signal are used. In this manner, the acquired set of scattering curves are all corrected for fluctuations due to changes in synchrotron bunch current and instrumental drift and can thus be directly compared. Furthermore, by carrying out all simulations (see below) for an identical liquid unit cell, a direct connection to excited-state fractions, temperature increase *etc.* is formed. Fig. 1(*a*) shows two recorded scattering

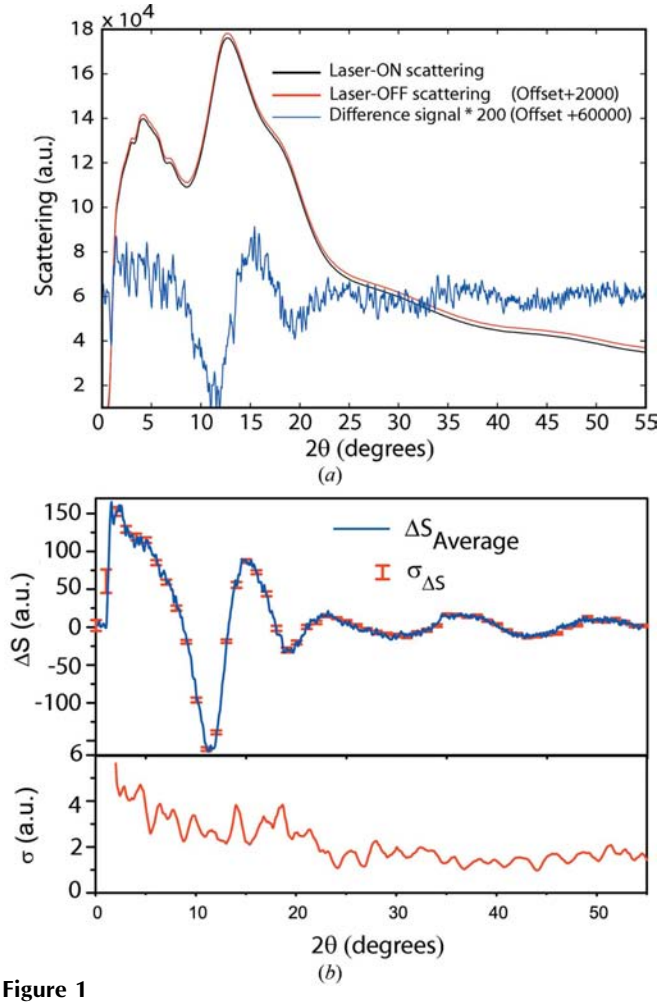

Figure 1
(*a*) Pair of On/Off scattering curves scaled to the calculated scattering from a 'liquid unit cell' (see text) and the corresponding difference signal $\Delta S(2\theta)$. The red scattering curve has been offset and the difference curve has been offset and scaled for clarity. (*b*) Average of 45 individual $\Delta S(2\theta)$ obtained in a single experiment, along with the estimated noise level $\sigma(2\theta)$.

curves for a 12 mM PtPOP (aq.) solution, scaled to the calculated scattering from a single unit cell.

From the scaled scattering curves $S(2\theta)_{\rm scaled}$, difference signals are formed by appropriate subtraction of laser-Off signals from laser-On signals. Commonly, an alternating sequence of laser-On and laser-Off scattering patterns is acquired such that each laser-On signal is bracketed by two laser-Off signals in order to minimize the effects of drift. In this case, the difference signal is constructed as $\Delta S(2\theta) = S(2\theta)_{\rm On} - [S(2\theta)_{\rm Off1} + S(2\theta)_{\rm Off2}]/2$, but more intense beams and faster detectors can allow for different and faster acquisition schemes. Fig. 1(a) shows an example of a $\Delta S(2\theta)$ curve based on a single On–Off repetition for a single time delay (100 ps).

In order to reduce noise, it is common to repeat the data-acquisition sequence between 10 and 50 times and it has proven crucial for our work to develop a robust method for rejecting outliers in such sets of $\Delta S(2\theta)$ curves. These outliers often arise due to unstable jet conditions and/or precipitates formed in the investigated solution-state systems and manifest themselves as having, for example, increased or decreased intensity at high or low scattering angles or through diffraction rings giving rise to sharp peaks in the difference curves. In the course of the data pre-analysis, these are removed as discussed in detail in previous work (see online supplementary material of Haldrup *et al.*, 2009), based on a point-by-point application of the unbiased Chauvenet criterion (Taylor, 1997). By this, the standard deviation $\sigma_{\Delta S_k(2\theta_i)}$ of the ensemble consisting of j $\Delta S(2\theta_i)$ curves (j = 10–50) is calculated at each $\Delta S(2\theta_i)$ point, omitting the kth of the j $\Delta S(2\theta)$ curves. Subsequently, for each $\Delta S_k(2\theta_i)$ on this chosen curve, one evaluates whether the probability of finding a particular value of $\Delta S_k(2\theta_i)$ is below 0.5, given the standard deviation $\sigma_{\Delta S_k(2\theta_i)}$ determined at this value of $2\theta_i$. If this is the case for a fraction of points exceeding a pre-set threshold, the difference curve is excluded from the later analysis. This provides a well defined and robust way of determining outlier curves, and using the unbiased version of the criterion [omitting the kth curve when calculating $\sigma_{\Delta S_k(2\theta_i)}$] provides robustness against individual outliers with large amplitudes. A constant cutoff threshold of 2.5% has been found to work very well in terms of identifying outlier curves, the fraction of which usually lie in the 5–10% range, depending on experimental conditions. A more rigorous treatment where the set of j values of the fraction of points on each curve failing the criterion is subjected to a similar analysis and rejection process is under development. Fig. 1(b) (top) illustrates an average curve based on 45 repetitions for the same time delay. The total data-acquisition time for 45 cycles and a single time delay with the present ID09B set-up (Frelon detector) is $\sim$400 s.

While the noise level for each point on an averaged $\Delta S(2\theta)$ curve can, in principle, be estimated from a knowledge of the detector counts on each pixel and applying the laws of error propagation to each subsequent data-analysis step, this is in practice an unwieldy process. In our previous work, an *ad hoc* procedure for estimating the noise level was implemented, where the standard deviation at each point, obtained from the ensemble of difference curves, was scaled to the 'local' standard deviation estimated from five-point intervals. Here, we propose to use a more robust and straightforward method, first introduced by Dent *et al.* (1991) in an EXAFS application and based on the observation (also discussed below) that the information-containing part of the signal has low-frequency oscillations on which high-frequency noise is superposed. By this method of error estimation, a low-order polynomial is fitted to the data points in a narrow interval around each data point and the standard deviation estimated from the set of residuals in this interval. It is found that in the present case an interval length of 20 points and a second-order polynomial fit are able to accurately capture the highest frequencies contained in $\Delta S(2\theta)$ (see below) and the resultant calculated $\sigma_{\Delta S}(2\theta)$ are shown for some data points in Fig. 1(b), top, and plotted for all data points in the lower graph of Fig. 1(b). The angle dependence of the noise reflects the complex interplay between scattered intensity, number of pixels and pixel efficiency, the latter related to variations in beam–phosphor interaction length as a function of angle.

Having obtained low-noise, averaged difference scattering curves that are unbiased by outliers, the final step in the pre-analysis procedure is to convert $\Delta S(2\theta)$ to $\Delta S(Q)$, $Q = (4\pi/\lambda)\sin\theta$. This is done by discretizing the X-ray (pink-beam, see below) spectrum $I(\lambda) = \sum_i I_i\lambda_i$, calculating $\Delta S(Q)_i$ for each discrete wavelength λ_i and forming the weighted average $\Delta S(Q) = \sum_i I_i \Delta S(Q)_i$.

3. Data analysis

Starting from $\Delta S(Q)$ for either one or several time delays, the objective of the analysis described here is to determine the structural changes in the liquid volume subjected to the pump pulse and subsequently probed by the X-ray beam. Contained within $\Delta S(Q)$ are the scattering contributions from every atom–atom pair in the system being affected by optical excitation of PtPOP. As mentioned in §1, these contributions can be divided into three groups according to either structural changes within the bulk solvent, in the oriented solvent shell around the excited molecule or in the excited PtPOP molecules. To a good first approximation (Christensen *et al.*, 2009), the shell term can be ignored for the systems discussed in this work, while the contribution from the changes in the scattering properties of the bulk solvent due to the impulse heating from the laser and fast internal conversion processes needs to be considered in some detail.

As discussed in detail in previous work (Cammarata *et al.*, 2006), the bulk solvent contribution $\Delta S_{\rm Solv}$ to the total difference scattering signal can be accurately described in terms of two solvent differentials, describing $\Delta S_{\rm Solv}$ as a function of the temperature rise ΔT due to impulse heating and the associated expansion causing a change in density $\Delta\rho$. The total hydrodynamic response is thus given through

$$\Delta S_{\rm Solv} = \Delta T \frac{\partial(\Delta S)}{\partial T}\Big|_\rho + \Delta\rho \frac{\partial(\Delta S)}{\partial \rho}\Big|_T. \qquad (1)$$

The solvent differentials $\frac{\partial(\Delta S)}{\partial T}|_\rho$ and $\frac{\partial(\Delta S)}{\partial \rho}|_T$ are determined in a separate experiment through direct laser excitation combined

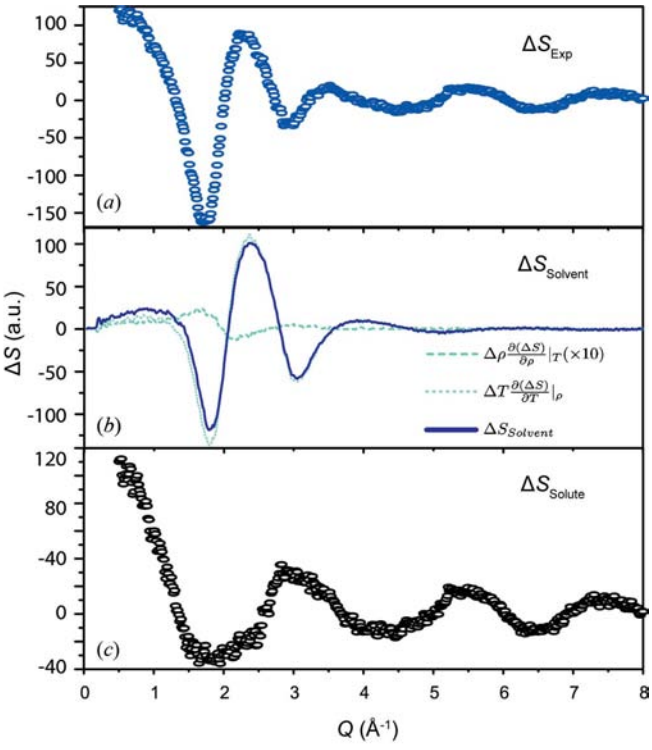

Figure 2
(*a*) $\Delta S(Q)$ for a 12 m*M* solution of PtPOP in water 100 ps after excitation. (*b*) Solvent contribution determined from the two hydrodynamic differentials (see text). As little or no solvent expansion is observed at sub-nanosecond timescales after excitation the contribution from $\Delta\rho\frac{\partial(\Delta S)}{\partial\rho}|_T$ has been multiplied by ten. (*c*) Solute-only difference signal obtained after subtracting the bulk solvent contribution.

with static scattering measurements. This method has been shown to provide a significantly more accurate fit to experimental data than, for example, molecular dynamics approaches (Cammarata *et al.*, 2006). The contributions to $\Delta S(Q)$ from bulk solvent changes and structural changes can furthermore be considered as independent, $\Delta S = \Delta S_{\rm Solute} + \Delta S_{\rm Solvent}$, and in the remaining part of this work we will focus the discussion on the solute contribution to $\Delta S(Q)$ and its interpretation in terms of structural changes in the investigated solute systems. Fig. 2 shows $\Delta S(Q)$ as obtained and with the contribution from bulk solvent changes subtracted.

3.1. Maximum-likelihood framework

Referring to Fig. 2, the basis for the structural analysis is a slowly varying curve with three to four slightly perturbed oscillations. As this constitutes a fairly information-poor starting point, *ab initio* derivation of the structural changes in a molecule as complicated as PtPOP is not a viable strategy. Instead, the analysis is carried out in terms of comparison of structural models, where prior information from, for example, crystallographic investigations and spectroscopy can be included in the analysis as has also been done in similar analysis of time-resolved EXAFS data (van der Veen *et al.*, 2009).

One way of quantitatively comparing structural models is to parameterize the structural variations within a maximum-likelihood framework, where a relative likelihood L is assigned to each combination of M parameters $(P_1, \ldots, P_M)$ proposed to explain a set of N data points $(x_{i,1\ldots N}, y_{i,1\ldots N})$. Assuming that the noise σ_i follows a Gaussian distribution, the relative likelihood $L(P_1, \ldots, P_M)$ that some set of parameters describes the data set can be calculated from a (reduced) χ^2 estimator (Press *et al.*, 1986),

$$L(P_1, \ldots, P_M) \propto \exp(-\chi^2),$$
$$\chi^2 = \sum_Q \frac{[S_{\rm Sim}(Q) - S_{\rm Data}(Q)]^2}{\sigma_Q^2}/(N - M - 1). \qquad (2)$$

In the present context, a ground-state structure could be identified based on crystal data and steady-state measurements on concentrated solutions of PtPOP (Christensen *et al.*, 2009). The target for the investigation was the excited-state structure, and based on suggestions in the literature (Novozhilova *et al.*, 2002; Coppens *et al.*, 2005; Yasuda *et al.*, 2004) a set of putative excited-state structures were derived from the established ground-state structure and parameterized in terms of the distance $d_{\rm PtPt}$ between the Pt atoms and the perpendicular distance $d_{\rm PpPp}$ between the square-planar phosphorous units. Solute-only difference signals $\Delta S_{\rm Solute}(Q)$ were calculated by subtracting the simulated scattering from the ground-state structure from simulated scattering from an excited-state structure, both calculated from the orientation-averaged Debye expression (Als-Nielsen & McMorrow, 2001; Christensen *et al.*, 2009; Kim *et al.*, 2009*b*). Including the fraction of excited-state molecules α (and the two hydrodynamic variables ΔT and $\Delta\rho$) as free parameters, the objective of the structural analysis thus becomes to determine the most likely values of $d_{\rm PtPt}$ and $d_{\rm PpPp}$ as well as the associated uncertainties from calculations of χ^2 for each combination of $\{d_{\rm PtPt}, d_{\rm PpPp}, \alpha, \Delta T, \Delta\rho\}$.

3.2. Correlations and experimental fixing of variables

The output from the maximum-likelihood global fitting introduced above is a five-dimensional likelihood distribution. In the present case, there is a single, well defined maximum corresponding to one five-parameter combination resulting in the best fit of the model to the data. Fig. 3(*a*) shows the very good agreement between this best-fit model and the data; the bulk solvent contribution has been subtracted from both the data and the model for clarity.

The determination of experimental uncertainties is illustrated graphically in Fig. 3(*b*), where the projection of the full-likelihood distribution onto a plane spanned by two fit parameters, α and $d_{\rm PtPt}$, illustrates the strong correlation between these two parameters. For an unbiased determination of the experimental uncertainty associated with each of the five fit parameters, the full likelihood distribution must be projected onto each of the five axes as shown in the two-dimensional case in Fig. 3(*b*). From this procedure, the most likely value of each parameter can be determined and the uncertainty estimated as the interval centred on the most

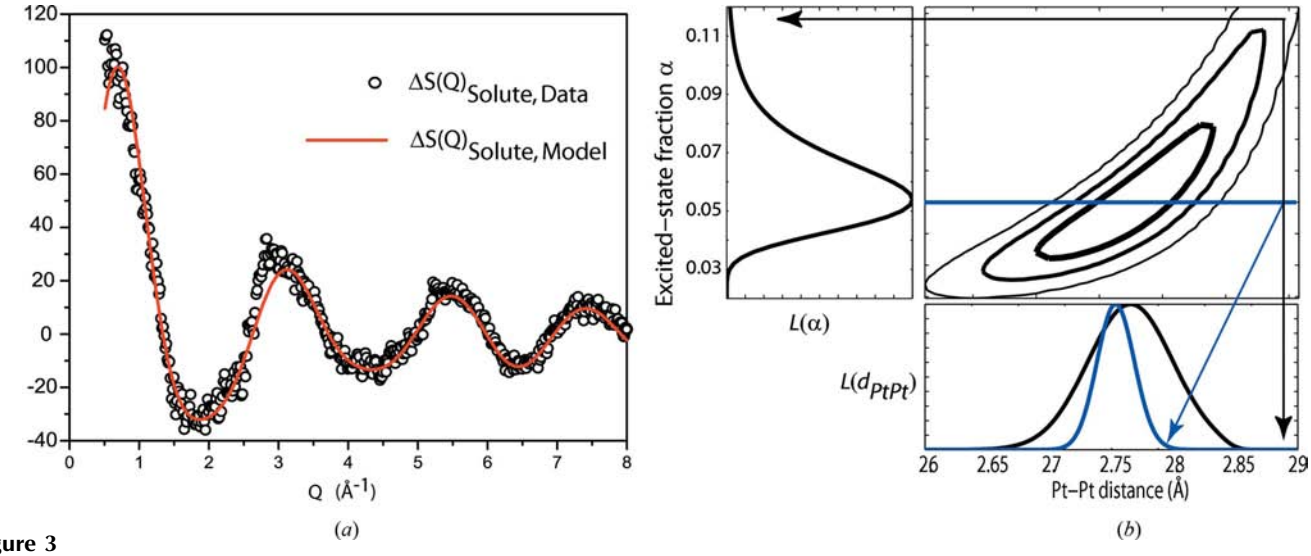

Figure 3
(*a*) Data (black circles) and fit (red curve), both shown without the solvent contribution. (*b*) Projection of the five-dimensional likelihood distribution onto the α–d_{PtPt} plane. Black arrows indicate the projection of $L(\alpha, d_{PtPt})$ onto the two single-parameter axes and the corresponding likelihood distributions $L(\alpha)$ and $L(d_{PtPt})$. The blue line+arrow highlight a section through the two-dimensional distribution, corresponding to independent determination of the excitation fraction to 5% and the corresponding, much narrower, likelihood distribution $L(d_{PtPt})_{\alpha=5\%}$.

likely value and enclosing 68% of the likelihood distribution, in analogy with the σ determined for Gaussian distributions. The contours on the two-dimensional projection contain 68, 95 and 99.7% of the likelihood distribution.

In terms of quantifying the experimental accuracy in the determination of structural parameters, we find that applying a measure of the uncertainty that has a clear and well defined connection to other statistical measures has distinct advantages in terms of interpretation. The uncertainty estimate based on the 68% interval is such a measure through the direct connection to the σ obtained for Gaussian distributions. It is a considerably more conservative estimate than, for example, the alternative confidence-level approach applied in the otherwise analogous data analysis in two recent EXAFS papers on the structure of the excited states of PtPOP (van der Veen *et al.*, 2009) and $Fe^{II}(bpy)_3$ (Gawelda *et al.*, 2009). Using the method of these EXAFS studies, variations in χ^2 as one moves away from the minimum, we arrive at an accuracy of $\sigma_{d_{PtPt}} = 0.012$ Å on d_{PtPt}, compared with $\sigma_{d_{PtPt}} = 0.05$ Å with the likelihood-based method presented above.

A further advantage of the global-likelihood-mapping approach compared with more standard search-based methods for χ^2 minimization is robustness towards local minima as well as access to the global covariance matrix. In this respect, an interesting observation from Fig. 3 is that, in cases where strong experimental correlation between two parameters is observed, very significant reductions in parameter uncertainties can be achieved if one or other of such variables can be well determined by external means. In the present case, such strong correlation is observed only between α and d_{PtPt}, and if, for example, the excitation fraction could have been determined during the experiment, and assuming it was found to be equal to the most likely value determined in the preceding analysis, the quite broad d_{PtPt} distribution given by the thicker black line would be replaced by the much narrower distribution depicted by the thin blue line and obtained by taking a section through $L(\alpha, d_{PtPt})$ rather than the full projection. Thus, a global analysis approach such as the one presented in this and related works may yield important clues to improvements in experimental techniques.

3.3. Information content in difference signals

From the results shown above and as presented by several different groups in the literature (Ihee, 2009; Kim *et al.*, 2009*b*), it is clearly possible to separate solvent and solute contributions and to extract structural information about the solute molecules by using time-resolved X-ray scattering on liquid systems. However, to the best of the authors' knowledge, it remains an essentially open question exactly how much structural information can be deduced from the slowly varying difference signals that form the basis of the structural analysis. Also, it is at present not clear what experimental avenues can most fruitfully be pursued in order to minimize the uncertainty of the various structural parameters investigated.

Regarding the information content of the data in time-resolved solution-state scattering experiments, other workers (*e.g.* see online supplementary information of Cammarata *et al.*, 2008) have employed theoretical approaches based originally on the basic information-theory work of C. E. Shannon (Weaver & Shannon, 1949; Brillouin, 1962), but later adapted for small-angle X-ray scattering (Moore, 1980; Taupin & Luzzati, 1982; Svergun & Stuhrmann, 1991; Svergun *et al.*, 1996) and for analysis of EXAFS data (Lee *et al.*, 1981; Stern, 1993). Within this approach, the information content of a set of data points was estimated by considering the Fourier transform resolution of the data (given from the Q interval investigated, $\Delta Q = Q_{Max} - Q_{Min}$) and the maximum particle size considered in the analysis, D_{Max}, which corresponds to a

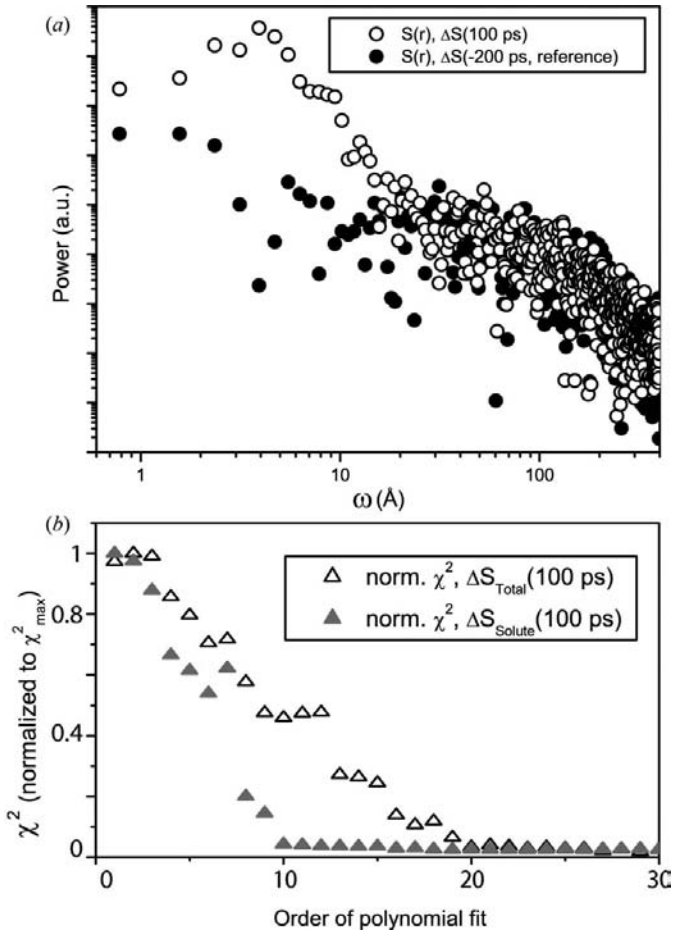

Figure 4
(*a*) Power spectra $S(\omega) = |\mathcal{F}(\Delta S)|^2$ for the 100 ps data and for a reference signal, where the pump pulse arrives after the probe pulse. (*b*) Normalized χ^2 as a function of the maximum polynomial order used to fit ΔS. Also shown is the result when the bulk solvent contribution to ΔS has been subtracted.

filter window width in real space. This led to the number of 'relevant independent points'/'Shannon channels' in the data to be estimated as $N = 2\Delta Q D_{\rm Max}/\pi$ [1]. Taking the experimental data on the PtPOP system presented above as an example, the maximum linear dimension of the PtPOP entity is of the order of 6 Å and data have been collected in the Q = 0.5–8 Å region; thus the number of Shannon channels becomes ∼28. This represents the theoretical upper limit on the number of parameters that can be determined in the experiments considered here.

To investigate further the number of independent parameters that can in principle be determined from $\Delta S(Q)$, Fig. 4(*a*) shows the power spectrum $S(\omega) = |\mathcal{F}(\Delta S)|^2$ calculated from the Fourier transform of ΔS (100 ps) and ΔS (−200 ps), with the latter serving to illustrate the 'intrinsic noise level' as the laser pump pulse in this case arrives after the X-ray probe pulse. Comparing the two signals, a merging is observed at ω ∼15–20 Å, above which the two power spectra are both completely dominated by noise. The position of the merging corresponds to only the first 20 Fourier components having significant amplitudes, *i.e.* the signal can be well described with 20 independent parameters. This can be further checked for consistency by a back-transform (not shown) exclusively based on the 20 first Fourier components, and we find that this acts as an efficient low-pass filter accurately reproducing the original signal with the high-frequency noise components removed.

The data presented in Fig. 4(*b*) further support this estimate of information content in the data by showing the (normalized) χ^2 of polynomial fits to the data as a function of the order of polynomial used. Again, ∼20 independent parameters are needed to accurately represent ΔS. If the bulk solvent contribution to the total difference signal is removed, we find that about ten independent parameters are required to represent the data, indicating that, for this particular experiment, a maximum of ten independent parameters regarding the structural change of the solute can in principle be determined. The rational choice of input structures discussed in the preceding section corresponds to fixing a number of these 'free' parameters. From the quality of the model fit to the data, we estimate that the actual number of free parameters is not significantly larger than the five parameters varied in the structural analysis and we also note that the ground-state starting structure appears to be well chosen.

4. Effects of X-ray bandwidth and signal strength

The world-leading beamline in terms of published time-resolved X-ray scattering studies on liquid systems is ID09B at the ESRF. The majority of these published results have been based on data taken with the pink-beam option, where the first harmonic of the u17 undulator is used to provide a very intense beam on the sample (10^9 photons pulse^{-1}) (Wulff *et al.*, 2002; Cammarata *et al.*, 2009). Fig. 5(*a*) shows the triangular intensity distribution as a function of energy and, defining the bandwidth of the spectrum as BW = FWHM/$E_{\rm Max}$, it has a bandwidth of 4.6% centred at 17.95 keV. However, it has a very long low-energy tail and Fig. 5(*a*) shows the smearing associated with this tail by comparing difference signals from PtPOP, calculated using either the full spectrum, a monochromatic spectrum centred at 17.8 keV or a quasi-monochromatic spectrum (BW = 3.4%) obtained by placing a Ru/B_4C multilayer in the focused X-ray beam after the high-speed chopper.

Fig. 5(*b*) shows how this smearing also increases the uncertainty of the determination of structural parameters. A simulated difference signal was constructed by calculating the scattering from the PtPOP/PtPOP* structures determined previously (Christensen *et al.*, 2009), using the three spectral distributions discussed above and in each case assuming 5% photo-excitation and ignoring the solvent contribution to the difference signal. Noise with the same $\sigma(Q)$ distribution as determined in the experiment discussed above was then added to the simulated signal, and the result subjected to the same analysis procedure as presented above.

[1] It was later shown by A. Stern (Stern, 1993) that the correct expression should include an additional two channels in EXAFS analysis.

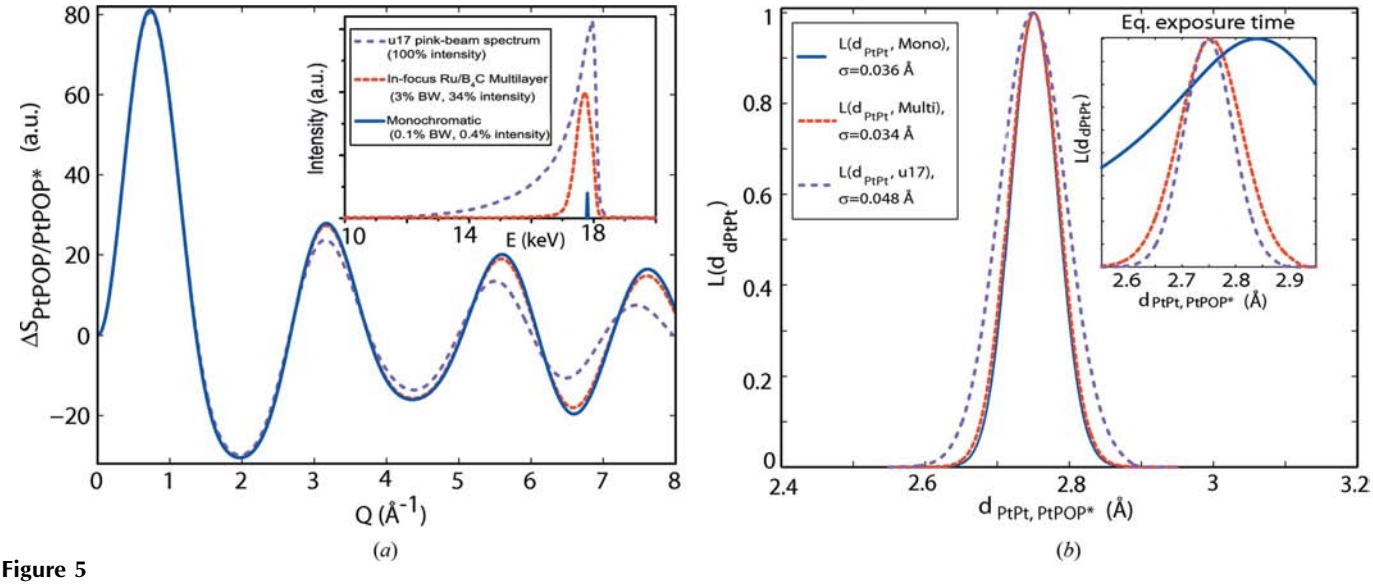

Figure 5
(*a*) u17, multilayer and monochromatic spectra and corresponding simulations of difference signals. (*b*) Likelihood distributions for different spectral distributions, assuming identical S/N levels based on the 45-repetition data set. The inset shows the likelihood distributions when the simulated data represent identical exposure times, rather than noise levels.

As the three likelihood distributions $L(d_{\rm PtPt})$ in Fig. 5(*b*) show, there is a significant decrease (25%) in the uncertainty on $d_{\rm PtPt}$ when a monochromatic X-ray beam is used as probe, compared with a polychromatic pink-beam. However, this increase in accuracy comes at a cost of beam intensity, as the monochromatic beam at ID09B is 250 times weaker than the polychromatic beam. For this reason, the possibility of using a multilayer mirror with 34% transmission placed in the focused beam becomes a very interesting option, as this yields virtually the same increase in accuracy as the monochromatic beam at much less cost in intensity. The inset in Fig. 5(*b*) highlights this by showing the three likelihood distributions when the signal-to-noise (S/N) level is adjusted to represent equal exposure times with the ∼400 s pink-beam experiment as reference.

These simulated results are in good qualitative agreement with the trends observed by Ichiyanagi *et al.* (2009) in their experimental study of the effect of multilayer mirrors in time-resolved liquid scattering experiments. The three times smaller spectral width of the u17 undulator installed at ID09B compared with the NW14A beamline makes the smearing of the difference signal significantly less pronounced, but, as Fig. 5(*b*) shows, the polychromaticity still increases the uncertainty of the determination of structural parameters compared with the narrower multilayer spectra.

As the final experimental parameter to be investigated in this exploratory part of the work, Fig. 6 shows that, at small excitation fractions, even modest increases in this parameter lead to significant gains in terms of decreased parameter uncertainty. Recent experiments at ID09B using a high-intensity microsecond-long excitation pulse rather than the 2–3 ps stretched femtosecond pulse often used for this type of study have qualitatively confirmed a very marked increase in the S/N ratio. Thus, for systems with little or no dynamics on the single-microsecond timescale, this appears an interesting option to explore.

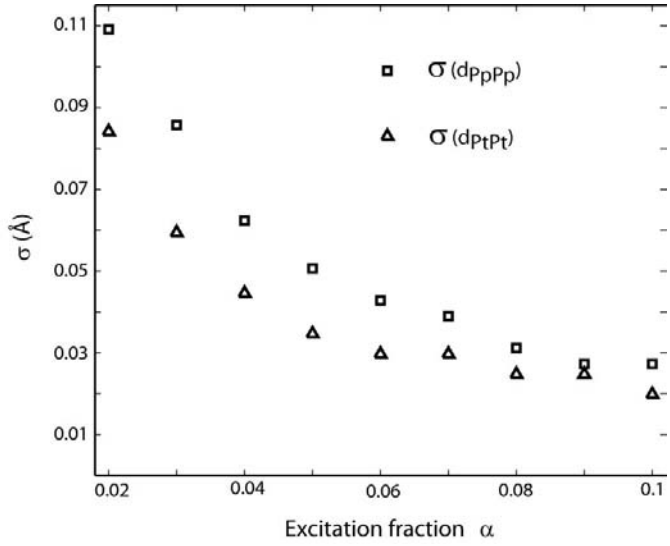

Figure 6
Uncertainty σ on $d_{\rm PtPt}$ (triangles) and $d_{\rm PpPp}$ (squares) as a function of the excitation fraction α, keeping the noise level constant at the level determined experimentally. A factor of two in terms of uncertainty can be gained by going from $\alpha = 0.04$ to $\alpha = 0.08$.

5. Conclusions and outlook

A general analysis methodology has been presented for the elucidation of excited-state structures for medium-sized molecules such as PtPOP and using time-resolved X-ray scattering from solution-state systems as the experimental tool. As is always the case in this type of study, careful averaging of many repeated experiments and robust outlier detection and rejection are crucial in order to achieve reliable, low-noise difference signals $\Delta S(Q)$. Based on such signals, it has been shown how structural analysis within a maximum-likelihood framework can yield information about structural parameters with high accuracy. In the present work we find

that the two key structural parameters d_{PtPt} and d_{PpPp} describing the structure of the $^3A_{2u}$ excited state of the PtPOP molecule could be determined to be $d_{PtPt} = 2.76 \pm 0.05$ Å and $d_{PpPp} = 2.87 \pm 0.09$ Å. These values agree with our previously reported results within the uncertainty intervals, and we attribute the slight differences in most probable values to the improved estimate of the noise, $\sigma(Q)$, highlighting the importance of determining this in an accurate and reproducible manner.

Within the context of this approach to structure determination, it appears reasonable to ask whether more parameters could have been included in the analysis, such as for instance a rotation of the Pt—Pt axis as has been suggested by some authors (Kim *et al.*, 2002). Through an analysis of the power spectra of $\Delta S(Q)$ as well as an investigation of what order of polynomial is required to fit the $\Delta S(Q)_{total}$ and $\Delta S(Q)_{solute}$, we find that approximately 20 independent parameters can be determined from the total signal. This is within the upper limit, 28 parameters, set by fundamental analysis of the theoretical information content of the data. From high-order polynomial fits to the entire data set, the 20 free parameters appear to be partitioned equally between bulk solvent and solute contributions to $\Delta S(Q)$, as ten parameters are required to describe the solute-only part of the data (Fig. 4*b*). Three of these parameters are explicitly incorporated in the maximum-likelihood structural analysis, d_{PtPt}, d_{PpPp} and the excitation fraction α. From the quality of the fits, the data appear to be very well reproduced by the model using these parameters, suggesting that only limited amounts of further information may be extracted under these particular experimental conditions although recent EXAFS results indicate that modification of the model to include changes in the ligand structure of the excited-state PtPOP* may be worthwhile (van der Veen *et al.*, 2009).

From the theoretical analysis of the information content of the data, extending the available Q range in future studies may allow for richer structure determination in systems with more complicated structural changes. This would necessitate more advanced algorithms for the structural analysis such as simulated annealing or genetic algorithms, however, as the current global approach scales poorly with the number of varied parameters.

Focusing on improving future experiments of the kind presented here, simulations of experimental data obtained with either a pink-beam undulator spectrum, a quasi-monochromatic multilayer spectrum or a monochromatic spectrum strongly favour the implementation of a multilayer mirror in the experimental set-up as will take place at ID09B in the near future (M. Wulff & L. Guerin, personal communication). Reductions of 25% or more in the uncertainties related to parameter determination appear quite feasible. Analysis of the correlation between the parameters varied in the present study also indicates that *in situ* monitoring of excitation fractions (when possible) would make even more significant gains in accuracy possible. Other immediate gains in terms of accuracy may be achieved by increasing the excited-state population in the investigated sample volume, as this significantly decreases the uncertainties of the structural parameters (see Fig. 6).

Looking further ahead, the analysis approach described in the current work appears well suited to take advantage of the coming pulsed X-ray sources with pulse lengths well down into the femtosecond regime, although a few significant modifications are needed as the assumption of complete rotational disorder of the excited-state population becomes invalid. Implementing this, structural elucidation of singlet as well as triplet excited states and direct investigations of bond breakage in medium-sized molecules in solution will be interesting targets for investigation.

We wish to express our great appreciation to Marco Cammarata and Michael Wulff from the ID09B beamline at ESRF, Grenoble, for fruitful discussions and expert assistance with the experiments. This work was supported by the Danish National Research Foundation's Centre for Molecular Movies and DANSCATT.

References

Als-Nielsen, J. & McMorrow, D. (2001). *Elements of Modern X-ray Physics*. New York: John Wiley and Sons.

Bratos, S., Mirloup, F., Vuilleumier, R. & Wulff, M. (2002). *J. Chem. Phys.* **116**, 10615.

Brillouin, L. (1962). *Science and Information Theory*. New York: Dover.

Cammarata, M., Eybert, L., Ewald, F., Reichenbach, W., Wulff, M., Anfinrud, P., Schotte, F., Plech, A., Kong, Q., Lorenc, M., Lindenau, B., Raebiger, J. & Polachowski, S. (2009). *Rev. Sci. Instrum.* **80**, 015101.

Cammarata, M., Levantino, M., Schotte, F., Anfinrud, P. A., Ewald, F., Choi, J., Cupane, A., Wulff, M. & Ihee, H. (2008). *Nat. Methods*, **5**, 881–886.

Cammarata, M., Lorenc, M., Kim, T., Lee, J. H., Kong, Q. Y., Pontecorvo, E., Russo, M. L., Schiro, G., Cupane, A., Wulff, M. & Ihee, H. (2006). *J. Chem. Phys.* **124**, 124504.

Christensen, M., Haldrup, K., Bechgaard, K., Feidenhans'l, R., Kong, Q., Cammarata, M., Russo, M. L., Wulff, M., Harrit, N. & Nielsen, M. M. (2009). *J. Am. Chem. Soc.* **131**, 502–508.

Collet, E., Lemee-Cailleau, M.-H., Cointe, M. B.-L., Cailleau, H., Wulff, M., Luty, T., Koshihara, S.-Y., Mathias Meyer, L. T., Rabiller, P. & Techert, S. (2003). *Science*, **300**, 612–615.

Coppens, P., Vorontsov, I. I., Graber, T., Gembicky, M. & Kovalevsky, A. Y. (2005). *Acta Cryst.* A**61**, 162–172.

Dent, A., Stephenson, P. & Greaves, G. (1991). *Rev. Sci. Instrum.* **63**, 856–858.

Ejdrup, T., Lemke, H. T., Haldrup, K., Nielsen, T. N., Arms, D. A., Walko, D. A., Miceli, A., Landahl, E. C., Dufresne, E. M. & Nielsen, M. M. (2009). *J. Synchrotron Rad.* **16**, 387–390.

Gawelda, W., Pham, V.-T., van der Veen, R. M., Grolimund, D., Abela, R., Chergui, M. & Bressler, C. (2009). *J. Chem. Phys.* **130**, 124520.

Hajdu, F. (1972). *Acta Cryst.* A**28**, 250–252.

Haldrup, K., Christensen, M., Cammarata, M., Kong, Q., Wulff, M., Mariager, S. O., Bechgaard, K., Feidenhans'l, R., Harrit, N. & Nielsen, M. M. (2009). *Angew. Chem. Int. Ed.* **48**, 4180–4184.

Haubold, H.-G., Gebhardt, R., Buth, G. & Goerigk, G. (1994). *Structural Characterization of Compositional and Density Inhomogeneities in Resonant Anomalous X-ray Scattering, Theory and Applications*, edited by G. Materlik, C. F. Sparks & K. Fischer. Amsterdam: Elsevier Science.

Henriksen, N. E. & Møller, K. B. (2008). *J. Phys. Chem. B*, **112**, 558–567.
Ichiyanagi, K., Sato, T., Nozawa, S., Kim, K. H., Lee, J. H., Choi, J., Tomita, A., Ichikawa, H., Adachi, S., Ihee, H. & Koshihara, S. (2009). *J. Synchrotron Rad.* **16**, 391–394.
Ihee, H. (2009). *Acc. Chem. Res.* **42**, 356–366.
Ihee, H., Lorenc, M., Kim, T. K., Kong, Q. Y., Cammarata, M., Lee, J. H., Bratos, S. & Wulff, M. (2005). *Science*, **309**, 1223–1227.
Kim, C. D., Pillet, S., Wu, G., Fullagar, W. K. & Coppens, P. (2002). *Acta Cryst.* A**58**, 133–137.
Kim, T. K., Lee, J. H., Wulff, M., Kong, Q. & Ihee, H. (2009*a*). *ChemPhysChem*, **10**, 3140.
Kim, T. K., Lee, J. H., Wulff, M., Kong, Q. & Ihee, H. (2009*b*). *ChemPhysChem*, **10**, 1958–1980.
Kong, Q., Lee, J. H., Plech, A., Wulff, M., Ihee, H. & Koch, M. H. J. (2008). *Angew. Chem. Int. Ed.* **47**, 5550–5553.
Kukura, P., McCamant, D. W., Yoon, S., Anf, R. A. & Mathies, D. B. W. (2005). *Science*, **310**, 1006–1009.
Lee, P. A., Citrin, P. H., Eisenberger, P. & Kincaid, B. M. (1981). *Rev. Mod. Phys.* **53**, 769–806.
Moore, P. B. (1980). *J. Appl. Cryst.* **13**, 168–175.
Novozhilova, I., Volkov, A. & Coppens, P. (2002). *J. Am. Chem. Soc.* **125**, 1079–1087.
Nozawa, S. *et al.* (2007). *J. Synchrotron Rad.* **14**, 313–319.
Press, W. H., Flannery, B. P., Teukolsky, T. A. & Vetterling, W. T. (1986). *Numerical Recipes – The Art of Scientific Computing.* Cambridge University Press.
Rice, S. F. & Gray, H. B. (1983). *J. Am. Chem. Soc.* **105**, 4571–4575.
Schotte, F., Lim, M., Jackson, T. A., Smirnov, A. V., Soman, J., Olson, J. S., Phillips, G. N. Jr, Wulff, M. & Anfinrud, P. A. (2003). *Science*, **300**, 1944–1947.
Stern, E. A. (1993). *Phys. Rev. B*, **48**, 9825–9827.
Svergun, D. I. & Stuhrmann, H. B. (1991). *Acta Cryst.* A**47**, 736–744.
Svergun, D. I., Volkov, V. V., Kozin, M. B. & Stuhrmann, H. B. (1996). *Acta Cryst.* A**52**, 419–426.
Taupin, D. & Luzzati, V. (1982). *J. Appl. Cryst.* **15**, 289–300.
Taylor, J. R. (1997). *Introduction to Error Analysis.* Sausalito: University Science Books.
Techert, S. & Zachariasse, K. A. (2003). *J. Am. Chem. Soc.* **126**, 5593–5600.
Veen, R. M. van der, Milne, C. J., Nahhas, A. E., Lima, F. A., Pham, V.-T., Best, J., Weinstein, J. A., Borca, C. N., Abela, R., Bressler, C. & Chergui, M. (2009). *Angew. Chem. Int. Ed.* **48**, 2711–2714.
Weaver, W. & Shannon, C. E. (1949). *The Mathematical Theory of Communication.* Urbana: University of Illinois Press.
Wulff, M., Plech, A., Eybert, L., Randler, R., Schotte, F. & Anfinrud, P. (2002). *Faraday Discuss.* **122**, 13–26.
Yasuda, N., Uekusa, H. & Ohashi, Y. (2004). *Bull. Chem. Soc. Jpn*, **77**, 933–944.
Zewail, A. (2000). *J. Phys. Chem. A*, **104**, 5660–5694.
Zheng, J., Kwak, K., Xie, J. & Fayer, M. D. (2006). *Science*, **313**, 1951–1955.