Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 27 June 2004
Accepted 24 January 2005

Diffraction and imaging from a beam of laser-aligned proteins: resolution limits

J. C. H. Spence,[a]* K. Schmidt,[a] J. S. Wu,[a] G. Hembree,[a] U. Weierstall,[a] B. Doak[a] and P. Fromme[b]

[a]Department of Physics and Astronomy, Arizona State University, Tempe, AZ 85287-1504, USA, and [b]Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287-1504, USA. Correspondence e-mail: spence@asu.edu

The effect of the limited alignment of hydrated molecules is considered in a laser-aligned molecular beam, on diffraction patterns taken from the beam. Simulated patterns for a protein beam are inverted using the Fienup–Gerchberg–Saxton phasing algorithm, and the effect of limited alignment on the resolution of the resulting potential maps is studied. For a typical protein molecule (lysozyme) with anisotropic polarizability, it is found that up to 1 kW of continuous-wave near-infrared laser power (depending on dielectric constant), together with cooling to liquid-nitrogen temperatures, may be needed to produce sufficiently accurate alignment for direct observation of the secondary structure of proteins in the reconstructed potential or charge-density map. For a typical virus (TMV), a 50 W continuous-wave laser is adequate for subnanometre resolution at room temperature. The dependence of resolution on laser power, temperature, molecular size, shape and dielectric constant is analyzed.

1. Introduction

In a recent publication (Spence & Doak, 2004), the use of diffraction patterns from aligned molecular beams has been proposed as an approach to the structure determination of proteins that are difficult to crystallize. As in previous work on small organic molecules (Stapelfeldt & Seideman, 2003) such as amino acids (Lindinger *et al.*, 1999), polarized laser light is used to align the molecules in the beam. By cooling a beam of hydrated proteins to low temperature, sufficient alignment for diffraction purposes might be possible. We therefore suggest an arrangement consisting of continuous, orthogonal, intersecting electron (or X-ray), molecular and laser beams, as shown in Fig. 1. Surplus vitreous ice surrounding the proteins is allowed to sublimate along the molecular beam flight path. The limited coherence of the electron beam ensures that no interference occurs between the wavefields scattered by different molecules within the electron beam. Thus the diffraction pattern is a sum of the intensities of the identical patterns from the many hydrated molecules within the beam at any one time. The two-dimensional electron diffraction pattern accumulates continuously at the detector for a fixed laser and molecular orientation before being read out. Repeating this process for many new orientations (each with a new laser polarization direction) would allow a tomographic reconstruction of the molecular charge density or electrostatic potential. Since the oriented molecules traversing the 10 μm beam (in 200 ns) are constantly replenished, radiation damage at atomic resolution can be shown to be negligible. In steady state, the oriented molecular arrangement within the beam acts like a stationary nanocrystal in determining the exposure time. (This time is estimated to be about 80 s per orientation for electron diffraction. By using a 'shower-head' array of nozzles to compensate for the lower scattering cross section of X-rays, a similar exposure time could be obtained using a medium-energy synchrotron undulator.)

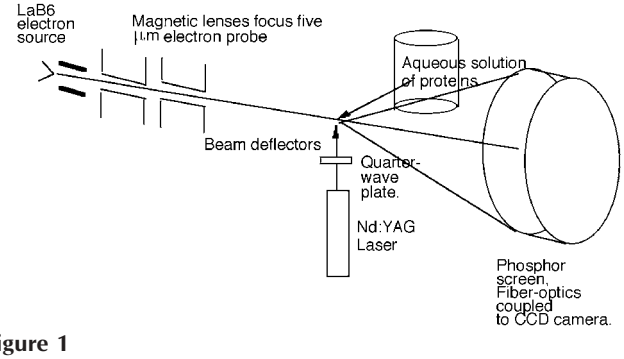

Figure 1
Electron diffraction camera for single-molecule electron diffraction from a liquid jet. A continuous electron beam is used. Hydrated organic molecules emerge from a nozzle into vacuum where droplets form and cool by evaporation to form a beam of doped iceballs traversing the electron beam. A continuous-wave fiber laser is used to align them. The laser polarization is rotated after each data readout for a new molecular-beam orientation.

This arrangement would result in great simplification and reduction of data-collection times over similar femtosecond pulsed X-ray schemes that have been proposed (Neutze *et al.*, 2000). These are based on the finding from simulations that the onset of radiation damage at atomic resolution occurs about 10–20 fs after significant diffraction has occurred. The method aims to collect patterns from successive randomly oriented proteins in a stream crossing an X-ray beam in synchrony with the X-ray pulses. The need for area-detector read-out (a slow process) after recording the diffraction pattern from each molecule in that method is eliminated in our method, as is the need to determine the orientation between successive patterns from randomly oriented molecules. By comparison with existing cryo-electron-microscopy methods, this approach could greatly improve throughput while avoiding tedious sample preparation, and may be able to handle smaller molecules at higher resolution. Since both methods deal with proteins in vitreous ice, they will be affected equally by conformational variations, which might be minimized by attaching a small molecule to the proteins. By comparison with NMR, the new method, if successful, would offer much faster throughput since the molecular beam is fed from a solution of proteins in room-temperature water to which cryoprotectant is added to encourage vitrification. This solution enters a vacuum chamber through a μm-sized aperture, where necking, droplet formation and freezing occurs to produce a stream of vitreous iceballs containing proteins, as in similar previous work (Bartell & Huang, 1994; Charvat *et al.*, 2002).

In this paper, we investigate in more detail the degree of alignment possible in laser-aligned molecular beams at low temperature for the case of large molecules, and determine the limits thus imposed on the resolution of reconstructed charge-density maps. The phase problem for the continuous distribution of scattering from individual molecules is solved using recently developed iterative methods, which have now been applied to both X-ray (Miao *et al.*, 1999; Marchesini *et al.*, 2003) and electron diffraction data (Weierstall *et al.*, 2001), most recently at atomic resolution (Zuo *et al.*, 2003).

The aim of this work is to determine the experimental conditions (laser power, temperature *etc.*) that would allow the observation of the secondary structure of proteins by applying the iterative inversion algorithm to simulated diffraction patterns with various degrees of misalignment.

2. Laser alignment

Detailed quantum-mechanical treatments of the laser-alignment problem for small-molecule beams have appeared in the literature (Friedrich & Herschbach, 1995; Stapelfeldt & Seideman, 2003), including the case where elliptical polarization is used to provide 'three-dimensional' orientation (see later discussion). Here we first describe the simplest classical model of one-dimensional alignment for a large molecule with axial symmetry, which exposes the dependence of the effect on experimental parameters. This is then used to determine the resolution limits for a given degree of alignment. Detailed calculations of the anisotropic, frequency-dependent, electronic molecular polarizability for small molecules, on which the alignment effect depends, have been reported elsewhere (Fraschini *et al.*, 1996). (We assume, as in the small-molecule work, that resonances due to nuclear vibrations, as studied in infrared absorption spectroscopy, are avoided.) In the adiabatic approximation, and away from such resonances, the orientation energy of a molecule within a polarized light field due to a difference in induced electronic polarizability $\Delta\alpha = \alpha_{\zeta\zeta} - \alpha_{\xi\xi}$ along principal axes is (Seideman, 2001)

$$H = -\tfrac{1}{4}\sum_{\rho,\rho'} E_\rho \alpha_{\rho,\rho'} E'_\rho = -(2\pi I/c)\Delta\alpha\cos^2(\theta), \qquad (1)$$

where I is the laser power density (intensity), c the speed of light and θ the polar Euler angle between the symmetry axis of the molecule and the electric field vector $\mathbf{E}$ of the laser, here taken to be plane-polarized. Only the frequency-dependent electronic contribution to $\Delta\alpha$, and the RMS value of the electric field, is effective – the time average of the linear (Stark) term which depends on the permanent dipole moment is zero. Setting this potential energy H equal to thermal energy $(1/2)kT$ by equipartition, we obtain, in the harmonic approximation, a RMS angular variation of

$$\langle \Delta\vartheta^2 \rangle = T/(3 \times 10^{-8} I \Delta\alpha), \qquad (2)$$

where I is the laser power density (intensity) in W cm^{-2} and $\Delta\alpha$ is given in nm^3. The same result is obtained in this limit by evaluating the expectation value of the thermal average of $\cos^2(\theta)$, with Boltzman-factor weighting, as given previously for quantum systems (Seideman, 2001). Since $\Delta\alpha$ is proportional to molecular volume, the mean misorientation angle decreases steeply as the inverse cube of the molecular dimension, inversely as the degree of anisotropy and laser intensity, and is proportional to the temperature.

Accurate estimates of the anisotropic frequency-dependent electronic polarizability tensor α for proteins are difficult to obtain, particularly in the near-infrared (NIR) region around 1000 nm wavelength where powerful continuous wave (CW) lasers are available. Calculations for amino acids have recently been published (Song, 2002). Experimental vibrational near-infrared spectroscopy of small organic molecules shows the dramatic reduction in density of third-harmonic absorption lines due to nuclear motion around 1 μm wavelength as temperature is reduced to 15 K (Boraas *et al.*, 1994), while optical spectra for proteins appear to be featureless in this spectral region (Arakawa *et al.*, 1997). Well away from resonances, the better known zero-frequency values of electronic polarizability may be used. A considerable literature exists on modeling dielectric properties of proteins (Porschke, 1997). Here we model a large molecule as a dielectric prolate spheroid of volume V and ellipticity 0.8, for which the polarizability may be evaluated exactly by the methods of classical electrostatics in terms of the dielectric constant K (Bohren & Huffman, 1983). We take $K = 4$ for protein (Simonson, 2003), giving $\Delta\alpha = \gamma V$ with typical shape factor $\gamma = 0.3$. From equation (2), the relation between temperature and laser power then becomes

$$T = 3 \times 10^{-8} \langle \Delta\theta^2 \rangle \gamma V I \ (\mathrm{W\,cm^{-2}}). \qquad (3)$$

Fig. 2 shows a plot of temperature against laser power, based on this equation, for a virus, a macromolecular assembly (ribosome) and a large protein of length 11 nm, all with $\Delta\theta$ fixed at 4°. We see, for the laser power density of 2.5×10^{12} W m^{-2} (corresponding to a modern 50 W fiber laser operating at 1 μm wavelength, with 5 μm diameter focus) that a temperature of only about 230 K is needed to align a large macromolecule like the ribosome, while 60 K may be needed for the protein. Our first experiments are commencing with phage viruses or TMV, which can be aligned at room temperature in water rather than ice, as seen from the figure. In fact, much higher near-infrared CW single-mode laser power densities up to 50×10^{12} W m^{-2} can be obtained from fiber lasers if all of the output power is focused into a 5 μm-diameter spot.

In the simplest approximation, discussed in more detail below, the spatial resolution d to be expected in a charge-density map reconstructed from diffraction pattern intensities averaged over a small range $\Delta\theta$ of orientations is $d = L\Delta\theta/2 =$ 0.7 nm for a molecular length $L = 20$ nm with $\Delta\theta = 4°$. We note that some secondary structure of proteins (*e.g.* α helices) becomes visible at between 1 and 0.7 nm resolution, but higher resolution (perhaps 0.4 nm) is needed to see β sheets. More accurate estimates of resolution loss with misalignment follow in the next section.

The time taken for a large molecule to align can be roughly estimated from its natural oscillation period and damping time. This can be estimated by modeling the motion as a damped torsion pendulum, with spring constant k given by the second derivative of the energy in equation (1). This model is based on Langevin's treatment of Brownian motion, and the galvanometer mirror fluctuation problem (Uhlenbeck & Goudsmit, 1929). The oscillation period $2\pi(I/k)^{1/2} = 13$ ns for the ribosome (I is the moment of inertia), while the Stokes damping time is $I/(6\pi r^3\eta) = 1.8$ ns, with η the viscosity of dry nitrogen. (This might be used for medium-energy X-ray experiments at atmospheric pressure.) In the low-pressure free-molecular-flow regime, allowance must be made for variation of viscosity with pressure if the mean free path exceeds molecular dimensions. These times should be compared with the transit time of 200 ns for a molecule traveling at 50 m s^{-1} across a 10 μm diameter beam. Thermal fluctuations, given by equation (2), remain after the decay of these damped oscillations.

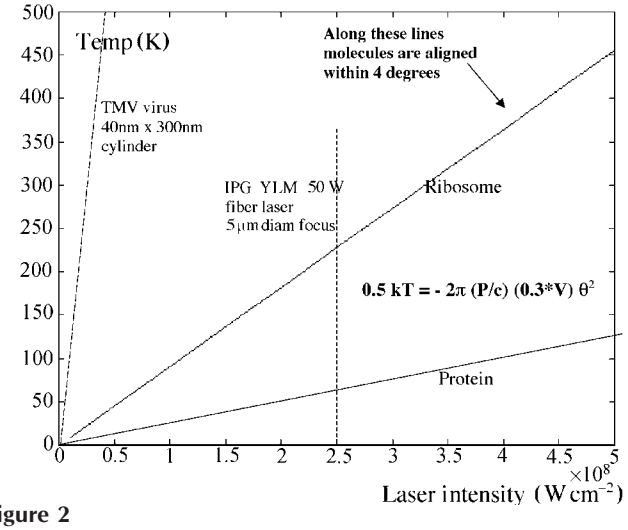

Figure 2
Plot of temperature against laser intensity [equation (3)] for several molecules (virus, length 300 nm; ribosome, length 35 nm; large protein, length 11 nm) for a fixed degree of alignment. The vertical line indicates a 50 W CW fiber laser.

Since the electronic and vibrational polarizabilities are additive, considerable enhancement of the alignment effect can be obtained by operating a tunable laser at a frequency near a vibrational resonance (Friedrich & Herschbach, 1995). Anomalous dispersion then enhances both the polarizability and its anisotropy. The treatment above includes only anisotropic molecular shape effects arising from electronic polarizability – the addition of vibrational anisotropy near resonance and an isotropic polarizability tensor will produce alignment forces even for spherical molecules. This enhancement corresponds to multiphonon processes and, for sufficiently high laser power, would result in dissociation of the molecule. We calculate the temperature rise of water during the transit time to be 4 K at a laser power of 2.5×10^{12} W m^{-2} and 1 μm wavelength.

These considerations refer to the alignment of molecules about a single axis using plane-polarized light or, with elliptically polarized light, to fixing the direction but not the sense of molecular axes. (The energy depends only on $\mathbf{E}^2$, not the sense of $\mathbf{E}$.) Using elliptical polarization, the energy, H, is therefore unchanged by a twofold rotation about any axis, resulting in four degenerate orientations. Two-dimensional diffraction patterns obtained from a mixture of all these orientations at small diffraction wavelength are subject to the projection approximation (a flat Ewald sphere, in which three-dimensional twofold symmetries become mirror lines or inversion symmetry when projected onto two dimensions) and inversion symmetry (Friedel's law). For a molecule without symmetry, the combination of all these operations produces patterns of *mm* symmetry from an equal population of all four equivalent molecular orientations. Several methods have been proposed to fix molecular orientation absolutely, one of which has recently been tested (Sakai *et al.*, 2003). Here the Stark linear term in the Hamiltonian, sensitive to any permanent dipole moment (but with zero time average for alternating fields), is used by the addition of a static field to the experiment, whose sense determines the molecular orientation. Alternatively, a fractional-cycle laser pulse that does not change sign might be used (Seideman, 2004).

The effects of recoil due to the momentum transfer associated with diffraction have been considered. In the worst case where all recoil appears as angular momentum, it may be shown that the rotation angle in the interval between electron arrivals is less than 0.01° at an electron beam current of 1 μA, producing random orientational fluctuations of this magnitude that are well beyond the spatial resolution of the experiment.

As discussed in our earlier paper (Spence & Doak, 2004), each aspect of this proposal has been demonstrated experimentally in the literature, except the extension to large molecules. See, for example, measurement of degree of laser alignment for small molecules (Stapelfeldt & Seideman, 2003), electron diffraction from a beam of iceballs (Bartell & Huang, 1994), electron diffraction from laser-aligned molecules (Hoshina *et al.*, 2003), and inversion of an electron diffraction pattern from one nanotube to an atomic resolution image by iterative solution of the phase problem (Zuo *et al.*, 2003). The practicality of the method depends on the recent development of very high power CW fiber lasers and on the favorable dependence of the alignment error on the inverse cube of the molecular dimension.

3. Resolution limits due to misalignment

In this section, we present simulated continuous diffraction patterns for a small organic molecule (in two dimensions) and a protein (in three). The intensity of these patterns is summed over a small angular range of orientations, and the results used as the input to a phasing program based on the Fienup–Gerchberg–Saxton HIO algorithm. This solves the phase problem, allowing the molecular potential to be recovered, whose resolution is examined for various degrees of misalignment. The phasing algorithm assumes that the complex scattering distribution is the Fraunhoffer far-field diffraction pattern of the electrostatic potential of the molecule, to which it is related by simple Fourier transform. (Curvature of the Ewald sphere and multiple scattering are neglected.) The algorithm iterates between real and reciprocal space, imposing known constraints in each domain. Initially, we treat a simple two-dimensional projection of a molecule with orientational error due only to rotation about the projection axis. For the case of the three-dimensional lysozyme molecule, the diffraction data are first assembled into a three-dimensional reciprocal-space volume, and the iterations are based on three-dimensional Fourier transformation. In our case, the constraints are the measured Fourier moduli and the known sign of the scattering potential. Note that the misoriented diffraction data are inconsistent (especially at high angles) with the true potential we are trying to reconstruct. The approximate boundary of the molecule (its support function) is also a powerful convex constraint, and it has recently been shown that this can be obtained iteratively from the known autocorrelation function of the molecule (Marchesini *et al.*, 2003) or by using a modification of the Ozlányi–Sütő flipping algorithm (Oszlányi & Sütő, 2004). Having reconstructed the potential from the phased (complex) diffraction pattern, we then estimate the resolution in the potential map and plot it as a function of increasing misalignment. Fig. 3(*a*) shows the electrostatic potential for copper phthalocyanine (CuPhTh), while Fig. 3(*b*) shows the corresponding 'oversampled' diffraction pattern. [The sampling interval is optimal for the scattered *intensity* according Shannon's theorem: it is half the Bragg angle for scattering from a periodically repeated molecule (Marchesini *et al.*, 2003).] For comparison, Fig. 3(*c*) shows a misaligned pattern obtained from many diffraction patterns summed over a Gaussian distribution of orientations with 5° standard deviation. The patterns have been rotated slightly about the *c* axis (normal to the page and parallel to the electron beam) and no other rotations are considered for this first simplified arrangement.

While diffraction patterns from misaligned molecules may contain very high resolution (high-angle) scattering, not all of this is meaningful, since the resulting retrieved potential will not faithfully represent the molecule. (The HIO algorithm finds the best potential consistent with the misaligned diffraction data and other constraints. 'Best' is defined by the minimum Euclidean distance ε in Hilbert space between suitably constrained sets, where ε is the HIO error metric). Because of its popularity in crystallography, we show the crystallographic *R* factor

$$R(u_o) = \frac{\sum_{g<u_o} \left| |F_g^e| - |F_g^m| \right|}{\sum_{g<u_o} |F_g^m|}, \qquad (4)$$

where the sum extends to a given resolution limit $d = 1/u_o$. $|F_g^e|$ are the measured moduli of the (misaligned) Fourier coeffi-

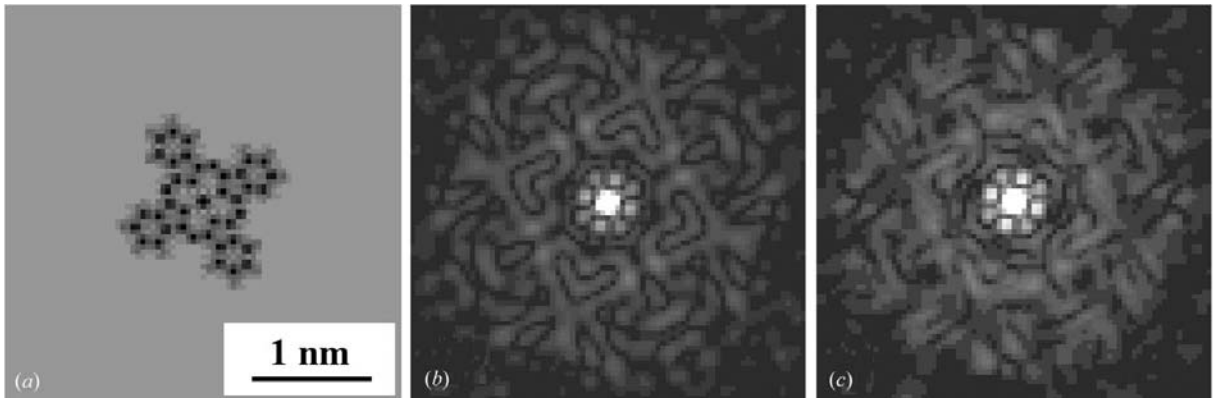

Figure 3
(*a*) Electrostatic potential map for copper phthalocyanine. (*b*) Simulated oversampled diffraction pattern of the supercell. (*c*) Sum of diffraction patterns normally distributed over a 5° angular range.

cients (*e.g.* Fig. 3*c*) used in the HIO algorithm, while $|F_g^m|$ are the true values from the known structure (*e.g.* Fig. 3*b*). R takes no account of the phases of the Fourier coefficients and may be related uniquely to structure only if atomicity is assumed. Experience from protein crystallography suggests that an R factor of about 0.2 is the maximum acceptable value, however we choose $R = 0.15$ as our condition for fidelity. We therefore wish to find the largest misalignment possible that gives the resolution $d = 0.8$ nm needed to see secondary protein structure, subject to the constraint $R < 0.15$.

In order for alignment based on shape anisotropy to be possible, a molecule must have an elongated shape. (Other sources of anisotropy are also possible, as mentioned above.) Such a molecule will oscillate about its center of mass. If the long axis has length L, then, for single-axis alignment, atoms at the end of a spheroidal molecule move through the largest distance during oscillation of $d = \Delta\theta' L/2$, where $\Delta\theta'$ is the full angular range of motion and d may be considered the average resolution (smearing). (This is consistent with the idea that smearing due to misalignment, which increases radially in reciprocal space, does not limit resolution if it occurs over distances smaller than the width of the shape transform.) Then $\Delta\theta'$ may be associated with the standard deviation $\langle \Delta\theta^2 \rangle^{1/2}$ from equation (2). However, this expression for d depends on the dimensionality of the alignment and the shape of the molecule, since shape affects the number of atoms undergoing the largest motions, and most atoms move by a smaller amount. We therefore define $d = k\Delta\theta L$, where $0.5 < k < 1$, and use simulations to find the value of k for simple cases. An analogy with the Debye–Waller factor for crystals can be made. However, while uncorrelated molecular rotations in crystals may affect the temperature factor (Pauling, 1930), in our case, with the beam coherence width comparable to one molecule and no interference between different molecules (as occurs in crystals), the azimuthal smearing of our two-dimensional diffraction pattern about the normal to the pattern does not affect the (radial) temperature factor, but does affect image resolution through its influence on phases.

Fig. 4 shows a plot of R factor against resolution d for several misalignment angles of CuPhTh diffraction patterns. The resolution d has been limited by terminating the diffraction pattern at scattering angle $\theta = \lambda/d$. The region below the horizontal line indicates the domain in which faithful reconstructions ($R < 0.15$) can be obtained with sufficient (or better) resolution than that needed to see secondary structure ($d <$ 1 nm) in a protein. Note that the shape of these curves will depend on any *a priori* information, such as convex or non-convex constraints used in the phasing algorithm. For proteins, this might (but did not) include knowledge of the protein

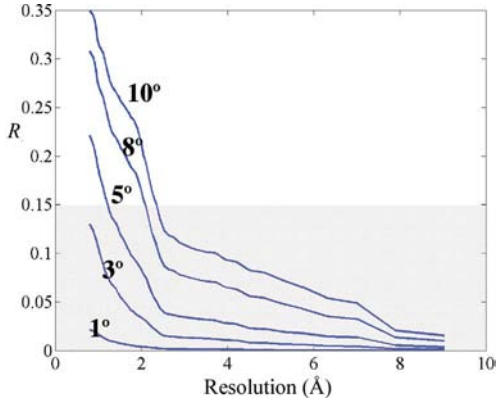

Figure 4
Crystallographic $R(u_o)$ factor plotted against resolution $d = 1/u_o$ for CuPhTh for several misalignment angles $1 < \Delta\theta < 10°$. R measures the agreement between the Fourier moduli of the estimate and those of the model up to a specified resolution d. We consider $R = 0.15$ to be the minimum acceptable value for a faithful reconstruction.

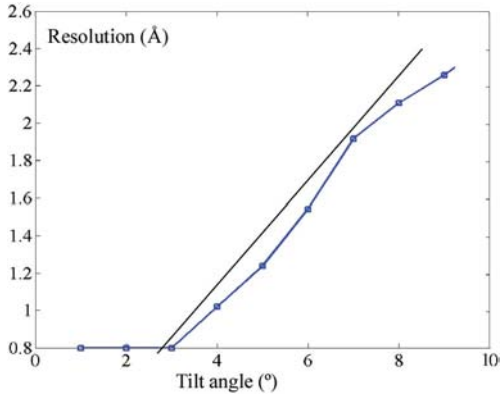

Figure 5
Resolution against tilt angle for a single CuPC molecule, from Fig. 3, at $R = 0.15$. The straight line represents the approximation $d = k\Delta\theta L$, where $L = 2$ nm, the approximate molecular diameter for CuPhTh, and $k = 0.7$.

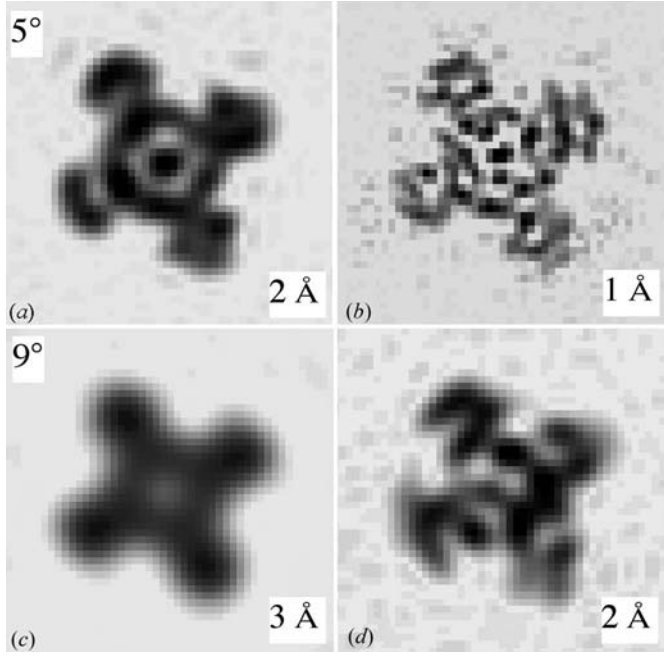

Figure 6
Reconstructed images (projected electrostatic potentials) of a CuPhTh molecule inverted from diffraction patterns with 5° misorientation (top row) at (*a*) 2 Å and (*b*) 1 Å resolution. Lower row shows the images with 9° misalignment at (*c*) 3 Å and (*d*) 2 Å resolution. Although (*a*) and (*d*) are same resolution, (*a*) is a more faithful representation of the molecule as a result of the reduced alignment error.

sequence, the known 'grey-level' histogram constraint for the potential (often taken to be the same for all proteins), an atomicity constraint, or any of the standard density modification constraints used in crystallography (Zhang *et al.*, 2001).

Taking values of misalignment from Fig. 4 along $R = 0.15$, we obtain the plot in Fig. 5 showing resolution d against misalignment angle $\Delta\theta$ for a faithful reconstruction. We see that, if additional constraints are not used, the maximum misalignment that can be tolerated for 0.2 nm resolution for this small molecule is about 8°. We also consider our simple intuitive expression $d = kL\Delta\theta$, relating resolution to misalignment. This is shown as a straight line on Fig. 5, and is a good fit with $k = 0.7$ and $R = 0.15$. The achievement of 7° misalignment for such a small molecule would require lower temperature and/or higher laser power than that shown for the molecules in Fig. 2.

Fig. 6 shows reconstructed images from 5 and 9° misaligned diffraction patterns at several resolutions. An interpretable image can be reconstructed at 5° and 2 Å resolution, whereas there are many erroneous peaks in the image at higher (1 Å) resolution. The reconstructed 9° misaligned image is shown in Figs. 6(*c*) and (*d*), at resolutions of 3 and 2 Å, respectively. Similarly, whereas the 3 Å image is correct, the 2 Å image is incorrect. This indicates the importance of using both a goodness-of-fit index such as R together with the resolution parameter (the angular cut-off in the diffraction data used as input for HIO).

Fig. 7(*a*) shows the projected potential along the [001] direction for a simple protein, lysozyme (5lyz), with coordinates taken from the Protein Data Bank [tetragonal, $P4_32_12$ (No. 96), cell constants $a = 79.1$, $b = 79.1$, $c = 37.9$ Å, 8008 atoms]. The diffraction pattern shown in Fig. 7(*b*) was calculated for an isolated molecule within a supercell, whereas the crystal contains eight molecules per cell. To provide about two-times oversampling, we used a larger unit cell that is 1.2 times bigger than the original one: $a = 94.9$, $b = 94.9$, $c = 45.5$ Å. A single 5lyz molecule, which has 1102 non-H atoms, was placed in the center of the larger cell. We assume elliptically polarized light, with the major axis of the ellipse parallel to the electron beam and to the c axis of the molecule. However, for simplicity, no rotation about this axis is considered. (This would produce effects similar to those seen in the case of CuPhTh above.) The molecular c axis is assumed to fall within a solid angle centered on c. The tilt angle is represented by three Euler angles, ϕ, θ and ψ. The three Euler angles are defined with respect to a coordinate origin at the center of the mass of the molecule, which is assumed to remain stationary as the molecule tilts around. [The z axis lies in the electron-beam direction, which coincides with the long axis of the protein molecule when the tilt error is zero. x and y lie in the plane normal to the z axis, forming a right-handed coordinate system. ϕ is the first rotation angle around the z axis, the second rotation θ is the angle about the x axis, and the third rotation ψ is about the z' axis. (A rotation of the z axis by θ about the x axis generates the z' axis.) This convention can be found at http://mathworld.wolfram.com/.] Since a single 5lyz molecule loses the symmetry it has in the crystal, ϕ ranges from 0 to 360° and θ from the largest negative misaligned angle to the largest positive misalignment angle (*e.g.* from −5 to 5°). In addition, $\psi = -\phi$, since we do not consider rotation around the electron-beam direction. Fig. 7(*c*) shows a simulated diffraction pattern projected along [001], taken from molecules whose largest misalignment angle is $\Delta\theta = 5°$. (The alignment angles were uniformly distributed about the beam axis, which is also the c axis of the molecule. The misalignment solid angle was $\Delta\Omega = \pi\Delta\theta^2 = 0.02$ sr.) The three-dimensional Fourier iterations were performed using a $97 \times 97 \times 47$ voxel array, which took 2 s per iteration on a 2.6 GHz PC and required about 60 iterations to converge. A charge-flipping variant of the HIO algorithm was used, which has been shown to correspond to Fienup's output–output algorithm (Wu *et al.*, 2004). This algorithm uses a density threshold to find the support (Oszlányi & Sütő, 2004).

Fig. 8 shows a plot of R [equation (4)] against resolution d for several misalignment angles for the 5lyz molecule. Again, we set $R = 0.15$ as the threshold. The shaded region indicates the domain in which faithful reconstructions ($R < 0.15$) can be obtained with sufficient (or better) resolution needed to see secondary structure. A plot of resolution against misalignment angle is shown in Fig. 9. (The best-fit value of $k = 1$ is expected to be an overestimate because a uniform distribution of angles was used.) From the curve, we find that, if the largest mis-

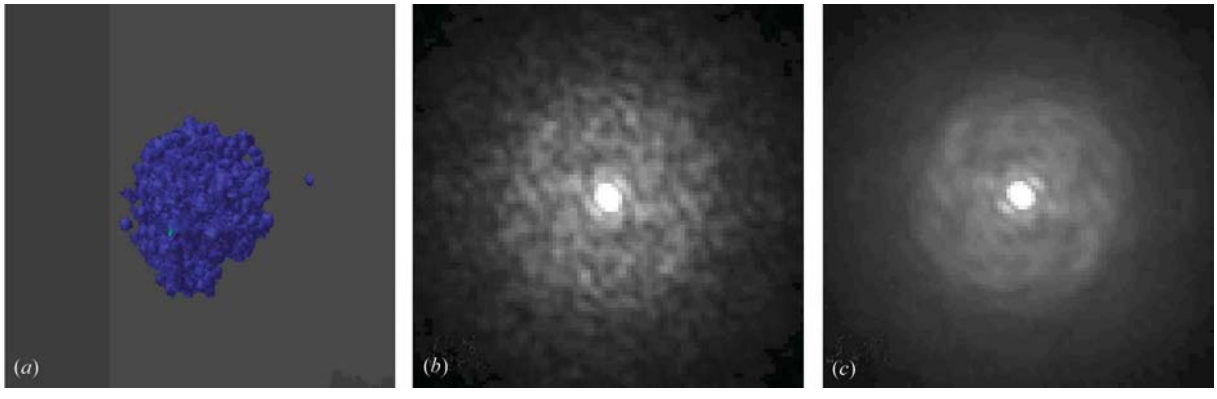

Figure 7
(*a*) Calculated potential of 5lyz projected along the [001] direction at 1.4 Å resolution. The lateral supercell dimension for the projection is 94.9 Å. (*b*) The corresponding oversampled diffraction pattern along the [001]* direction. (*c*) Sum of diffraction patterns distributed over 5°.

aligned angle is 5°, a faithful structural image can be obtained at 6 Å resolution. Fig. 10(*a*) shows a reconstructed image at 6 Å, while its image at 1.4 Å is shown in Fig. 10(*b*). Compared to Fig. 7(*a*), we see that the 1.4 Å potential map has lost considerable information. Fig. 11 shows the lysozyme structure with α helices indicated.

The introduction of noise influences the phase-recovery process mainly by causing the algorithm to stagnate, rather than by limiting resolution; however, prior to stagnation, resolution loss (as defined above) does occur. This process has been investigated both theoretically (Fienup, 1997) and in experimental reconstructions of non-periodic images from both electron (Zuo *et al.*, 2003) and X-ray (Marchesini *et al.*, 2003) data. For these protein images, we have simulated the effect of noise using a parameter noise = (signal/SNR) × random, where SNR is the signal-to-noise ratio and random is a random number from −0.5 to 0.5 (Miao *et al.*, 1998). The noise modulated by the parameter SNR was added to simulated diffraction patterns of copper phthalocyanine, taken from molecules whose largest misalignment angle is $\Delta\theta = 5°$ (Fig. 1*c*). Fig. 12(*a*) shows the influence of the noise on resolution, calculated using the method as described above. (We use a crystallographic *R* factor for data falling within a

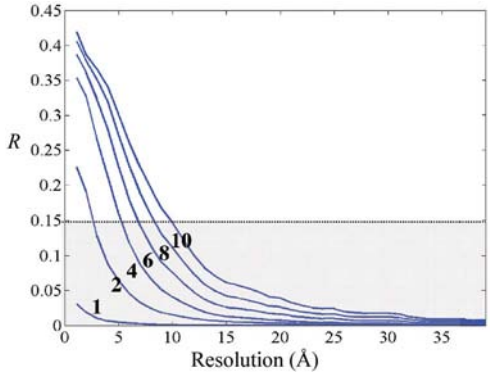

Figure 8
Crystallographic $R(u_o)$ factor plotted against resolution $d = 1/u_o$ for several misalignment angles $1 < \Delta\theta < 10°$ for the 5lyz molecule. *R* measures the agreement between the Fourier moduli of the estimate and those of the model up to a specified resolution *d*. We consider $R = 0.15$ to be the minimum acceptable value for a faithful reconstruction as shown.

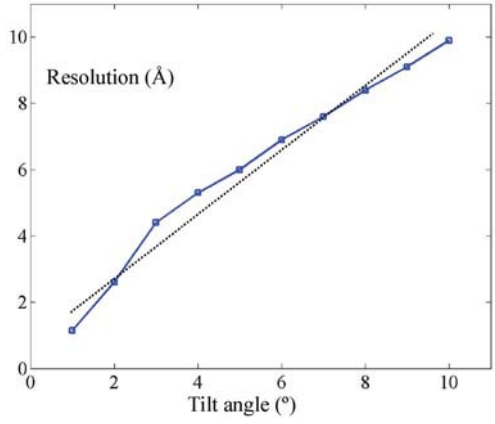

Figure 9
Plot of resolution against misaligned error for 5lyz protein molecule. A line fit can also be found, which represents the approximation $d = \Delta\theta L$, so that $k = 1$ in this case (but see text). The optimal *L* found from the curve is 4.5 nm, the approximate longest length of the molecule.

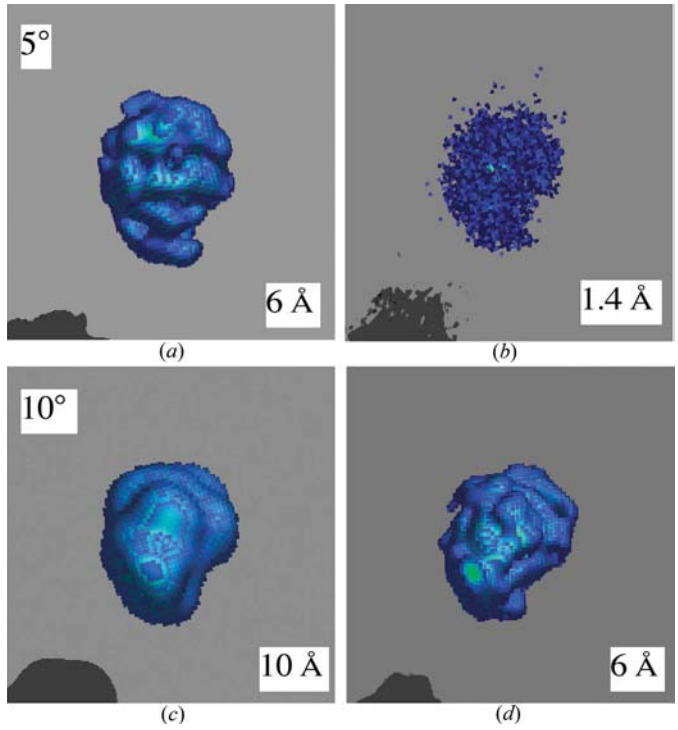

Figure 10
Reconstructed images (projected electrostatic potentials) of lysozyme (5lyz) molecule inverted from diffraction patterns with 5° misorientation at (*a*) 6 Å and (*b*) 1.4 Å resolution, respectively. Reconstruction using 10° diffraction pattern at (*c*) 10 Å and (*d*) 6 Å are also shown.

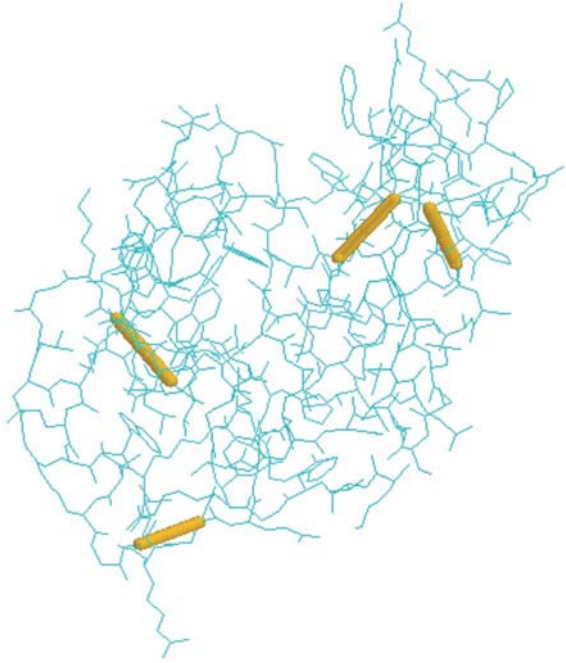

Figure 11
Lysozyme structure. One molecule, showing α helices.

certain resolution circle in the diffraction pattern and an acceptance threshold of $R = 0.15$.) Similar noise was also added to the pattern from a lysozyme molecule with the largest misalignment angle of 5° as shown in Fig. 7(*c*). A plot of resolution against SNR is shown in Fig. 12(*b*) for lysozyme. In both cases, we found that successful inversion required a SNR value greater than 10.

4. Discussion

For a large macromolecular assembly such as the ribosome, with $k = 0.5$ and $L = 35$ nm, a misalignment of less than 3.2° is required for 1 nm resolution. From Fig. 2, this requires a temperature a little below 230 K for a 50 W CW laser with 5 µm focus. For the TMV virus, modeled more accurately as a prolate spheroid with radii $a = 150$, $b = 10$ nm, the resulting eccentricity is $e = 0.9977$, giving a shape factor in the polarizability calculation (Bohren & Huffman, 1983) of $\gamma = 10$ for equation (3). For such a rod of length $L = 300$ nm oscillating about its center of mass, the resolution may be taken as $\Delta\theta L/2$, so that a misalignment $\Delta\theta = 0.38^\circ = \langle\Delta\theta^2\rangle^{1/2}$ is needed for 1 nm resolution. Equation (3) then indicates that a temperature of 320 K is required to achieve this with the same laser conditions. (The pitch of the TMV helix is 2.3 nm and could be resolved under these conditions.) We note that this temperature is above the freezing point of water, so that diffraction from TMV in water droplets may be possible.

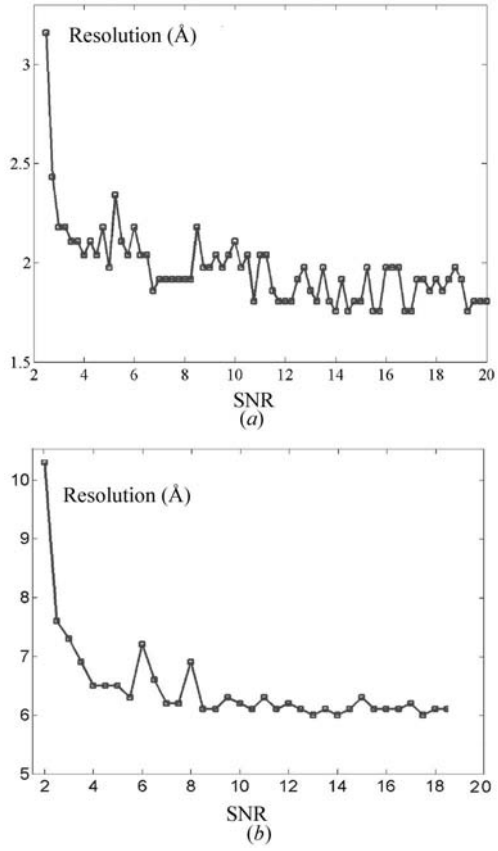

Figure 12
(*a*) Plot of resolution against SNR (with random noise) for a single CuPhTh molecule with a maximum misalignment error of 5°. A resolution of about 1.8 Å can only be obtained if the signal-to-noise ratio SNR is larger than 10. (*b*) Plot of resolution against SNR for lysozyme, with a maximum misalignment error of 5°. A resolution of about 6 Å can be obtained if the SNR is greater than 10.

For smaller proteins, lower temperatures or higher laser powers will be needed. For example, lysozyme may be modeled as a prolate spheroid with radii $a = 2.25$, $b = 1.3$ nm, giving eccentricity $e = 0.8$ and $\gamma = 0.3$ if a dielectric constant of 15 is used for hydrated material. Then a temperature of 20 K is needed to achieve 0.7 nm resolution using a 50 W CW NIR laser. This falls to 6 K if the dielectric constant is reduced to 4. Similar fiber lasers are now commercially available at powers of up to 1 kW. At 1 kW, with a dielectric constant of 15, this resolution for lysozyme can be achieved at 366 K, above room temperature (or at 115 K for a dielectric constant of 4).

In general, for an oblate spheroid of radii $a > b$, combining $d = a\Delta\theta$ with equation (3), we find the resolution d to be

$$d = \left(\frac{1}{b}\right)\left(\frac{3aT}{3 \times 10^{-8} I 4\pi\gamma}\right)^{1/2},$$

where a, b and d are expressed in nm and I in W cm^{-2}. For a given laser power, temperature and aspect ratio a/b, the resolution improves (d decreases) inversely as the square root of the molecular size.

It would appear that data analysis could best be undertaken in two stages. First, the HIO algorithm, which does not assume atomicity and require atomic resolution data, is used to locate the main α helices, as shown for lysozyme in Fig. 11. These act as a framework for the structure and define much of the fold. Then an attempt to further refine the data could be made using the methods of powder diffraction, in which a model structure is parameterized in terms of atom positions, and the parameters varied for best fit to the diffraction data in the manner of a Rietveld refinement. The procedure will nevertheless be limited by the resolution and noise present in the data. A protein structure has been solved successfully using powder data for the first time recently in the case where the structure of a related transformed structure was known (Von Dreele *et al.*, 2000).

Proteins labeled with a heavy atom such as selenium might also be used to further improve the phasing process. The anisotropy of the protein can also be increased by the addition of other molecules to the protein (Glaeser, 2004). We will consider this approach, and the possibility of angular deconvolution, in a subsequent paper. Deconvolution could be used to measure the alignment of the molecules and to enhance the resolution of the image inversion. We note the improved angular resolution of fiber diffraction data (corresponding to our single-axis alignment) over powder diffraction data. For certain structures, simplifications arise due to symmetry. For example, for the simplest helical structures, the diffracted intensity may be independent or only weakly dependent on rotation about the primary axis, so that single-axis alignment may suffice. The methods of fiber diffraction may be useful. In addition, further constraints may be applied in particular

cases. It is now known that 80% of α helices in membrane proteins can be predicted from the sequence alone, and this information might be incorporated into a histogram constraint. These α helices constitute 50% of globular proteins. We have also considered the use of small crystallites of protein, containing, say n^3 molecules in each crystallite. Since there is coherence across the entire crystallite, the scattered intensity for $n = 3$, for example, would be increased by $27^2 = 729$ times if the data collection is not angle-resolved, as in electron diffraction. However, the formation of these crystallites in solution, with identical n for each crystallite, may be very difficult. Our first experiments are about to start using virus particles, with high n-fold symmetry, giving such an intensity advantage, in addition to the possibility of accurate alignment at room temperature due to their large volume.

In summary, we find that subnanometre resolution in diffraction patterns from a laser-aligned molecular beam of small proteins at liquid-nitrogen temperature will require kilowatt CW laser power in the infrared, whereas tens of watts are needed for large macromolecular assemblies and viruses. The conditions of power and temperature depend strongly on the molecular size, shape and (poorly known) dielectric constant, but most importantly on size.

We are most grateful for discussions with Professors R. Glaeser, J. Frank, T. Seideman and R. Packard. Work supported by NSF award SGER DBI-0429814.

References

Arakawa, E. Tuminello, P., Khare, B. & Milham, M. (1997). *Biospectroscopy*, **3**, 73–81.

Bartell, L. S. & Huang, J. (1994). *J. Phys Chem.* **98**, 7455–7456.

Bohren, C. E. & Huffman, D. R. (1983). *Absorption and Scattering of Light by Small Particles.* New York: Wiley.

Boraas, K., Lin, Z. & Reilly, M. (1994). *J. Chem Phys.* **100**, 7916–7919.

Charvat, A., Lugovoj, E., Faubel, M. & Abel, B. (2002). *Eur. Phys. J.* D**20**, 573–577.

Fienup, J. (1997). *Appl. Opt.* **36**, 8352–8360.

Fraschini, E., Bonati, L. & Pitea, D. (1996). *J. Phys Chem* **100**, 10564–10568.

Friedrich, B. & Herschbach, D. (1995). *J. Chem Phys.* **99**, 15686–15696.

Glaeser, R. (2004). Personal communication.

Hoshina, K., Yamanouchi, K., Ohshima, T., Ose, Y. &Todokoro, H. (2003). *J. Chem Phys.* **118,** 6211–6223.

Lindinger, A.,Toennies, J. P. & Vilesov, A.F. (1999). *J. Chem. Phys.* **110**, 1429–1435.

Marchesini, S., He, H., Chapman, H., Hau-Riege, S., Noy, A., Howells, M., Weierstall, U. & Spence, J. (2003). *Phys. Rev. B*, **68**, 140101(R)–140104(R).

Miao, J., Charalambous, C., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–343.

Miao, J., Sayre, D. & Chapman, H. N. (1998). *J. Opt. Soc. Am.* A**15**, 1662–1669.

Neutze, R., Wouts, R., Spoel, T. van der, Weckert, E. & Hadju, J. (2000). *Nature (London)*, **406**, 752–753.

Oszlányi, G. & Sütő, A. (2004). *Acta Cryst.* A**60**, 134–141.

Pauling, L. (1930). *Phys. Rev.* **36**, 430–442.

Porschke, D. (1997). *Biophys. J.* **66**, 241–247.

Sakai, H., Minemoto, S., Nanjo, H., Tanji, H. & Suzuki, T. (2003) *Phys. Rev. Lett.* **90**, 083001–083003.

Seideman, T. (2001). *J. Chem Phys.* **115**, 5965–5973.

Seideman, T. (2004). Personal communication.

Simonson, T. (2003). *Rep. Prog. Phys.* **66**, 737–787.

Song, X. (2002). *J. Chem Phys.* **116**, 9359–9364.

Spence, J. & Doak, T. (2004). *Phys. Rev. Lett.* **92**, 198102–198104.

Stapelfeldt, H. & Seideman, T. (2003). *Rev. Mod. Phys.* **75**, 543–552.

Uhlenbeck, G. E. & Goudsmit, S. (1929). *Phys. Rev.* **34**, 145–152.

Von Dreele, R. B., Stephens, P., Smith, G. & Blessing, R. (2000). *Acta Cryst.* D**56**, 1549–1553.

Weierstall, U., Chen, Q., Spence, J., Howells, M., Isaacson, M. & Panepucci, R. (2001). *Ultramicroscopy*, **90**, 171–185.

Wu, J. S., Weierstall, U. Spence, J. C. H. & Koch, C. T. (2004). *Opt. Lett.* **29**, 2737–2740.

Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *International Tables for Crystallography*, Vol. F, edited by M. Rossmann & E. Arnold, ch. 15.1, pp. 311–324. Boston/Dordrecht: Kluwer Academic Publishers.

Zuo, J. M., Vartanyants, I. A., Gao, M., Zhang, M. & Nagahara, L. A. (2003). *Science*, **300**, 1419–1421.