

X-ray lasers for structural and dynamic biology

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2012 Rep. Prog. Phys. 75 102601

(<http://iopscience.iop.org/0034-4885/75/10/102601>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 131.169.239.71

The article was downloaded on 14/09/2012 at 09:06

Please note that [terms and conditions apply](#).

X-ray lasers for structural and dynamic biology

J C H Spence^{1,2}, U Weierstall^{1,2} and H N Chapman^{3,4}

¹ Department of Physics, Arizona State University, Tempe, AZ 85287, USA

² Lawrence Berkeley Laboratory, Berkeley, CA 94720, USA

³ Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany

⁴ University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany

E-mail: spence@asu.edu

Received 25 January 2012, in final form 23 July 2012

Published 13 September 2012

Online at stacks.iop.org/RoPP/75/102601

Abstract

Research opportunities and techniques are reviewed for the application of hard x-ray pulsed free-electron lasers (XFEL) to structural biology. These include the imaging of protein nanocrystals, single particles such as viruses, pump–probe experiments for time-resolved nanocrystallography, and snapshot wide-angle x-ray scattering (WAXS) from molecules in solution. The use of femtosecond exposure times, rather than freezing of samples, as a means of minimizing radiation damage is shown to open up new opportunities for the molecular imaging of biochemical reactions at room temperature in solution. This is possible using a ‘diffract-and-destroy’ mode in which the incident pulse terminates before radiation damage begins. Methods for delivering hundreds of hydrated bioparticles per second (in random orientations) to a pulsed x-ray beam are described. New data analysis approaches are outlined for the correlated fluctuations in fast WAXS, for protein nanocrystals just a few molecules on a side, and for the continuous x-ray scattering from a single virus. Methods for determining the orientation of a molecule from its diffraction pattern are reviewed. Methods for the preparation of protein nanocrystals are also reviewed. New opportunities for solving the phase problem for XFEL data are outlined. A summary of the latest results is given, which now extend to atomic resolution for nanocrystals. Possibilities for time-resolved chemistry using fast WAXS (solution scattering) from mixtures is reviewed, toward the general goal of making molecular movies of biochemical processes.

(Some figures may appear in colour only in the online journal)

This article was invited by R H Austin.

Contents

1. Introduction	2		
2. Instrumentation	3		
3. Data analysis—nanocrystals	7		
4. Data analysis—single particles	11		
5. The phase problem in SFX	14		
6. Radiation damage with femtosecond pulses	15		
7. Time-resolved nanocrystallography	18		
8. Snapshot SAXS	19		
9. Sample preparation	20		
10. Key issues—challenges and solutions	21		
		<i>10.1. The optimum beam energy and identification of resolution-limiting factors</i>	21
		<i>10.2. Sample preparation</i>	22
		<i>10.3. Injector development</i>	22
		<i>10.4. Time-resolved nanocrystallography</i>	22
		<i>10.5. Software developments and phasing</i>	23
		<i>10.6. Detectors</i>	23
		<i>10.7. Snapshot solution scattering (fast WAXS)</i>	23
		Acknowledgments	24
		References	24

1. Introduction

The recent invention and development of the hard x-ray free-electron laser (XFEL) has opened up new opportunities for structural biology. Before the turn of the century, it was believed that true single-molecule imaging using scattered radiation would never be possible, because the radiation dose needed to achieve sufficient high-angle elastic scattering would, as a result of inelastic processes, destroy the molecule (Breedlove and Trammel 1970). Theoretical work had suggested that short pulses might outrun radiation damage (Solem 1986), but no experimental results existed. If we consider a small x-ray beam which forms a near-delta function in time, into which we may pack as many photons as possible, it is clear that damage-free elastic scattering could be obtained regardless of dose, resolution and sample size, down to the single-molecule level.

As a result of recent experiments, we now know that if the dose is delivered quickly enough, it is indeed possible to outrun radiation damage (Chapman *et al* 2006a). In this way we can obtain sufficient image-forming elastic scattering before radiation damage dominates or even begins, thus allowing the possibility of molecular movies by a snapshot ‘diffract-and-destroy’ method (Neutze *et al* 2000). Since the damage, which occurs after termination of the incident pulse, may destroy the sample, this method requires a constantly refreshed supply of identical particles, such as molecules or perhaps viruses. So far atomic resolution by this method could only be attained by taking advantage of the coherent amplification of Bragg scattering from nanocrystals. But it is now clear that only the need for engineering advances in XFEL and sample injector technology (brightness, beam diameter, repetition rate, hit rate, water background) prevents single-molecule imaging, not the more fundamental problem of radiation damage. If molecular snapshots are recorded in many random orientations and the molecules assume a limited number of conformations, then the snapshots might be sorted according to their orientation and conformation (Frank 2006) and merged to form a three-dimensional molecular movie (Huldt *et al* 2003). This sorting process is only possible if the conformational and orientational changes can be distinguished, as demonstrated in simulations (Fung *et al* 2009).

Emma *et al* (2010) provides a report on the capabilities of the first hard-x-ray laser, the Linac Coherent Light Source (LCLS) at SLAC near Stanford, USA, while others are now under construction or commissioning around the world. These spatially coherent light sources operate in a pulsed mode which provides time-resolved ‘snapshot’ x-ray images of proteins, both in nanocrystalline and single-particle form, in their native environment and at room temperature (RT). In addition it is now clear that these sources can indeed generate femtosecond x-ray pulses brief enough to terminate before radiation damage (which ultimately destroys the sample) sets in. In this way we may break the nexus between damage, sample size, dose and resolution (Howells *et al* 2009), thus avoiding the need to freeze samples for damage protection. (The effects of electronic damage, which occurs during a pulse, are discussed later in this review.) When applied to protein nanocrystals, we

will refer to this serial, destructive-readout, ‘diffract-before-destroy’ method as serial femtosecond nanocrystallography (SFX), to distinguish it from single-particle methods, where samples such as viruses are used.

Macromolecular crystallography (MX) at synchrotrons, the most successful technique for protein structure determination, provides charge-density maps of proteins limited in resolution by both crystal quality and radiation damage. The process of finding the correct conditions for growing the large, well-diffracting protein crystals required for MX can take years. MX samples are usually frozen to reduce radiation damage, and the crystallization process usually (but not always) allows only a single protein conformation to be studied. The results from other techniques, such as cryo-electron microscopy and atomic force microscopy, make it increasingly clear that this shortcoming of MX is limiting our view of protein interactions. Recent MX at RT has shown how flash cooling to reduce radiation damage can bias hidden structural ensembles in protein crystals and remodel the conformational distribution of 35% of side-chains, while eliminating the packing defects necessary for functional motions. Thus MX at RT can reveal motions crucial for catalysis, ligand binding and allosteric regulation (Fraser *et al* 2010). Despite valuable progress in time-resolved protein crystallography discussed in section 7, what is urgently needed is a time-resolved technique, which can image individual proteins at subnanometer resolution in three dimensions, in their native environment, unaffected by damage from the imaging radiation.

Both the serial crystallography SFX method recently demonstrated at the LCLS (Chapman *et al* 2011) and single-particle (virus) imaging experiments at this XFEL (Seibert *et al* 2011) address the limitations of MX in structural biology caused by crystal quality and radiation damage.

The idea that the early coherent elastic scattering might provide a high-resolution x-ray hologram of organic material, before it is destroyed, was first analyzed in detail by Solem (1986), who predicted that 10 nm resolution might be possible using 1 ps pulses, and who described a ‘self-shuttering’ mechanism. Doniach (1996) discussed time-resolved holographic crystallography using XFELs, while detailed simulations by Neutze *et al* (2000) provided estimates of resolution for various pulse durations and intensities by tracking the atomic motion following the photoelectron cascade, which vaporizes a sample. Experiments using the FLASH soft x-ray XFEL at the Deutsches Elektronen-Synchrotron (DESY) in 2006 (Chapman *et al* 2006a) demonstrated this ‘diffract-before-destroy’ principle. In these experiments radiation damage is reduced or eliminated by using an x-ray pulse so intense and brief that it terminates before damage processes affect the length scale of interest, yet contains sufficient photons to produce a useful diffraction pattern from the initial burst of elastically scattered photons (Barty *et al* 2011). Using hard x-rays at the LCLS, single x-ray pulses of 30–70 fs duration containing about 7×10^{11} photons of 9 keV have been found to produce diffraction patterns from micrometer-sized crystals of Lysozyme (Boutet *et al* 2012) at RT extending beyond 2 Å. The corresponding radiation dose in that case was 33 MGy/pulse, similar to the Henderson ‘safe

dose' (Henderson 1995) which normally limits resolution in MX at cryogenic temperatures. This was about 30 times higher than the tolerable dose for RT MX measurements with synchrotron radiation (Southworth-Davies *et al* 2007), and yet this dose was only limited by the beamline configuration at the time. The highest doses reported so far were at 3 GGy/pulse, carried out with 6 Å wavelength x-rays which limited the attainable resolution to 7.5 Å (Chapman *et al* 2011).

The limitation due to crystal quality is addressed through the use of microcrystals, down to submicrometer size, which have been used in these works, and by the single-particle approach, which avoids altogether the need for crystallization.

SFX and single-particle imaging also promise to improve the efficiency of the overall process of solving protein structures, and carrying out parameter studies as required for drug discovery. While it may take many years of tedious trials to find the conditions needed to grow large crystals suitable for MX, it seems likely that 'invisible' submicrometer crystals suitable for SFX can be grown much more readily. (Crystal growers frequently observe 'showers of microcrystals' in their growth solutions, and the mother liquid itself might be used for SFX. Only recently was the structure of the important G-protein-coupled receptor solved by conventional MX, using microcrystals (Rasmussen *et al* 2011).)

Finally, the use of an XFEL promises advances in time-resolved protein crystallography, and in snapshot imaging of molecular reactions in solution. Results have so far been obtained from several membrane proteins, soluble proteins and enzymes, as summarized in the final section of this review.

In the following we will review the methods, and attempt to identify the key issues in the development of these capabilities, which may eventually yield high-resolution molecular movies, showing molecular machines at work (Frank 2011). As some of the first practitioners utilizing XFELs for the elucidation of structure and dynamics of macromolecules and their assemblies, we aim to provide the reader with recent insights and experience gained from the limited amount of beamtime that has so far been available. The field has been reviewed in its application to membrane proteins recently by Fromme and Spence (2011).

2. Instrumentation

This short review will not discuss XFEL physics other than the experimental parameters needed for structural and dynamic biology—for an explanation of the self-amplified spontaneous emission mode (SASE) in which an XFEL operates, see Margaritondo and Ribic (2011). For this topic it is important to note that the LCLS and other planned hard x-ray FELs produce pulses that are typically 10 to 200 fs in duration with a pulse energy of up to about 5 mJ. This corresponds to a peak x-ray power of up to 50 GW. At 8 keV, 5 mJ corresponds to about 4×10^{12} photons, and 6×10^{13} photons at 500 eV photon energy. Beamline transmissions from the source to the sample are typically 20%. The pulses are almost fully spatially coherent and are quasi-monochromatic with a bandwidth of about 0.1% (although the wavelength may jitter by 0.3% from shot to shot). Figure 1 shows the general experimental arrangement used

in the first experiments at the LCLS for the study of protein nanocrystals. Bioparticles are sprayed in single-file across the pulsed x-ray beam, and diffraction patterns were read out on a split detector after each pulse, 120 times per second in order to record both high-angle (low angular resolution) and low-angle (high angular resolution) data. Since a beamstop would quickly be ablated by the x-ray beam, a gap or hole in the detector is required to allow the unscattered beam and small-angle scattering to pass to a down-stream beam dump. A second detector (not shown), also split into two panels, is placed behind the first to receive the small angle scattering. The distance between the first detector and the interaction region is about 10 cm. The scattering chamber design, the detectors and the particle delivery systems will be briefly covered here, with emphasis on the limitations imposed by the experimental conditions.

The first instrument used for soft x-ray diffraction from bioparticles at the LCLS was designed by the Advanced Study Group (ASG), a collaboration of Max Planck institutes in Germany. A full account of the CFEL-ASG (CAMP) multipurpose chamber design is given in Struder *et al* (2010), which also describes the pnCCD photon-counting detector used for this work. This detector consists of 1024×1024 , pixels of $75 \mu\text{m} \times 75 \mu\text{m}$ size arranged in two subpanels, which detect 50 eV–25 keV photons—later designs will move to 2048×2048 pixels. A dynamic range of about 1000 photons per pixel is possible at 2 keV, with a quantum efficiency of greater than 0.8, and a maximum frame readout rate of 200 Hz. The readout noise is less than 20 electrons, allowing the detection of single photons. The chamber accommodates time-of-flight ion and electron detectors, which detect fragments from the vaporized samples.

A second instrument, the coherent x-ray imaging (CXI) instrument, devoted to hard x-ray imaging at LCLS, is described in Boutet and Williams (2010). This paper discusses the x-ray focusing optics needed (both KB mirrors and refractive lenses), the design of the beamline hutch, beam attenuators, beam profile monitor, detector requirements and the fragment time-of-flight ion detector. The paper includes a discussion of fast powder diffraction used for studies of nucleation and growth of crystallites. The chamber can be operated under vacuum or atmospheric conditions. The x-ray beam will eventually be focused down to a diameter of $0.1 \mu\text{m}$, and is currently $2 \mu\text{m}$. The detector consists of a number of subpanels or tiles, which may be reconfigured, and small gaps between them (Philipp *et al* 2010).

The repetition rate of current hard x-ray FELs (typically 120 Hz) and the requirement for sample hydration are placing severe demands on the apparatus for the delivery of bioparticles to the x-ray beam. Sample hydration is required for structure determination to be of biological significance in most, but not all, samples. The x-ray interaction chamber is typically pumped down to vacuum pressure to reduce background scattering, or can be backfilled with helium. A single-file droplet beam injected into vacuum (as shown in figures 1 and 5) cools rapidly by evaporative cooling at a rate of about 10^6 K s^{-1} , or at a lesser rate of about 10^4 K s^{-1} if surrounded by a coaxial sheath gas. If the time between

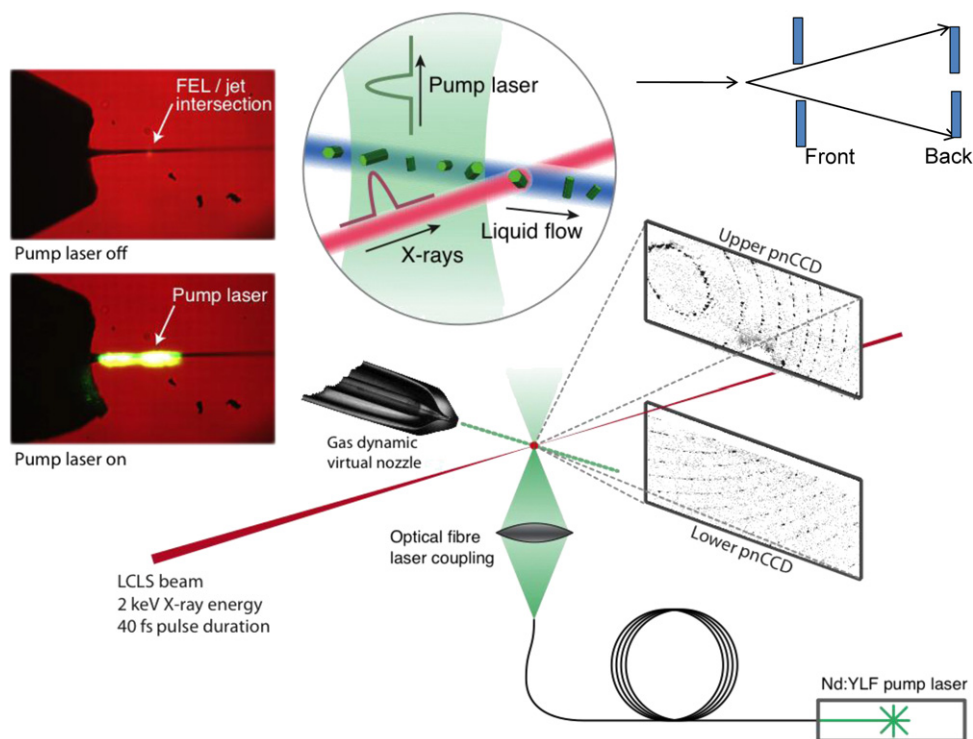


Figure 1. General arrangement used for serial femtosecond nanocrystallography at LCLS (SFX). Hydrated bioparticles are sprayed in single file, in vacuum, across the pulsed x-ray beam. The method of optical excitation of the particles is also shown, using a pump laser. The inset images show (top right) the geometric arrangement for a second, low-angle detector and (at left) experimental images of the particles producing a bright flash (top inset) as they are vaporized by the beam or (below) illuminated by the visible-light pump laser. For a $10 \mu\text{s}$ delay between pump laser and x-ray pulse, the particles travel about $130 \mu\text{m}$. Some arrangements allow on-demand triggering of particle injection using a piezo device. Reproduced with permission from Aquila *et al* (2012). Copyright 2012 The Optical Society.

injection and observation is long enough for freezing to occur, experience from cryo-electron microscopy indicates the need for a transformation to vitreous, not crystalline ice to preserve the structure of embedded bioparticles. In most microcrystal experiments, however, the x-ray exposure is carried out about $100 \mu\text{m}$ from the nozzle tip where the temperature drop is much less than 10 K (as determined by the isotropic high-angle x-ray diffraction from water (DePonte 2012)). This concept of serial crystallography using a continuous stream of particles running across a beam was described in Spence and Doak (2004). The delivery of hydrated bioparticles (protein nanocrystals or viruses, for example) may or may not be synchronized with the x-ray pulses, but, for a free-running liquid or gas-phase jet sprayed across the x-ray beam, one must consider the loss of precious protein (e.g. human protein) running to waste between shots, and wasted photons which miss particles. The particle injector must be able to run for days without clogging and have high ‘luminosity’ (good collimation, small beam area, high particle flux)—ideally it should generate a single-file beam of particles with optimum spacing, each with a water jacket of minimum thickness (to reduce x-ray background) but thick enough to provide an adequate chemical buffer environment for the proteins. Background is minimized if the x-ray beam diameter is about equal to the particle size (less than $0.5 \mu\text{m}$ for a virus) and the water jacket (considered to be an essential part of a protein) is as thin as possible. Since the x-ray pulse is so fast that it freezes all motion, the probability of a particle hit simply depends on the average number of particles per interaction volume $V = D_x^2 D_p$, where D_x is the x-ray

beam diameter and D_p is the particle beam diameter (which is assumed to be larger than the x-ray beam). The hit rate, for an x-ray repetition rate R and particle number density n is then

$$H = nVR. \quad (1)$$

Not all of these hits will produce useful diffraction patterns. The x-ray hit rate decreases as the x-ray beam diameter D_x is made smaller, an important practical problem, but increases with repetition rate R . The flow rate is $F = vA$, where A is the particle beam cross sectional area and v the particle velocity. For a Poisson distribution of interparticle spacings, a maximum single-particle hit rate is obtained with about 37% of pulses failing to hit particles (Bevington and Robinson 2002). For some experiments, simultaneous hits on two particles may be acceptable, such as small crystals whose diffraction patterns can be separated in subsequent analysis, or for the correlated fluctuation methods discussed below. Hit rate is often expressed as a percentage of shots collected that contain useful patterns, or $H' = H/R = nV$.

At the LCLS, currently two ways of injecting bioparticles into the x-ray beam are being used, as an aerosol in the gas phase, and in solution in a liquid jet. For both types of injector, a general focusing principle is used which relies on the reduction in cross sectional area A of a stream of particles embedded in liquid or gas. This reduction is associated with an increase in particle velocity v , where the product Av is approximately constant. In this way clogging problems can be avoided for the liquid jet through use of a large diameter liquid capillary and subsequent gas focusing.

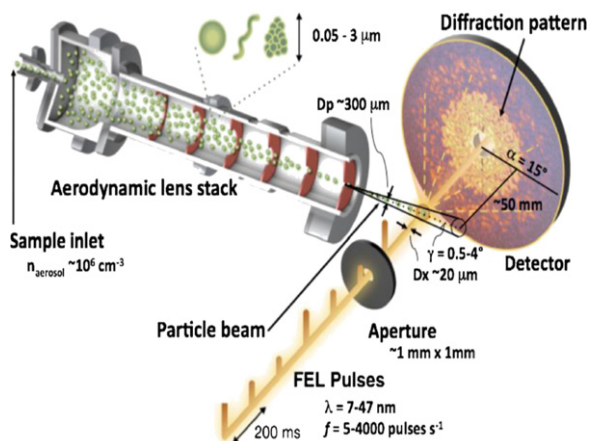


Figure 2. Gas-phase injector. Bioparticles are injected from a nebulizer, electrospray or GDVN nozzle into a stack of gas-focusing lenses from which they emerge to be intercepted by chance by the XFEL x-ray pulses. Time-of-flight spectroscopy may be incorporated and water background is greatly reduced in this arrangement since the particles dry while drifting through the lens stack. Reproduced from Bogan *et al* (2010).

Figure 2 shows a schematic of a gas-phase type injector (Wang *et al* 2005). It uses electrospray (followed by an electrostatic discharging device) or a nebulizer to create a gaseous suspension of wet particles (up to $10\ \mu\text{m}$ in size), which are then led into a stack of aerodynamic gas-focusing elements (Bogan *et al* 2010). Time-of-flight mass spectrometry of the species ionized by the XFEL beam may be incorporated. A differential mobility analyzer might also be used to pre-select monodispersed particles. The background from water scattering is reduced as particles dry out during the relatively long travel time between nebulizer and x-ray beam; however, this drying process may also concentrate salts. Focused particle beams of about $20\ \mu\text{m}$ diameter are readily achieved using this method, with particle velocity $v \sim 150\ \text{m s}^{-1}$. The particle concentration n is then adjusted for maximum hit rate within the chemically allowable limits. However, a synchronized on-demand mode, in which one particle is ejected in response to one pulse from the XFEL photocathode, has not been demonstrated. Hit rates from free-running gas focused systems of this type at the LCLS have varied from much less than one percent in early work up to perhaps 10% on average, using a $20\ \mu\text{m}$ bioparticle beam focus, with occasional bursts of up to 40% hit rate. Despite the low hit rate this type of injector has so far been the choice for single-particle work at LCLS, because many viruses are relatively insensitive to their environment (compared, for example, with membrane proteins), and because of the need to reduce water background scattering (the difference between the x-ray refractive indices of protein and water is very small). The liquid jet injector, which will be discussed next, has also been used to obtain virus diffraction patterns, albeit with higher water background.

For studies of protein nanocrystals, the much higher intensity of the Bragg peaks lies far above the water background and has allowed the use of a liquid jet injector (Faubel *et al* 1988, Weierstall *et al* 2012), as shown in

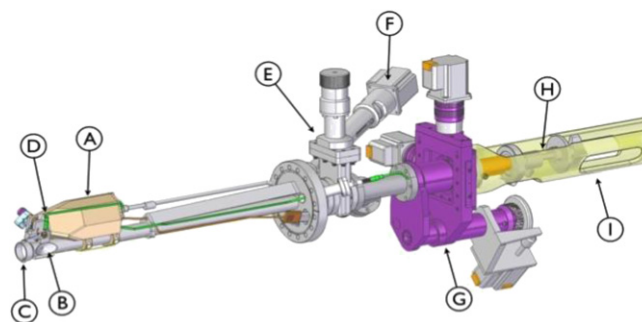


Figure 3. Liquid-stream bioparticle injector. X-ray beam emerges from cone at B, orthogonal to liquid flow from H to C along nozzle rod. In-vacuum CCD microscope at A looks down through prism on interaction region (producing images in figure 1), while fiber-optic line D delivers pump laser light. The spray nozzle may be pulled back behind a gate valve at G for exchange without breaking chamber vacuum. Manipulators at F and H provide precision motion for the jet, microscope optic axis and pump laser. Waste protein and buffer is collected at C. Reproduced with permission from Weierstall *et al* (2012). Copyright 2012 American Institute of Physics.

figure 3. The gas-dynamic virtual nozzle (GDVN) producing the jet consists of a hollow glass capillary with $40\ \mu\text{m}$ ID, centered within a larger glass capillary tube. A buffer solution containing the protein crystals is fed through the inner capillary via a HPLC pump or a pressurized liquid reservoir. All the flexibility of microfluidic switching may be taken advantage of, to allow rapid changes or mixing of solutions. High pressure gas flows in the interstitial space between the tubes and emerges to speed up the liquid, hence focusing it to a diameter of about $5\ \mu\text{m}$ as it emerges (the cone of focused liquid can be seen in figures 4 and 5). After travelling a distance in vacuum, the liquid stream breaks up into a single-file droplet beam due to the Plateau-Rayleigh instability. For a liquid jet in vacuum without gas focusing, the average droplet diameter after breakup was shown by Rayleigh to be 1.89 times that of the liquid column (Rayleigh 1879). The addition of gas focusing in the GDVN reduces the droplet size at high gas pressures to about the same size as the liquid column (Weierstall unpublished). The smallest droplet diameter so far produced with a GDVN nozzle is $0.3\ \mu\text{m}$ (DePonte *et al* 2011), about the same size as a large virus. For the simpler Rayleigh jet (Weierstall *et al* 2008), the droplet breakup may be triggered by a piezo actuator and so synchronized with the XFEL. (A piezo-driven GDVN may also be possible but is more complex.) The liquid jet produced with a GDVN has a typical jet velocity of $10\ \text{m s}^{-1}$ (e.g. a $2\ \mu\text{m}$ diameter jet flowing at $7\ \mu\text{l min}^{-1}$). The x-ray beam may be focused at any point along the stream, where the falling temperature provides a valuable experimental variable, and produces supercooled water and eventually iceballs (Bartells 1986). The gas-focusing effect on the liquid jet depends on both focusing gas pressure and liquid pressure. The liquid capillary has to be centered with micrometer-level precision inside the concentric gas capillary. The cylindrical symmetry and relatively large dimensions do not favor semiconductor lithographic techniques, so that many different schemes have been attempted for the fabrication of these GDVN devices over several years. Current practice is to pass a fiber, with a hand-ground conical tip, into a capillary

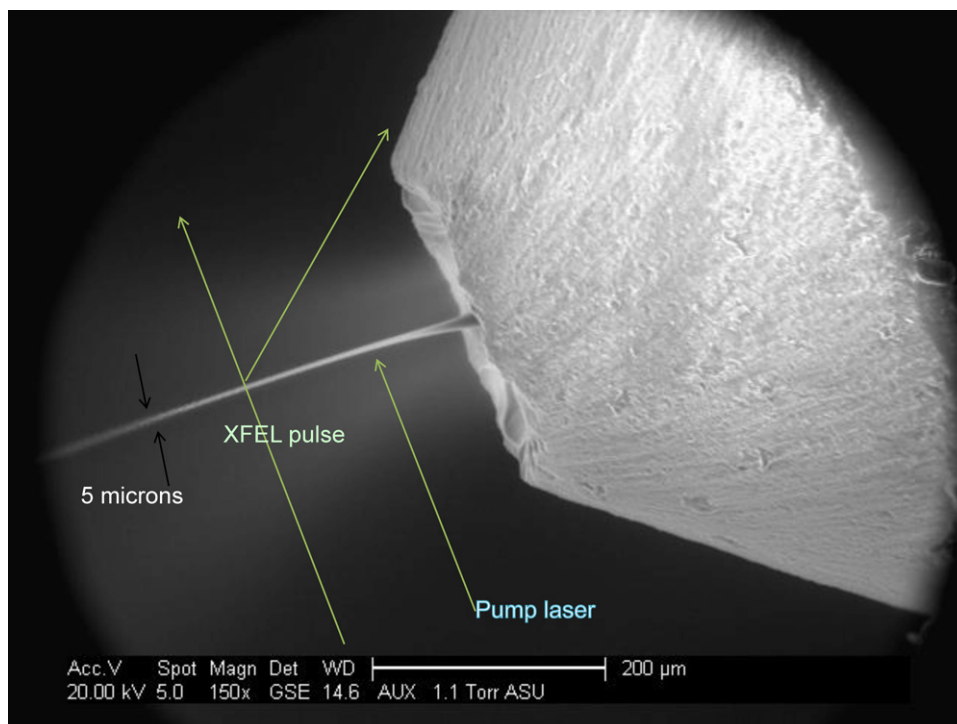


Figure 4. Liquid jet nozzle seen operating inside an environmental SEM (submicrometer droplets cannot be resolved in an optical microscope). The hollow fiber-optic carrying the fluid terminates just to inside (to the right) of the ground cone on this glass capillary tube. From this is seen a bright diverging stream of gas, which is focusing the liquid stream. The positions of the XFEL and pump laser beams are shown. The droplets freeze over a distance of about 1 cm as they cool by evaporation into vacuum, travelling at about 10 m s^{-1} . A flow rate of $10 \mu\text{l min}^{-1}$ is common.

tube with square internal cross section. The end of this square outer glass capillary tube is heated in a propane flame, where the opening is thermally polished and closes down to form a round aperture at the end. The sidewalls leading to this aperture retain the square internal section, and so center the liquid capillary, as shown in figure 5. Fabrication details are given in DePonte *et al* (2008) and Weierstall *et al* (2012), where methods of centering the liquid capillary, the design of the entire injector housing, in-vacuum camera, pump laser mount and manipulator are also discussed. Unlike conventional MX, in SFX one works with bioparticles in RT fluids. Sustained high hit rates of up to 40% have been obtained with the liquid phase injector, because, for an x-ray beam focused close to the nozzle, the emerging liquid jet acts as a localizing medium where the bioparticles can be found with certainty. As one example from early work, a total of 1.8×10^6 shots produced 112 000 nanodiffraction patterns containing more than 10 Bragg spots from photosystem I nanocrystals (PSI), a 6% hit rate, of which 40% were indexable (Kirian *et al* 2011a, 2011b). Images of the liquid jet intersected by the XFEL beam and a pump laser beam may be obtained using the in-vacuum miniature CCD microscope shown in figures 3 and 16, which is crucial for motorized alignment of the jet. The current XFEL beam diameter is about $1 \mu\text{m}$. The liquid jet, microscope optic axis, pump laser and x-ray beam have to intersect each other at a point, with the pump laser slightly upstream along the jet, and therefore mechanical alignment with micrometer-level precision is necessary. Optical fluorescence from protein buffer vaporized by XFEL pulses can be observed with the microscope and assist alignment; however, this signal

is very weak when using harder x-rays due to the lower absorption.

Many variations on this basic scheme have been developed. Following up earlier work on mixing fluids in cells for analysis by SAXS (Park *et al* 2006), a ‘mixing jet nozzle’, using a dual-bore glass fiber, has also been developed. Here two solutions, such as a substrate and an enzyme, may be mixed at the nozzle so that x-ray snapshots can be obtained along the jet as a function of reaction time, as a reaction proceeds. A nozzle for high viscosity liquids (up to the viscosity of toothpaste) is another injector type under development. Injection of high viscosity liquids is essential if membrane proteins in the lipid cubic phase or sponge phase are to be analyzed (SFX from membrane proteins in sponge phase is demonstrated in Johansson *et al* 2012). These injectors work with reduced jet velocity and low flowrate, thereby reducing sample consumption. The jet diameter is usually larger than with GDVN sources ($\sim 20 \mu\text{m}$), which increases the x-ray scattering background. The smallest liquid jets that have been generated with a GDVN source (300 nm diameter) were observed in a TEM (DePonte *et al* 2011) and ESEM, and these may have application in high-energy fast electron diffraction systems (Vredenburg and Luiten 2011) and XFEL diffraction of virus particles at soft x-ray wavelengths of the water window (between the carbon and oxygen absorption edges). Of particular interest for XFELs is the on-demand mode (as used in ink-jet printers), in which one droplet is generated with a piezo actuator synchronized to the XFEL pulse. This would eliminate the loss of protein which runs to waste between shots when using the GDVN source. But

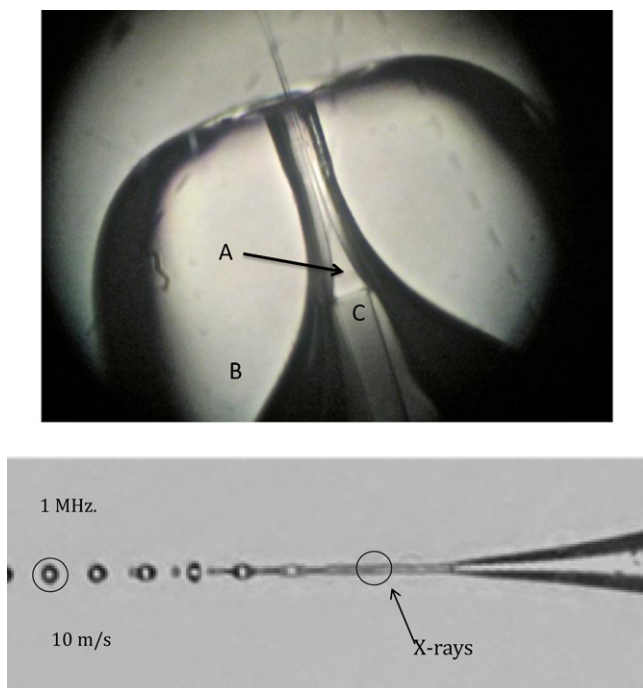


Figure 5. Gas-dynamic virtual nozzle (upper) and breakup of a Rayleigh droplet beam (lower). In the upper figure, a cone of liquid is seen at A being focused as it speeds up under the influence of a coaxial high pressure gas jet running between the outer glass capillary tube B (inner diameter $40\ \mu\text{m}$) and the inner hollow fiber-optic line C carrying the buffer and protein mixture. The stream emerges into vacuum where it will break up into droplets as shown below. The x-ray beam may be positioned in either the continuous-flow or droplet region, along which temperature falls, producing micrometer-sized balls of ice.

so far the smallest droplets created with a drop on demand source have about $30\ \mu\text{m}$ diameter (Weierstall *et al* 2008), too large for use with single particles or micrometer-sized crystals. Another disadvantage of drop on demand technology is that a helium-filled chamber near atmospheric pressure would be needed to avoid freezing of the liquid at the nozzle in vacuum. This is different from the GDVN nozzle and the high viscosity nozzle where the fluid at the nozzle exit does not freeze in vacuum since it is surrounded by a coaxial gas flow. Fluidic switching of GDVN nozzles, in which the liquid flow is turned on and off at millisecond rates, producing single-file slugs of liquid rather than droplets, appears promising as a means of improving efficiency to reduce wasted protein. Recent research is also devoted to the formation of flowing thin liquid bilayer films (Beerlink *et al* 2008) which might pass across the XFEL beam, to provide a continuous flow of membrane proteins such as G protein-coupled receptors (GPCRs), ion-gating channel membrane proteins or perhaps two-dimensional crystals.

The large amount of protein needed for these first SFX experiments will be reduced in the near future in several ways. Firstly, advances discussed below in data analysis will provide converged structure factors with far less data, perhaps a tenth of that used during 2010. Less data are also required if the bandwidth of the FEL pulses could be made broader, so that Bragg peaks are more fully integrated on each shot. Increasing the bandwidth from 0.1% to 2% is expected to give a reduction

in required data by another factor of 10. The SwissFEL facility currently under design aims to provide such pulses. Secondly, the fluidic switching method and low flowrate high viscosity liquid jets discussed above will eliminate most protein which runs to waste between shots. It is not clear at present whether the fluidic switching rate or the detector readout frame rate will provide the bottleneck for data collection. The ‘toothpaste’ jets, suitable for membrane proteins in lipid cubic phase, with their very low flow rates (e.g. $35\ \text{nl min}^{-1}$) require very small volumes of protein. The consumption of protein could also be reduced by using even higher intensity pulses obtained by increasing the pulse power of XFELs or by tighter focusing. This will allow strong diffraction signals to be collected from smaller crystal volumes. Note that decreasing the x-ray spot size will reduce the hit rate, yet still give a net decrease in overall required protein. Finally, the repetition rate of the European XFEL will be over 200 times higher at the LCLS, potentially giving a reduction by 200 times in required sample volume (but note the accompanying detector challenges, as described in section 10). Allowing a $50\ \mu\text{m}$ spacing between shots to avoid the effects of the previous x-ray pulse requires an x-ray repetition rate no higher than 0.2 MHz with the current $10\ \text{m s}^{-1}$ jets. Running the jet at higher speed with proportionally higher pulse rate results in no further loss of sample, since the consumption is proportional to the flow rate times the total measurement time.

3. Data analysis—nanocrystals

Reconstruction of a three-dimensional molecular image from single-particle data requires a determination of the relative orientation of the diffraction patterns from the thousands of randomly orientated particles. This process is simplified for microcrystals, where crystallographic indexing determines the molecular orientation relative to the laboratory frame, allowing Bragg reflections with the same index from different microcrystals to be summed. For the continuous diffraction patterns from single-particles, the process is much more difficult. We first discuss orientation determination for microcrystals.

During microcrystal injection, a filter in the feed line to the injector sets an upper limit on crystal size. The size of the smallest crystals can be determined directly from the fine fringes observed running between Bragg reflections (‘shape transforms’, as shown in figure 6). The number of crystal periods along direction g between facets of the crystallite shown in figure 6 is two more than the number of fringe maxima between the origin and Bragg reflection g . If needed the smallest nanocrystals (some have been observed containing only 6 unit cells on a side) may therefore be selected and excluded from the analysis. We assume that all crystals are smaller than one mosaic block, so that mosaicity effects are not considered initially. There is evidence from line-width analysis in protein powder x-ray diffraction that such microcrystals are more perfect than larger crystals (Von Dreele 2007). However, much work remains to be done to confirm this hypothesis. Because the coherence width of the XFEL spans the entire crystal (not one unit cell, as in MX), the

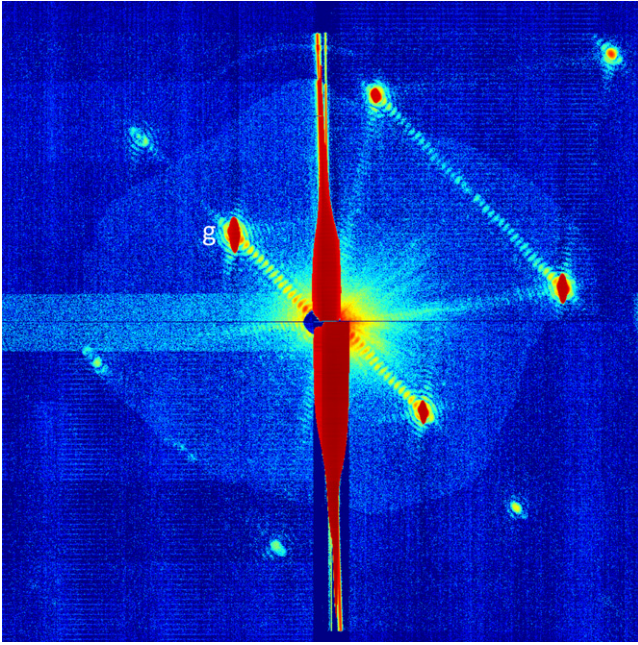


Figure 6. Shape transforms. Single 40 fs XFEL diffraction pattern from a single nanocrystal of Photosystem I recorded in the liquid jet at 2 keV on a rear detector. The thick streak running up the page through the center results from diffraction by the continuous column of liquid. From the number of subsidiary minima we can determine that this nanocrystal consisted of just 17 unit cells between facets along direction g . Reproduced with permission from Chapman *et al* (2011). Copyright 2011 Nature Publishing Group.

fine structure in the patterns from the smallest crystals may provide just the required information on crystal perfection, strains (Cha *et al* 2010) and growth mechanisms (Vekilov 2004) in future research. In addition to internal strains, an additional complication arises from the high proportion of surface molecules in these nanocrystals, which may assume conformations different from the bulk.

For all microcrystal work at XFELs, the snapshot diffraction patterns contain ‘partial’ Bragg reflections (Rossmann *et al* 1979), unlike those at a synchrotron, where continuous crystal rotation by a goniometer provides the angular integration across the Bragg condition needed to obtain a structure factor. The indexing and merging of millions of snapshot diffraction patterns therefore provides new challenges involving the many terabytes of data which result from days of data collection at a rate of 120 diffraction patterns per second. Indexing, for example, must be automated (Leslie 2006), as human examination of patterns is too time consuming. The transfer of tens to hundreds of terabytes of data between laboratories may take weeks.

For the smallest nanocrystals, the intersection of their shape transform with the energy-and-momentum-conserving Ewald sphere will generate scattering in non-Bragg directions, complicating the use of autoindexing software and the use of the conventional mosaicity, energy-spread and beam-divergence corrections used in MX. For plane-polarized monochromatic incident radiation with wavevector \mathbf{k}_i ($|\mathbf{k}_i| = 1/\lambda$) and negligible beam divergence, the diffracted photon flux I (counts/pulse) at $\Delta\mathbf{k} = \mathbf{k}_i - \mathbf{k}_o$ produced by the n th

parallelepiped crystallite, consisting of $N = N_1 \times N_2 \times N_3$ unit cells, is given in the kinematic theory as (Kirian *et al* 2010)

$$I_n(\Delta\mathbf{k}, \mathbf{k}_o, \alpha, \beta, \gamma, N_i) = J_0 |F(\Delta\mathbf{k})|^2 r_e^2 P(\mathbf{k}_o) \frac{\sin^2(N_1\Psi_1)}{\sin^2(\Psi_1)} \times \frac{\sin^2(N_2\Psi_2)}{\sin^2(\Psi_2)} \frac{\sin^2(N_3\Psi_3)}{\sin^2(\Psi_3)} \Delta\Omega = c |F(\Delta\mathbf{k})|^2 S(\Psi), \quad (2)$$

where $F(\Delta\mathbf{k})$ is the structure factor of the unit cell. J_0 is the incident photon flux density (counts/pulse/area) and $\Delta\Omega$ is the solid angle subtended by a detector pixel. Here

$$\begin{aligned} \Psi_1 &= 2\pi a \sin(\theta) \cos(\alpha)/\lambda, \\ \Psi_2 &= 2\pi b \sin(\theta) \cos(\beta)/\lambda, \\ \Psi_3 &= 2\pi c \sin(\theta) \cos(\gamma)/\lambda, \end{aligned} \quad (3)$$

where θ is half the scattering angle, and α , β and γ define the crystal orientation as the angles which the scattering vector $\Delta\mathbf{k}$ makes with the directions of the real-space unit cell vectors \mathbf{a} , \mathbf{b} and \mathbf{c} . $\Delta\mathbf{k}$ is defined by the position of the detector pixel and x-ray wavelength, and defines a point in reciprocal space where the Ewald sphere intersects the shape transform. r_e is the classical radius of the electron, equal to 2.82×10^{-5} Å and $S(\Psi)$ is a function describing the Fourier Transform of the external shape of the nanocrystal (the ‘shape transform’). The x-ray radiation produced by the LCLS is plane polarized, so that the polarization factor for polarization along the unit vector $\hat{\mathbf{u}}$ becomes $P(\mathbf{k}_o) = 1 - |\hat{\mathbf{u}} \cdot \hat{\mathbf{k}}_o|^2$ (Whitaker 1953). An angular integration over the triple product in equation (2) is proportional to $N_1 N_2 N_3$ and the volume of the unit cell, so the measured diffracted counts are therefore proportional to the number of electrons in the crystal—thus for a nanocrystal of just 10 molecules on a side, one has a thousand times more signal than from a single molecule, due to the coherent amplification of Bragg diffraction. Single-molecule imaging at XFELs is thus far more difficult and our intention from the beginning was to start with nanocrystals, and work down by filtration to the few-molecule or single-molecule level.

The extraction of squared structure factors $|F(g)|^2$ from equation (2) requires an angular integration over crystal orientation around each Bragg condition (over the volume of the shape transform), in addition to a sum over crystal size, and normalization of the data. If sufficient redundancy in the hundreds of thousands of recorded patterns is available, this may be achieved by summing all intensity from many randomly oriented nanocrystals within a small volume around each reciprocal lattice point, thus adopting a ‘Monte Carlo’ approach to integration, based on an assumption of equal probability for all nanocrystal orientations. For full details, and discussion of data normalization, Lorentz factors, background subtraction, beam divergence, energy spread, possible flow alignment and the effect of the particle size distribution, see Kirian *et al* (2011a). This paper also shows that, with sufficient data, this angular integration results in a structure solution where the R -factor shows a minimum when plotted against integration volume. We require that the ratio of these sums over partial reflections for two different reflections converge to the correct ratio of structure factors. Intensities from

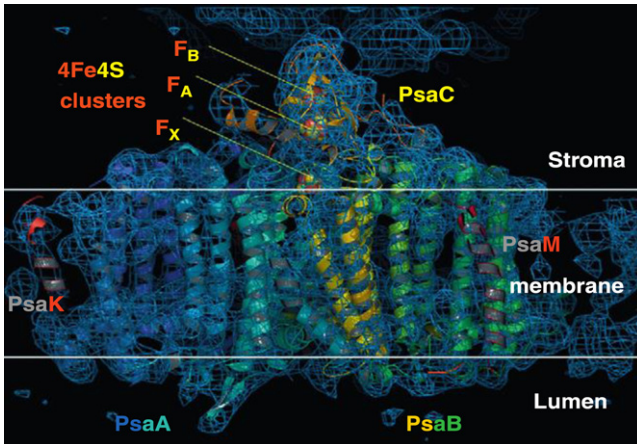


Figure 7. Charge-density map at 0.8 nm resolution, for Photosystem I (PSI) complex (1 MDa, two trimers per unit cell) reconstructed from tens of thousands of 2 keV XFEL snapshots, taken from size-varying nanocrystals in random orientations at 100 K. The cell membrane is indicated, with the Stroma side outermost toward the light. The crystals are hexagonal ($P6_3$, $a = b = 28.8$ nm, $c = 16.7$ nm) with 78% water content. Some of the 12 proteins making up this complex of 72 000 non-hydrogen atoms are labelled. This complex, together with Photosystem II, in all green plants is responsible for all the oxygen we breath (by splitting water in sunlight) and for CO_2 degradation. Reproduced with permission from Kirian *et al* (2011a). Copyright 2011 International Union of Crystallography.

different nanocrystals may therefore be scaled by identifying the same Bragg reflection in different nanocrystals, by relying on convergence of the Monte Carlo averaging process, or as follows.

- (1) Contributions from indexed partial reflections from crystals of different size are added to the same voxel in reciprocal space:

$$\langle I_n(\Delta\mathbf{k}) \rangle_{n,\delta,hkl} \approx J_0 r_e^2 P |F_{hkl}|^2 \langle |S_n(\Delta\mathbf{k})|^2 \rangle_{n,\delta,hkl} \Delta\Omega.$$

- (2) The particle size distribution may be divided out by dividing by the average shape transform

$$|F_{hkl}|^2 \approx \frac{\langle I_n(\Delta\mathbf{k}) \rangle_{n,\delta,hkl}}{J_0 r_e^2 P \langle |S_n(\Delta\mathbf{k})|^2 \rangle_{n,\delta} \Delta\Omega}.$$

Of particular importance is the calibration of the detector pixel positions and errors in the values of $\Delta\mathbf{k}$ assigned to each pixel—this process involves the optimization of many experimental parameters such as the detector working distance, x-ray wavelength, detector tilts and detector tile positions. Virtual powder patterns (the sum of many spot patterns) from a reference sample may be used to calibrate these parameters and to indicate when sufficient data have been collected.

The results of this Monte Carlo method have been compared with data from large crystals of Photosystem I obtained at a synchrotron, and gave R -factors of about 20% for the data collected at 1.9 keV at the LCLS in 2009 (Chapman *et al* 2011, Kirian *et al* 2011a, Barty *et al* 2011). Figure 7 shows the resulting density map, phased by molecular replacement (MR). This density map was limited to about 8 Å resolution

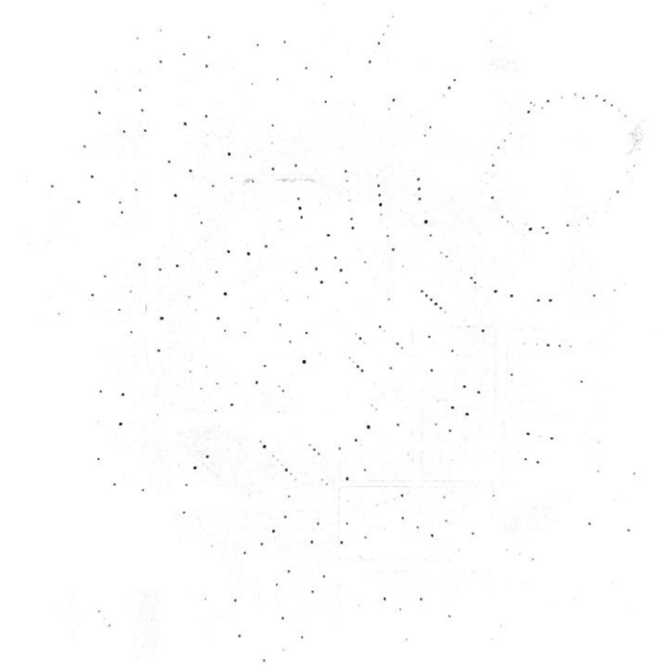


Figure 8. Single-shot 40 fs XFEL diffraction pattern from a single lysozyme nanocrystal recorded at 9.4 keV in the liquid jet at RT, extending to 0.18 nm resolution. The dose of 33 MGy is similar to the Henderson ‘safe dose’ for frozen samples, but 30 times higher than the tolerable dose for RT synchrotron data collection. Reproduced with permission from Boutet *et al* (2012). Copyright 2012 American Association for the Advancement of Science.

by the shortest x-ray wavelength available at the LCLS at that time. The beam energy of the LCLS has since been increased to 9 keV, which has allowed us to record data from nanocrystals to better than 2 Å resolution, using the liquid jet injector. Figure 8 shows a recent diffraction pattern from the model protein Lysozyme, recorded at 9.4 keV using 40 fs pulses, where the Bragg spots extend to the edge of the detector at better than 2 Å resolution. This work (Boutet *et al* 2012), which contains a detailed comparison of R -factors comparing the LCLS results with results from conventional synchrotrons, shows that the ‘diffract-and-destroy’ method of serial femtosecond nanocrystallography (SFX) extends to atomic resolution. The crystal size was about $1 \times 1 \times 3 \mu\text{m}$, limited using a filter in the fluid supply line to the jet, and the beam diameter $3.2 \mu\text{m}$. The energy per pulse was $600 \mu\text{J}$, corresponding to a dose of 33 MGy. Based on the energy deposited by this dose, an impulsive atomic velocity of 1 nm ps^{-1} can be calculated, suggesting negligible atomic displacement during a single 40 fs shot. Of the 1.5 million diffraction patterns collected, about 4.5% were useful, of which 18.4% could be indexed. (More efficient sample delivery methods are described below.) No significant differences could be found between density maps (phased by MR) from the SFX data and synchrotron data; however, if data from turkey lysozyme were used for phasing, differences between the experimental SFX hen and turkey maps were clearly evident. No features related to radiation damage were observed in difference maps between the SFX data and synchrotron data, while the Wilson B factor for the SFX data was similar to synchrotron values. The average number of partial reflections summed to each

full reflection for the atomic-resolution lysozyme results of figure 8 is about 400. A suite of programs (CrystFEL) has since been developed for the analysis of protein nanocrystal data by this Monte Carlo method, which can be downloaded from <http://www.cfel.de/> and which is described in a recent paper (White *et al* 2012). This paper contains a full description of the software, together with details of improved measures of data quality for this new type of nanocrystallography data, help files, documentation, and further discussion of imposed symmetries and automated indexing, using either the MOSFLM or DIRAX routines (which also finds cell parameters). For a new approach to autoindexing for very small crystals based on compressive sensing, see Maia *et al* (2011). The effect of strain, defects, surface atoms, and the resulting diffuse scattering on the merging of data from nanocrystals of different sizes, in random orientations, is discussed in Dilanian *et al* (2012), where the nanocrystals are treated as single particles for the purposes of data analysis.

The *ab initio* Monte Carlo approach avoids modeling, but requires a large degree of redundancy in the data. Equally accurate refinements might be obtained using far less data by adopting methods similar to the post-refinement approach used in MX (Rossmann and van Beek 1999). We note that the three-dimensional shape transform is identical around each lattice point for an unstrained nanocrystal, while a single two-dimensional diffraction pattern shows the various slices which the Ewald sphere cuts through identical copies of the shape transform at a different ‘height’ on each pattern (see figures 9 and 12). This redundant information on the shape transform $S(\Psi)$ from microcrystals of one size class would allow it to be modeled in simple form. It is, however, much more difficult than post-refinement in conventional MX, where each goniometer tilt gives new partial reflection information from the same set of full reflections, rather than each XFEL shot from a different crystal with different shape and size. The use of modeling has now been added to CrystFel in the program ‘partialator’. Here refinement parameters include incident intensity, crystal orientation and beam energy. Use of this refinement approach should allow collection of less data and so reduce the amount of XFEL beamtime required. CrystFEL also contains programs for automated indexing of microcrystal patterns, and for simulating them, based on equation (2). Improvements in the autoindexing success rate for nanocrystals are the subject of constant development. The combined hit detection and indexing rate, an important parameter which determines sample consumption, varies between 10% to over 50%.

As in conventional MX, a serious problem relates to twinning. While individual nanocrystals may not be twinned, indexing alone does not provide sufficient information in space groups that support merahedral twinning to allow merging of data from different microcrystals, without a 50% chance that they are merged in twin-related orientations. This problem is exacerbated when working with partial reflections from crystals of different size. For example in the hexagonal spacegroup $P6_3$, which supports merahedral twinning, a rotation by 180° normal to the c axis brings the indexed reciprocal lattice into coincidence with itself, but not the

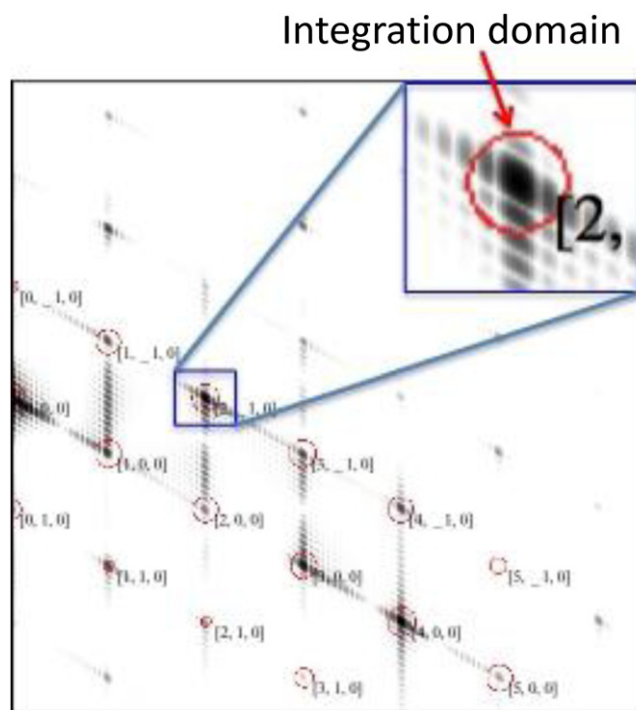


Figure 9. Simulation of a single-shot diffraction pattern from PSI at 1.8 keV, 1.5 mRad beam divergence, 0.1% bandwidth. The circle inset indicates the domain of integration used to merge data from different nanocrystals. The intensity variation shown is a slice on the Ewald sphere through the Fourier Transform of the external shape of the crystal, given by equation (2). Each reflection from the same crystal shows a different slice through the same transform if the crystals are unstrained. Each different crystal has a different transform. The structure factors depend on the volume of this transform.

structure factors, so that there are two ways to combine patterns from two different, untwinned nanocrystals. (The twinning operation takes reflection (h,k,l) to $(k,h,-l)$ in this case.) Future research may resolve this problem, perhaps based on the modeling methods used to deal with twins in MX, in combination with the Protein Data Bank (PDB), or using a model-free method based on expectation maximization, discussed further below.

Before any of this secondary data analysis can be undertaken, a crucial ‘hit-finder’ program (called ‘Cheetah’) must be run which can assemble diffraction patterns from the detector tiles in a commonly used format (e.g. HDF5), apply corrections for the differing gains and background corrections of these tiles, remove the streak due to diffraction from the water jet itself, reject blank frames, and identify the presence of Bragg spots. Many shots will miss nanocrystals entirely, or hit the side of a nanocrystal, causing a diffuse streak in the patterns and loss of diffracted intensity. (Entirely new effects appear if the coherent diffracted orders overlap when using beam divergence larger than the Bragg angle (Spence and Cowley 1978).) The efficiency of this ‘hit-finder’ program for primary data analysis is crucial to the future success of SFX. Improvements in the program may allow old data to be re-analyzed.

Given the current pace of software and hardware development, it seems likely that, despite the tens of terabytes

of data collected at XFEL beamtimes, it may soon be possible to complete the hit-finder analysis in realtime at the beamline. Developments in the post-refinement method applied to nanocrystals show that the number of patterns needed for convergence can be reduced from hundreds of thousands to about ten thousand, collected at perhaps 200 Hz, a rate limited by the readout speed of current detectors. In that case, future SFX analysis may soon allow complete data collection, analysis and MR phasing of a structure within a day (or more quickly at the higher repetition European XFEL).

4. Data analysis—single particles

The almost complete spatial coherence of an XFEL allows all the techniques for phasing and image reconstruction of coherent diffractive (lensless) x-ray Imaging (CDI) to be applied to femtosecond single-particle diffraction patterns. As a model sample we first consider a sphere of uniform electron density. The far-field coherent diffraction pattern of a sphere of radius R , embedded in a medium, is given by (Bohren and Huffman 1983, Starodub 2008)

$$I(q) = I_0 r_e^2 \Delta\Omega |\Delta\rho|^2 \left(4\pi \frac{\sin Rq - Rq \cos Rq}{q^3} \right)^2, \quad (4)$$

where I_0 is the incident fluence, $\Delta\rho$ is the difference of the complex electron density of the sphere to the medium, and $q = (4\pi/\lambda)\sin(\theta)$ is the photon momentum transfer for a scattering angle (pixel location) 2θ . The term in brackets is the 3D Fourier transform of a uniform sphere. The difference electron density can be written in terms of complex refractive index n

$$r_e^2 |\Delta\rho|^2 = \left(\frac{2\pi}{\lambda^2} \right)^2 |\Delta n|^2,$$

$$|\Delta n|^2 = |n_{\text{prot}} - n_{\text{water}}|^2 = (\delta_{\text{prot}} - \delta_{\text{water}})^2 + (\beta_{\text{prot}} - \beta_{\text{water}})^2,$$

where δ and β are the optical constants and the subscripts refer to the protein and the water medium. Far from absorption edges $\Delta\rho$ is independent of wavelength.

The intensity pattern $I(q)$ consists of circular rings (similar to Airy rings) with minima spaced in q approximately by π/R . The maximum intensity of the rings is given by

$$I_{\text{max}}(q) = I_0 r_e^2 \Delta\Omega |\Delta\rho|^2 16\pi^2 R^2 / q^4. \quad (5)$$

This q^{-4} dependence leads to the familiar problem in coherent diffractive imaging of having to simultaneously record strong intensity at low angles and much weaker intensities at higher angles, with the range of intensities often exceeding the dynamic range of the detector.

The fastest variation of diffracted intensity across the detector will be due to interference between x-rays scattered by points in the object located furthest from each other: the sphere diameter, in other words. Measuring this maximum fringe frequency requires at least two detector pixels per period, or a so-called Shannon sampling of $\Delta q = \pi/(2R)$. In this case the pixel solid angle is $\Delta\Omega = (\lambda/2\pi)^2 \Delta q^2$ so that equation (5) becomes

$$I_{\text{max}}(q) = I_0 r_e^2 |\Delta\rho|^2 \pi^2 \lambda^2 q^{-4}, \quad (6)$$

which is independent of sphere radius.

At resolutions approaching the atomic scale, we can approximate the structure as random, in which case there is no change in inter-object correlation with length scale. In this case the approximate intensity per Shannon sample is given by (Huldt *et al* 2003)

$$I_{\text{max}}(q) = I_0 r_e^2 \langle |f|^2 \rangle N_{\text{atom}} \lambda^2 / 4R^2, \quad (7)$$

where N_{atom} is the number of atoms in the object and f is the atomic scattering factor, in units of electrons. As observed in cryo-electron microscopy the transition from a steep falloff with q to a constant dependence occurs at a length scale below 1 nm. In crystals there is correlation at atomic scale lengths due to the repeat from unit cell to unit cell, which of course gives the Bragg amplification mentioned above. However, the molecules are not exactly identical, giving a multiplicative Debye–Waller factor to equation (7) that varies as $\exp(-Bq^2/8\pi^2)$ where B is the Debye–Waller factor.

Even with an incident photon count of 10^{13} photons per pulse, focused to a $0.1 \mu\text{m}$ diameter beam, the XFEL single-shot scattering from one biomolecule is seen from equations (5) to (7) to be extremely weak, falling off rapidly at the high angles needed to form a high-resolution ($<1 \text{ nm}$) image (Huldt *et al* 2003, Shneerson *et al* 2008, Starodub *et al* 2008). For example, a 500 kD molecule illuminated by 3×10^{12} photons in a single pulse scatters about 4×10^{-2} hard-x-ray photons per Shannon pixel per shot at a scattering angle corresponding to $d = 0.18 \text{ nm}$ resolution ($d = \lambda/(2 \sin(\theta/2))$ for scattering angle θ). For the much larger $0.44 \mu\text{m}$ diameter capsid of the mimivirus, as many as 2 000 000 photons may be scattered. We note, however, that the resolution in a reconstructed real-space image is normally independent of particle size. Using the GTM and EMC methods discussed below, however, it has been found that these methods can only reconstruct to a resolution of 1/8 of the particle size (Giannakis *et al* 2012). Clearly, large numbers of ‘single-shots’ must be combined from identical particles to form a three-dimensional image, in which the signal-to-noise ratio S will, according to Poisson statistics, improve after M patterns per orientation class to $S = MN_s / \sqrt{M(N_s + N_b)} \approx \sqrt{MN_s} (1 - N_b/2N_s)$ if the noise is small, with N_s counts per shot per pixel and N_b background counts per pixel per shot (assuming Poisson noise). However, this idealized picture is far from experimental reality, because any additional noise, such as readout noise, thermal noise, stray light or photons, which produces even one photon per pixel per shot, will completely dominate the signal, which is much smaller than one photon per pixel at large angles. In addition, one must contend with gain variations between detector subpanels, the limited dynamic range discussed in section 2, and occurrence of saturated pixels.

The coherent hard-x-ray low-angle scattering intensity from a large molecule is proportional to the square of the number of electrons (see equation (4) at $q \approx 0$), whereas the high-angle scattering intensity (in the atomic resolution regime mentioned above) is proportional to the number of atoms. Using a photon-counting detector with a typical dynamic range of perhaps 1000 photons, saturation of pixels from low-angle scattering must be avoided, and the use of multiple stacked detectors, each with a central hole passing radiation to the

next, has been proposed to increase dynamic range per shot. Then the solid angle subtended by a detector pixel decreases with increasing distance between sample and detector.

It may seem that for merging of thousands of diffraction patterns from similar randomly oriented single particles (such as a virus), the same methods as used in the cryo-electron microscopy (Cryo-EM) community could be used (Frank 2006). Here noisy low-dose projection images of many copies of a particle, lying in many random orientations, are recorded within the field of view of each image, and must be merged to produce a three-dimensional image (Spence 2003). However, the XFEL diffraction patterns also require solution of the phase problem, and, unlike real-space Cryo-EM images, there is no requirement for correction of electron lens aberrations, while an enantiomorphous ambiguity arises from the Friedel symmetry of low-resolution patterns, not present for real-space images. In addition, diffraction patterns have an origin, unlike images, and the background due to ice in Cryo-EM images must be treated differently from the background in a x-ray diffraction pattern due to diffraction from a water jacket surrounding the particle. Building on previous work on iterative phasing of continuous diffraction patterns, three main approaches have been developed for the reconstruction of a three-dimensional image (density map) from many randomly oriented snapshot single-particle x-ray diffraction patterns, and for dealing with the associated problems of particle inhomogeneity. We will give here only a brief outline of the general principles of these methods focusing on key issues.

The simplest method, the ‘common line’ approach, uses the fact that 2D diffraction patterns from two different projections of the same object must show the same intensity along one line passing through the origin, where the two planes in reciprocal space intersect (this assumes that the Ewald sphere is flat to a good approximation). The addition of a third plane in reciprocal space necessarily intersects both of these planes, and the lines along which it does so then fixes the angles between all three. Finding these ‘common lines’ of intensity from a set of patterns in random orientations is difficult in the presence of noise, and is further complicated by the inversion symmetry of the patterns at low resolution (Shneerson *et al* 2008). If allowance is made for curvature of the Ewald sphere, orientation determination is possible using just two patterns by finding common arcs (Bortel and Tegze 2011). An eigenvalue approach is given in Yang *et al* (2010).

The generative topographic mapping (GTM) approach is illustrated in figure 10, simplified for the case of a three-pixel (x, y, z) detector and single-axis rotation of a particle, in order to illustrate the principle of the method. With this simplification, a snapshot diffraction pattern can be represented as a three-dimensional vector, with each component representing the scattered intensity value at a pixel. Rotation of a particle traces out a loop (a one-dimensional manifold) in this three-dimensional space of intensities. Determining this manifold allows one to assign an orientation to each snapshot. In general the detector has N pixels and particle rotation about 3 axes generates a 3D-manifold in the N -dimensional Hilbert space of pixel intensities. The manifold is seen to be parameterized by a three-dimensional latent space defined by the three Euler angles defining the particle orientation.

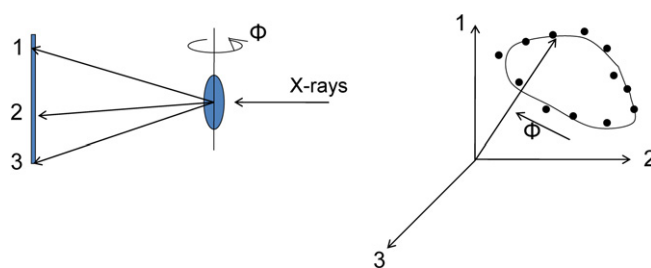


Figure 10. The principle of Hilbert-space methods of merging single-particle data, shown for the simplified case of a 3-pixel detector. Each pixel becomes a coordinate in a 3D vector space in which a single vector defines a diffraction pattern. The vector traces out a loop as a particle rotates about a single axis, allowing snapshots taken from random orientations to be sequenced correctly in the absence of noise. A separate loop can be expected for a different molecular conformation. Noise permitting, when combined with symmetry information, this allows orientation and conformational changes to be distinguished, in principle allowing the formation of a molecular movie from snapshots taken from random orientations.

Many difficulties arise, including the transformation from angular increment to coordinate change in N dimensions, the effects of noise and conformational changes, for which the reader is referred to the original papers (Fung *et al* 2009, Giannakis *et al* 2012, Schwander *et al* 2012). In the simplest case, a second conformation would define a second distinct loop; however, the effects of noise thicken the manifolds so that they may overlap. The key issue of distinguishing changes in particle orientation from conformational changes (essential in order to make a 3D ‘molecular movie’) is resolved using the fact that the operations associated with conformational change commute, while those associated with the rotation group do not. An important feature of this approach is that all the data are used for all the analysis, rather than selecting sub-classes (e.g. of orientation or conformation) for successive analysis. However, even in the absence of noise, a minimum number of scattered counts is needed to identify a particular orientation, proportional to the number of distinct orientations sought. The computational demands of this approach, like the EMC method discussed next, are considerable and set the limit on the size of the largest molecule which can be analyzed.

A third approach has been based on the principle of expectation maximization (EM) (Doerschuk and Johnson 2000, Loh and Elser 2009), and was demonstrated experimentally for inorganic particles by Loh *et al* (2010). The method is most simply explained in two dimensions (2D), for the case of a set of noisy 2D pictures $I(k)$ of the same non-symmetric object, which are known to lie in any one of four orientations $i = 1, 4$ differing by a 90° rotation about their normal. Here k is the image index and extends over the $N \times N$ pixels of the pictures. A model is first assumed, which may consist of random values, and is generated in each of the four orientations i (expansion). Assuming Poisson noise, the probability $P_i(k)$ is calculated that an experimental image $I(k)$ came from each model in orientation i . To avoid the occurrence of extremely small numbers in the first iteration, these probabilities are normalized to unity. The process is repeated (maximization) for each image, giving a set of

coefficients $P_i(k)$. Four new models are then formed from the weighted sum $M(i) = \sum_k P_i(k)I(k)$. Since the four initial orientation-generating operations applied to the model are known, it is then possible to return the four new models to the same orientation, average them, and use the average as a new estimate of the model. Iterations then continue from the first step. An experimental demonstration of the method using low resolution 2D x-ray shadow images has been demonstrated, using as few as 2.5 photons per image (Philipp *et al* 2012). A simplified variant of this method has been developed, in which maximization of the likelihood function is replaced by a simple search for the best-fitting orientation (Tegze and Bortel 2012).

The method has some similarity to Cryo-EM methods based on cross-correlation between experimental patterns to find similar orientations (Sjores 2010), but has the advantage that cross-correlations are computed between experimental patterns and a model, rather than between every possible experimental pair, so that the computation time is linear, rather than quadratic, in the number of patterns. Like the GTM method, this approach uses the entire body of data in each update of the model parameters; however, the GTM method does not constrain the data to fit a three-dimensional model density. The relationship between these algorithms is discussed in Moths and Ourmazd (2011), where it is suggested that the GTM and EMC approaches differ only in their method of regularization. A certain minimum number of counts per diffraction pattern will be needed to successfully assign each diffraction pattern to an orientation class.

In both these methods, solution of the non-crystallographic phase problem may be integrated with the problem of orientation determination. A considerable literature has been devoted to single-particle phasing by methods, which iterate between real and reciprocal space, imposing known information in each domain. This information includes the sign of the scattering potential, the measured diffraction intensities and perhaps the molecular envelope. In brief, iterative phasing, its connection to the Shannon sampling theorem (leading to the idea of ‘oversampling’), the importance of a knowledge of the single-particle envelope (the ‘support’ function), the relationship of this approach to density modification in crystallography, and to constrained optimization and iterated projection onto convex (or non-convex) sets, can all be traced through the key papers by Sayer (1952), Perutz (1954), Gerchberg and Saxton (1972), Fienup (1982) and Millane (1990). These methods use an optimizer, acting on a suitable metric, to iteratively project from a set of feasible solutions, subject to a known constraint, to the closest solution in another set, subject to a different known constraint. Reviews can be found in Spence (2008) and Marchesini (2007), which discuss issues of uniqueness and convergence, and provide references to many more recently improved algorithms. A software package (‘Hawk’) implementing these ideas together with the Shrinkwrap algorithm (Marchesini *et al* 2003) (which updates the support estimate during the iterations) is described in Maia *et al* (2010). In summary, given sufficiently high quality three-dimensional diffraction intensity data, the single-particle phase problem

can normally be solved by iterative methods for a ‘real object’, meaning that the three-dimensional diffraction intensity distribution has inversion symmetry, in the absence of any knowledge of the support function.

In practice many difficulties complicate the phasing of experimental CDI data, as discussed in detail in experimental papers using CW soft x-rays (see Miao *et al* 1999, for the first successful inversion of x-ray data, Chapman *et al* (2006a, 2006b), for a high-resolution three-dimensional inversion and detailed analysis of experimental conditions, and Nelson *et al* (2010) for lensless x-ray imaging of yeast cells at 11 nm resolution). Perhaps the most important problems when using methods derived from the Feinup Hybrid Input–Output (HIO) algorithm (Fienup 1982) are the lack of inversion symmetry in the x-ray diffraction data and the loss of information at very low angles due to the need for a beamstop. (The ‘curved’ Ewald sphere certainly eliminates inversion symmetry in two-dimensional patterns for the continuous distribution of scattering from single particles at high resolution. Friedel’s law does not relate points on the Ewald sphere reflected through the origin, except approximately at low angles for limited resolution.) Lack of inversion symmetry may also arise from other causes, such as stray light in the chamber or background scattering. The degree to which algorithms can deal with the resulting ‘complex objects’ varies, while simulations show that an accurate knowledge of the support function compensates for this loss of information (Fienup 1987). The problem is solved using Ptychography (see Rodenburg (2008) for a review); however, this mode is difficult to combine with the diffract-and-destroy method for inhomogeneous particles. (Ptychography solves both the twin image problem and the limited field-of-view problem of the original Gabor in-line holography method.) We note too the application of the crystallographic phasing method MAD to single-particle diffraction, described by Scherz *et al* (2008). In practice three-dimensional phase retrieval is much more robust than in 2D, and the experience from LCLS shows that the almost full spatial coherence of the illuminating beam (as compared with partial coherence at synchrotron sources) improves the success of phase retrieval.

Figure 11 shows a 70 fs single-shot diffraction pattern obtained at the LCLS from a single pseudo-icosahedral Mimi virus particle (diameter about $0.4 \mu\text{m}$), and its two-dimensional reconstruction, obtained using the Shrinkwrap algorithm included in the ‘Hawk’ software package (Maia *et al* 2010, Seibert *et al* 2011). The diffraction patterns shown were obtained using the gas-phase injector and the water jet, where for the latter the diffraction streak from the liquid column is seen across the pattern. (In later work, this streak has been oriented to run across the gap in the detector.) Both Mimi virus patterns were recorded on the rear (low angle) detector using scattered radiation, which had passed through the gap in the front detector, as shown in figure 1. The constraint, that the sample area may be estimated from its known autocorrelation function was also used, and the Seibert paper discusses the treatment of missing low-angle data using a modal analysis. The resolution, using 2 keV x-rays, is 32 nm, and the total number of scattered x-ray photons per shot is 1.7×10^6 , with 1.6×10^{10} photons per square micrometer incident per shot.

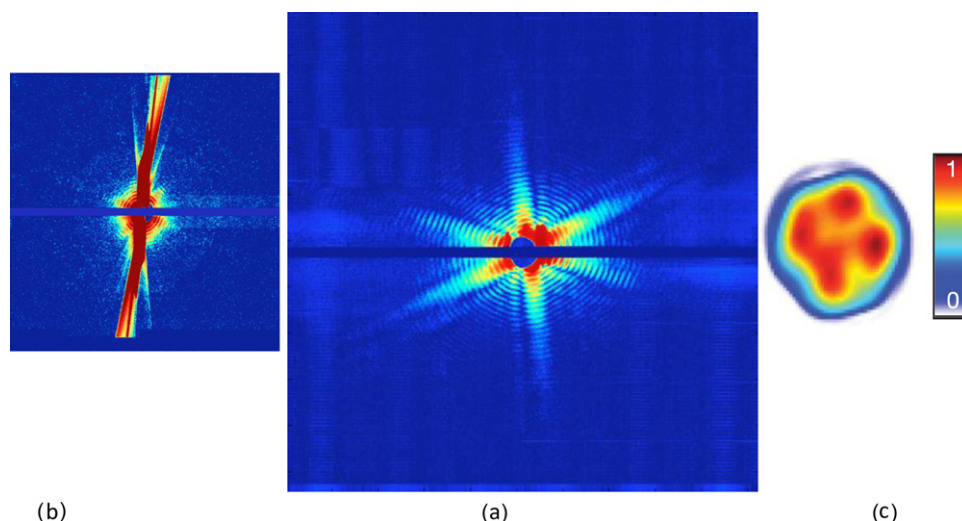


Figure 11. (a) Single-shot 70 fs diffraction pattern, using 2 keV x-rays, from a single Mimi virus (size about $0.3 \mu\text{m}$). (b) Pattern obtained on rear detector from chlorella virus particle in water jet, showing streak from liquid column. Data collection here in the water window reduces water background. (c) Reconstruction of projected image of virus from (a), showing 32 nm resolution and some internal structure. More recent work has merged many of these patterns to produce a three-dimensional image. Reproduced with permission from Seibert *et al* (2011). Copyright 2011 Nature Publishing Group.

The Chlorella virus pattern shown in figure 11(a) was recorded at an x-ray energy of 516 eV (in the water window) using the front detector and the water jet, the detector side edge extends to 7.8 nm resolution—scattering from the virus extends most of the way to this limit, and is currently being analyzed.

Unless the correlated fluctuation methods described in section 8 are used, a ‘hit-finder’ program similar to ‘Cheetah’ is also needed for single particles, to eliminate blank shots, patterns with simultaneous hits of several particles or cross-contamination from a previous sample or from empty water droplets and detector artifacts.

Particle inhomogeneity (which increases with particle size) is the most important problem for single-particle XFEL imaging, and may be solved in principle by the ability of the above methods to distinguish conformations, if sufficient high quality data are available. (A similar problem arises in Cryo-EM.) Similar particles coated with varying thicknesses and shapes of water or ice create an even greater difficulty, such as the wet ‘hair’ on a Mimivirus. Data from particles such as cells, formed from similar but differently arranged components, could only be merged for three-dimensional reconstruction if a suitable topological constraint can be found and implemented. Two-dimensional snapshot imaging of cells, using the 13.5 nm wavelength radiation at FLASH is described in Seibert *et al* (2010). A method for obtaining a three-dimensional reconstruction from a single shot is described in Schmidt *et al* (2008), using multiple incident beams from a beamsplitter. Several authors have pointed out that the curvature of the Ewald sphere provides limited three-dimensional information from a single shot (Spence *et al* 2002, Chapman *et al* 2006a, 2006b). Bergh *et al* (2008) describe other possibilities for extracting three-dimensional information from a single shot, such as Laue diffraction using harmonics, coherent convergent beam diffraction, and multiple-pinhole Fourier transform holography.

5. The phase problem in SFX

While the problem of orientation determination and the associated non-crystallographic phase problem are normally combined in single-particle analysis, there is a sufficiently distinct history of research on phasing in MX to justify a separate discussion of nanocrystal phasing. Nevertheless for a sufficiently small nanocrystal of just a few unit cells on a side, any of the single-particle methods could be applied to reconstruct an image of the entire nanocrystal, subject to the limitations of computing power. Such images would be invaluable for understanding protein crystal growth and surface effects, but unfortunately could only be obtained as two-dimensional projections from a stream of size-varying nanocrystals.

Historically, four phasing methods have been used in MX—direct methods (DM), multiple-wavelength anomalous diffraction (MAD), multiple isomorphous replacement (MIR) and molecular replacement (MR). DM is unsuitable for the large number of atoms in MX and requires atomic resolution data, while MIR requires chemical modifications to the protein, which may be difficult (replacement by heavy atoms such as mercury), leaving MAD and MR as the most popular methods. Replacement of methionine by the selenomethionine amino acid allows use of the selenium absorption edge for MAD, in which diffraction patterns are recorded at several beam energies around this edge. The known contribution from anomalous scattering to the phase of the structure factor allows solution of the phase problem. MR uses a known protein structure from a protein with similar sequence as a model, and now accounts for 67.5% of the structures in the PDB.

The application of these methods to SFX raises new issues and opportunities. The observation of sub-2 Å resolution diffraction patterns using single-shot 70 fs pulses gives confidence that useful structural information is contained in these patterns; however, changes in the ionization state

of the atoms during the pulse must be considered, if not in the atom positions. MR is the most straightforward method; however, for the smallest nanocrystals the influence of surface molecules should be considered, and work by both electron microdiffraction and CDI on strain mapping in inorganic catalyst particles suggests that cell parameters may change as the surface is approached. Crystal structures different from the bulk may also be favored for nanocrystals, as is the case for inorganic nanocrystals. The application of MAD methods to SFX has been treated in detail by Son *et al* (2011), who consider the time-dependent scattering factors which take account of ‘hollow atom’ formation during an XFEL pulse, and the electronic damage dynamics and resulting bleaching effect which result from the extreme x-ray intensities obtained with XFELs. The heavy atoms used for anomalous scattering undergo more rapid ionization than others, resulting in Auger decay, fluorescence, and photoelectron production, with a transition rate proportional to ionization cross section σ . Son *et al* (2011) solve the rate equations that couple the transitions between all subshells. They find, for example, that for Fe, where 27 783 electronic configurations must be considered during a 100 fs pulse, species such as Fe^{20+} are produced, modifying the dispersion correction of these heavy atoms used for phasing. By computing these modified dispersion corrections for SFX, they develop modified Karle-Hendrickson equations for femtosecond MAD. These assume that only heavy atoms (to whose absorption edge the incident x-rays are tuned) undergo electronic damage. (At 8 keV, for example, $\sigma_{\text{Fe}}/\sigma_{\text{C}} \sim 300$.) The scattering strength of these atoms is reduced (bleached) by different amounts above and below an edge, as their inner shells are emptied. Absorption decreases for these hollow atoms as x-ray fluence increases. Using the new K-H equation coefficients they calculate, and data collected above and below an edge, a new phasing method then becomes possible for SFX. In addition, the preferential and significant bleaching of the heavy-atom atomic scattering factors should allow phasing by a kind of isomorphous replacement, without any necessary change in chemical structure.

SFX has created another new opportunity to solve the phase problem, by using the interference fringes seen to run between Bragg spots from the smallest nanocrystals, as shown in figures 6 (Spence *et al* 2011). This approach owes its origins to work by Perutz (1954) using shrinkage for phasing and to Sayer’s (1952) ideas on oversampling, where the Bragg law and Shannon sampling are compared. From equation (2), we see that the number of unit cells N_1 between crystallite facets along direction g is $m+2$, where m is the number of subsidiary fringes seen between Bragg spots. From this equation we also see that the molecular transform modulus $|F(\Delta\mathbf{k})|$ (assuming one molecule per unit cell), which modulates the entire pattern, might be extracted from the measured intensity, using division by the shape transform. A better procedure, rather than phasing individual crystal projections, is to merge the data in three dimensions first. In this way division by small noisy quantities is avoided. This is shown for a simplified one-dimensional case in figure 12, where equation (2) is plotted for the case of two crystals, with 5 and 6 unit cells. (The

third slowly varying curve is the molecular transform $|F(\Delta\mathbf{k})|^2$ in equation (2).) Addition of these oscillations produces a smoother curve, which may then be sampled at a few points between the large Bragg peaks shown to provide the ‘oversampling’ needed by iterative methods to solve the phase problem for reconstruction of one unit cell, not the whole nanocrystal. We sum equation (2) over both crystal size and orientation in the three-dimensional diffraction volume. As shown, zero-crossings in the shape transform occur at different places for each nanocrystal of different size. The smoothly varying three-dimensional shape transform sum can then be divided into the measured intensity sum, thereby allowing the particle size distribution to be ‘divided out’. To provide parameterless normalization, the shape transform sum can also be obtained by summing corresponding pixels around each lattice point, so that the molecular transform envelope washes out. The sampling interval on the summed data may be much coarser than that needed to resolve subsidiary fringes on the larger nanocrystals. Having recovered the molecular transform modulus $|F(\Delta\mathbf{k})|$, the complex molecular transform $F(\Delta\mathbf{k})$ may be obtained by iterative phasing in a supercell. This results in a charge-density map of the unit cell contents. The effects of noise on this method have been studied in detail by Chen *et al* (2012). Simulations can be found in Spence *et al* (2011).

6. Radiation damage with femtosecond pulses

In conventional macromolecule crystallography (MX) using 12 keV x-rays, 98% of the photons are transmitted through the sample without interaction. Of the remaining 2%, 84% of the photons are annihilated in the production of a damaging photoelectron cascade, 8% undergo Compton scattering, and only the remaining 8% generate useful Bragg scattering. While these figures are not greatly altered when using femtosecond pulses, the effects of damage on the diffraction patterns is avoided if the pulse and hence the detection of scattered x-rays terminates before the atomic motion due to the heat generated by the photoelectron cascade gets under way.

The effects of damage depend on resolution, with fine detail being destroyed first (Howells *et al* 2009). For single particles, with 70 fs single-shot imaging of a single large virus at 2 keV, the 2D-resolution in recent work is about 32 nm, and lower for a three-dimensional reconstruction. Bergh *et al* (2008) provides a full review of the damage-resolution trade-off for single particles. Nanocrystals take advantage of redundancy, so that the damage is distributed over many molecules. Any theory of the time-evolution of electronic structure and nuclear motion in response to the intense electric fields of an XFEL pulse in nanocrystals must account for the experimental observation of 7.5 Å resolution Bragg peaks, using 70 fs pulses at 2 keV, and the Bragg spot-fading effects seen for high orders at longer pulse durations, shown in figure 13 (Chapman *et al* 2011, Lomb *et al* 2011), as well as the observation of sub-2 Å resolution with 40 fs 9 keV pulses (Boutet *et al* 2012). We have the remarkable finding that the tolerable dose to obtain macromolecular structures in the diffract-before-destroy mode has a demonstrated lower limit about 40 times greater than the Henderson-Garmen ‘safe dose’

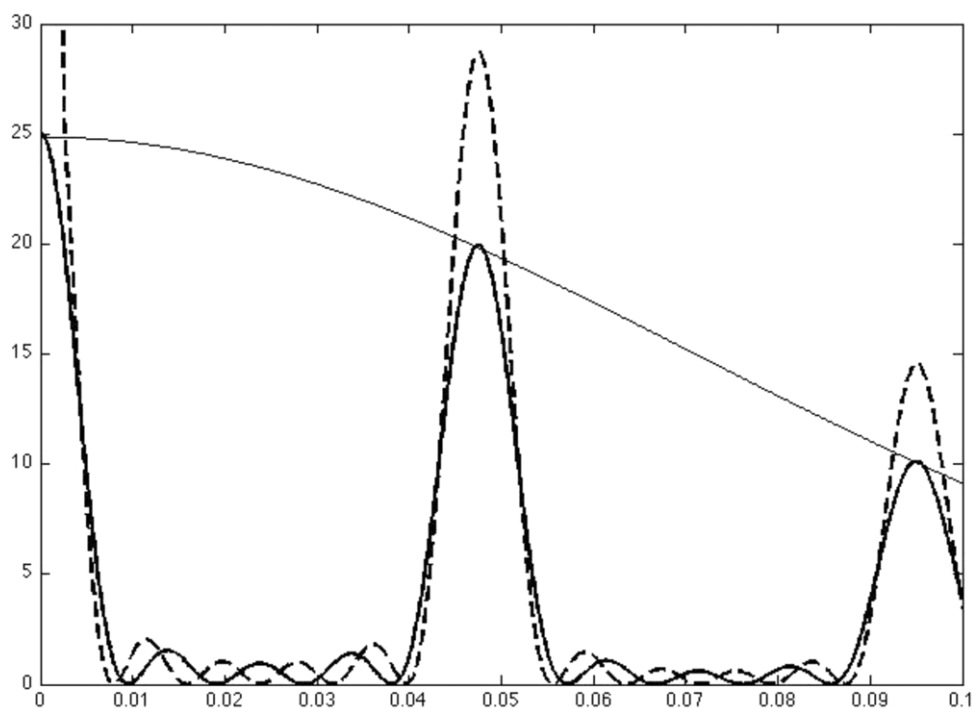


Figure 12. Oversampled nanocrystals in one dimension. Slowly varying curve is molecular transform $|F(\Delta k)|^2$ in equation (2) which is sought for phasing. Additional curves show one-dimensional diffraction (equation (2)) from nanocrystals of five (continuous curve), and six (dashed curve) unit cells (hence four subsidiary maxima between Bragg peaks shown). Addition of many size-varying crystals produces a smooth curve which may be sampled at a few points between the Bragg reflections to solve the phase problem by iterative methods, for reconstruction of the image of one unit cell.

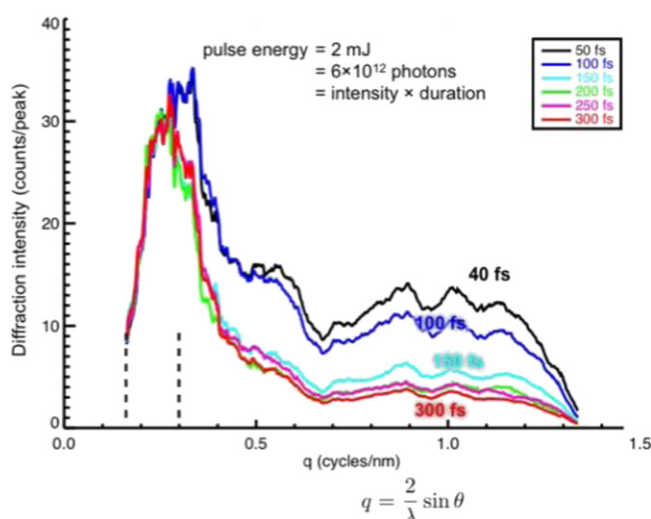


Figure 13. Scattered intensity plotted against scattering vector magnitude for PSI at 2 keV for various pulse durations. The difference between the 40 fs curve and results at 10 fs were found to be negligible. Reproduced with permission from Chapman *et al* (2011). Copyright 2011 Nature Publishing Group.

accepted in the MX community (Henderson 1995, Owen *et al* 2006) which corresponds to about 10 MGy per Angstrom of resolution (Howells *et al* 2009). Thus, the rate at which the dose is delivered is important (reciprocity failure). This experimental finding is only limited by the highest fluences currently generated by LCLS. Calculations, which are in extremely good agreement with these observations (Barty *et al*

2011), predict that the fluence can be increased by several orders of magnitude, essentially to the point where the intensity is so high that it strips all electrons from all atoms. A typical fluence per pulse at 2 keV is 4 kJ cm^{-2} , corresponding to a dose of 3 GGy. The peak irradiance is about $10^{17} \text{ W cm}^{-2}$ for a 70 fs pulse. Time-delay soft-x-ray holography experiments have been used to image the effect of a photoionization explosion (Chapman *et al* 2007). For reference, we note that a 10 (300) fs pulse of 2 keV x-rays corresponds, with Gaussian profile, to a 160 (1600) J cm^{-2} pulse containing 10^{11} (10^{12}) photons, focused here to a $10 \mu\text{m}^2$ area. A generic protein crystal containing 30% water may be modeled as $\text{H}_{160}\text{O}_{50}\text{C}_{52}\text{N}_{13}\text{S}$, with density 1.245 g cm^{-3} . (For urea the composition $\text{H}_4\text{CN}_2\text{O}$ has been used.)

Two main theoretical approaches have been used to understand these experimental results—a molecular dynamics model for the atoms and ions created by the x-ray pulse, (Neutze *et al* 2000) and a hydrodynamic plasma model (Hau-Riege 2011), based on non-local thermal equilibrium radiation transfer (the ‘Cretin’ code). An approach based on numerical solution of the Boltzman equation has also been described (Ziaja *et al* 2006). In the plasma dynamics code (Scott 2001), the temperature of the photoelectron plasma is tracked for a spherical ion–electron soup, which explodes after a few picoseconds. Each time step of the simulation gives atomic level populations, transition rates, heating rates and opacities (the intensity-induced transparency of ‘hollow’ ions with short lives), while Auger decay is obtained from detailed balance equations, using a screened hydrogenic model. Rates for impact ionization are obtained by integration over a

Maxwellian electron distribution. The time-dependence of the temperature can then be used to provide temperature factors in an x-ray diffraction calculation of the early stages of ionization. Ion positions are not tracked directly, but heated through coupling to the electron gas.

In the molecular dynamics approach (where Newton's equations are applied to a lattice of ions and atoms using a suitable interatomic potential), the time-dependent ionizations of a group of atoms are tracked, and these time-dependent ionicities are then used in a molecular dynamics simulation, which then provides atomic coordinates and ionicities for a time-dependent diffraction calculation. These two approaches have been compared (Caleman *et al* 2011b), using a Gaussian profile for the x-ray pulse, and found to give reasonable agreement. Both approaches suggest that, after 80 fs, about 20% of the atoms have been ionized, with significant strains appearing in the outer layers, producing detectable changes in the Bragg condition. At longer times after the pulse, temperatures can rise to tens of electron volts (perhaps 500 000 K) in the Warm Dense Matter region. Nuclear displacement, which mainly occurs after the incident x-ray pulse has terminated, occurs as a result of the photoelectron cascade, which follows ionization of a fraction of the atoms. (The Debye period for nuclear vibration is about 100 fs.) For a small biomolecule, photoelectrons will escape, while slower Auger electrons at much lower energy will be mainly responsible for secondary electron generation. Photoelectrons generated by hard x-rays travel at about 50 nm fs^{-1} . At 8 keV, where absorption is greatly reduced compared to 2 keV (but also elastic scattering is weaker), the ratio of elastically scattered photons to those annihilated by photoionization is 1/32 for oxygen, 1/26 for nitrogen and 1/20 carbon. The resulting loss of atomic electrons decreases scattering power by 17% for carbon, 14% for nitrogen and 12% for oxygen. The resulting highly charged particle may then undergo a Coulomb explosion, as first analysed in this context by Jurek *et al* (2004). The idea that radiation damage may be reduced by photoelectrons escaping into vacuum from very small crystals, rather than into surrounding bulk material, has been discussed previously in the MX literature (Nave and Hill 2005). For nanocrystals larger than the electron inelastic mean free path, however, photoelectrons will thermalize, leading to vaporization at very high temperatures. It has also been suggested that the initial plasma of electrons excited in a sample may constitute a spatially uniform charge density, which thus has little effect on the diffraction pattern. Differing aspects of the problem are emphasized by the different theoretical approaches, from photoionization and electron level occupancies at the single-atom level during the pulse, to molecular dynamics, or the tracking, perhaps by Monte Carlo methods, of the secondary electron multiple inelastic scattering, based on energy-loss functions. An approach which estimates the effect of electronic damage on loss of coherence in the scattered radiation (because the electronic structure is in motion) has been given by Quiney and Nugent (2011), which allows correction of this effect and the extraction of additional sample information.

Simulations show that a positively charged surface layer is formed, which peels off as the sample is destroyed from

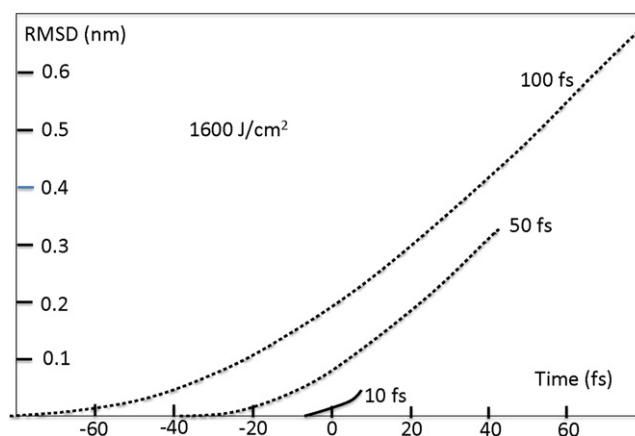


Figure 14. Simulations for urea at 2 keV based on a continuum model ('Cretin' code), showing atomic displacement during an XFEL pulse for three pulse durations. The fluence corresponds to 10^{12} photons focused to a $10 \mu\text{m}^2$ spot.

the outside in. A sacrificial tamper may therefore be used to slow the explosion, and this effect has been demonstrated experimentally using the FLASH soft x-ray FEL (Hau-Riege *et al* 2010). For nanocrystals in a liquid jet, the surrounding jet liquid will act as such a tamper, and the ionized water will provide a ready supply of electrons to neutralize the positively charged core of the bioparticle, as shown in simulations (Caleman *et al* 2011b) for urea nanocrystals. Here the higher absorption of water relative to protein is noted at 2 keV. It is possible, however, that the ionization process is spatially inhomogeneous, leading to clusters of the more readily ionized heavier atoms. The 'spot-fading' effect at longer pulse durations has been simulated for urea crystals (Caleman *et al* 2011a). In figure 14 (Caleman *et al* 2011b) we show the simulated RMS atomic displacement as a function of time during a 2 keV pulse, for several incident photon fluences, based on the Cretin code. The time-structure of the x-ray pulse is taken to be Gaussian, centered on $t = 0$. Here ion motions of up to 0.5 nm are predicted to occur in less than 100 fs (Current experimental results suggest less motion using higher energy x-rays, due to reduced absorption and damage, as discussed below).

The idea that the Bragg diffraction process, as a result of these damage processes, may self-terminate before the end of the x-ray pulse, has also been explored (Barty *et al* 2011). As the short-range order in a nanocrystal is destroyed initially, one expects a rising diffuse background and reduction in the intensity of Bragg peaks. This might be modeled as a pulse-duration dependent temperature factor. Lomb *et al* (2011) suggest that the time-dependence within the pulse may create an inhomogeneous (e.g. site dependent) damage process. The process by which Bragg diffraction is 'gated' by damage has been modeled by Barty *et al* (2011), using the theory of thermal diffuse scattering, assuming that Bragg scattering terminates when the RMS atomic vibration amplitude $2\pi\sigma_T$ equals a corresponding crystallographic d -spacing $d_{hkl} = 1/q$. This gives a turn-off time of $T_{\text{off}} = (2\pi q\sigma(T))^{-2/3}T$ where $\sigma(T)$ is the RMS atomic displacement at the end of a pulse of duration T . This model was compared with the virtual powder patterns

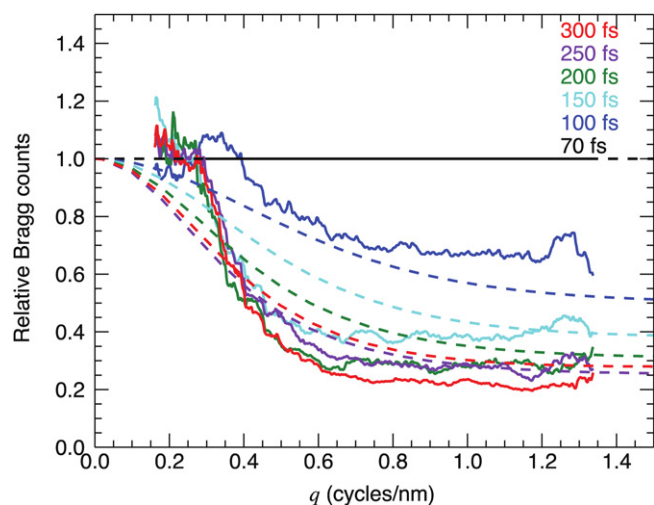


Figure 15. Simulations of spot-fading ('Cretin' hydrodynamic code, dashed lines) for PSI nanocrystals at 2 keV, compared with experiments using the liquid jet (continuous curves) for several pulse durations. The experimental data are the sum of Bragg spot patterns from 3792 nanocrystals in random orientations, making a 'virtual' powder pattern. The abscissa is given in units of $q = 1/d(hkl)$ in nm^{-1} . Bragg peaks may disappear before the end of the x-ray pulse due to loss of translational symmetry. Reproduced with permission from Barty *et al* (2011). Copyright 2011 Nature Publishing Group.

for several pulse durations, and compared with simulations using the Cretin hydrodynamic plasma model, as shown in figure 15. At 8 keV, the model predicts self-termination of Bragg diffraction after 20 fs for 0.3 nm resolution with micrometer-sized crystals. The total Bragg-scattered signal is proportional to the pulse irradiance I_0 times the turn-off time. The Cretin model shows that $\sigma^2(T)$ is proportional to I_0 , so the signal varies as $I_0 T_{\text{off}} \sim I_0^{2/3}$. That is, increasing the pulse fluence by proportionately extending the pulse duration (that is, keeping the pulse irradiance constant) does not have any effect on the Bragg signal. However, increasing the fluence by increasing the number of photons and maintaining the pulse duration causes the Bragg signal to turn off sooner. Nevertheless, the reduction in time does not overwhelm the increase in irradiance. This implies that the initial intensity of the XFEL pulse may be more important than the duration of the pulses. The analysis is also valid for the case of single-particle diffraction.

7. Time-resolved nanocrystallography

Time-resolved protein crystallography has been developed extensively at synchrotrons (Coppens *et al* 2009, Graber *et al* 2010) and requires a fast optical trigger to initiate a molecular reaction in synchrony with a later x-ray pulse. In this 'pump-probe' mode, the small differences between the structure factors of the optically illuminated ('pumped') sample and those of the unpumped sample are sought. A controllable delay is introduced between optical excitation and x-ray snapshot. The Laue mode (large bandwidth) is preferred, since it provides more Bragg spots per shot, and assists with the angular integration of partial reflections needed to extract structure factors. Reversible reactions are normally studied—the study

of irreversible reactions is only possible if the crystalline state survives for longer than the excitation. Recent work at LCLS seeks to take advantage of sub-picosecond time resolution to study the early stages of biochemical oxidation reactions, to study the microsecond timescale important for structural biology, or to take advantage of the reduced radiation damage possible using very short pulses. Since broadband radiation is not available at an XFEL, the normal Laue method cannot be used, (although a chirped beam with 1.5% bandwidth may be possible, and the SwissFEL design aims for 4%) and the small intensity changes sought due to pumping may require independent monitoring of the incident fluence and energy spread per shot for normalization. The fluence can be attenuated to a value below the damage threshold (or the beam defocused), so that data might be collected from large, single-crystal fixed samples mounted in a goniometer in a geometry which matches the absorption length of the pumping light to the path length of the x-ray beam in the sample. However, the sample must then be scanned to allow collection of pumped and unpumped diffraction patterns at 120 Hz. In addition, damage from one illuminated region may affect the next, and the fast goniometer motions and pump laser pulses must somehow be synchronized with the XFEL pulses. If the diffract-before-destroy method is used, a continuous supply of fresh sample is needed, irreversible reactions can be studied, and the Monte Carlo (or post-refinement) method may be used to integrate complete reflections, resulting in a simpler experimental arrangement. However, this requires the installation of a pump laser into the liquid injector shroud (the housing surrounding the jet), as shown in figures 16 and 1. Full details of a suitable hydrated bioparticle injector with pump-laser facilities for use with an XFEL are given in Weierstall *et al* (2012). Note that the pump laser has broad focus, so that particles hit by the XFEL pulse, will, at an earlier time, have been illuminated by the pump laser as they travel at perhaps 10 m s^{-1} in the liquid stream. An optimum interleaving of pump laser, delays and XFEL pulses can then be arranged to minimize effects of particle settling in the delivery tubes (where flow velocity is low) and variations in the XFEL fluence. The data are then collected successively for each delay, followed by a diffraction pattern with negative delay, in which the pump laser occurs after the XFEL pulse, to provide otherwise identical conditions for the ground (dark) state structure. This sequence of pump-probe delays is then repeated until sufficient accuracy is obtained. For each delay between pump and probe, sufficient data must be collected to provide structure factors for a three-dimensional reconstruction from the randomly oriented stream of nanocrystals, while each such data set for one delay provides one frame of a molecular movie. In many cases a starting model for the ground state may exist in the PDB, sufficient for MR phasing.

Figure 17 shows the first results of this approach for samples of Photosystem 1-ferredoxin (Aquila *et al* 2012). An electron transfer reaction following illumination at 532 nm results in undocking of the ferredoxin molecule from PSI on a timescale of tens of microseconds. Nanocrystals were grown overnight at the beamline using the batch method with $35 \mu\text{M}$ P700 and $38.5 \mu\text{M}$ ferredoxin in the presence of 25%

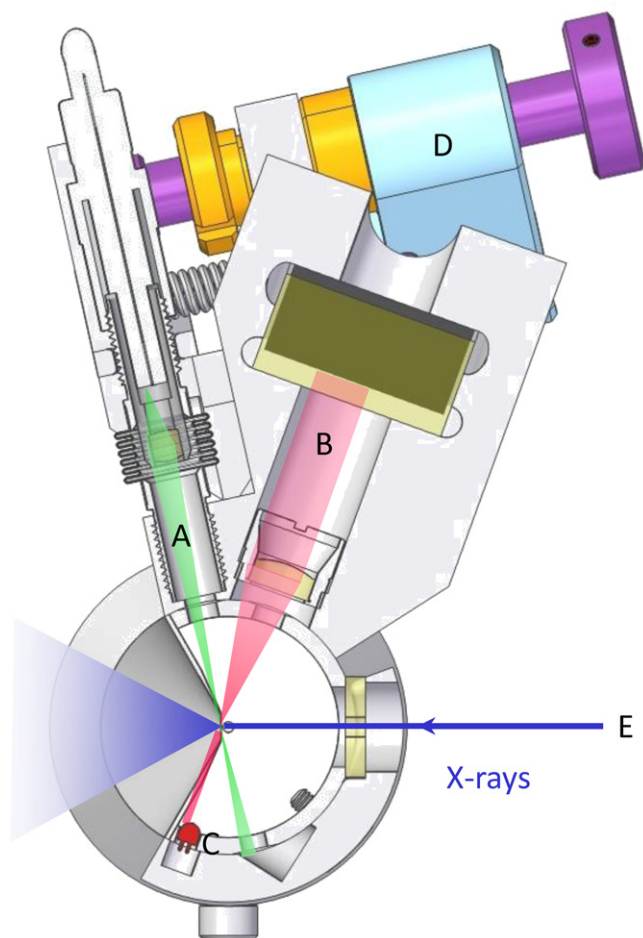


Figure 16. View along the liquid jet axis showing pump laser light (A), optical microscope optic axis (B), LED back-illumination for microscope (C) piezo motion to traverse pump laser focus across liquid stream (D), and XFEL beam (E). Reproduced with permission from Weierstall *et al* (2012). Copyright 2012 American Institute of Physics.

PEG 400, 100mM HEPES at pH 7.5, and 150mM CaCl_2 in complete darkness at 20 °C. Light excitation of PSI leads to charge separation in which its primary donor chlorophyll, P700, is excited and an electron transferred by a chain of electron carriers across the membrane over a distance of 60 Å. After ferredoxin, the soluble electron transfer protein, docks to the acceptor site of PSI, it is reduced in two steps, with time constants of 500 ns and 2 μs (Setif 2001, Ptushenko *et al* 2008) as determined by time-resolved optical spectroscopy and is followed by undocking of reduced ferredoxin from Photosystem I. Ferredoxin leaves its binding site to convey the electron to FNR (ferredoxin-NADP⁺ reductase), which finally reduces NADP⁺ to NADPH. This electron transfer reaction is one of the key steps in oxygenic photosynthesis. It has been shown previously by EPR spectroscopy that the PSI-fd co-crystals contain a photoactive complex of PSI with ferredoxin (Fromme *et al* 2002) and crystals dissolve in less than a second upon excitation with visible light at RT. Figure 17 shows the differences between the virtual powder diffracted intensities from the dark (unexcited) protein nanocrystals and those illuminated by the 532 nm pump laser. In these

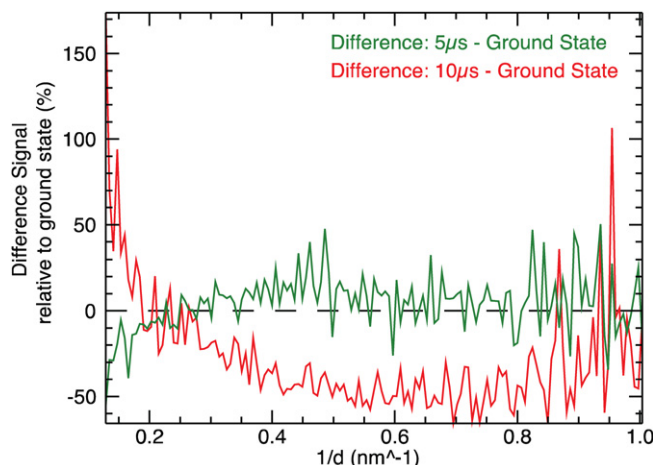


Figure 17. First results from time-resolved nanocrystallography, using 2 keV snapshot x-ray diffraction and a liquid jet in the ‘diffract-and-destroy’ mode. The difference between virtual powder patterns from PSI-ferredoxin co-crystals in their illuminated excited state and their dark ground state is shown. The excited state snapshots were recorded either 5 or 10 μs after brief illumination by a 532 nm pump laser, mimicking the effect of sunlight falling on plants. Reproduced with permission from Aquila *et al* (2012). Copyright 2012 The Optical Society.

preliminary results for this ‘pump–probe in a jet’ method, the differences clearly exceed the noise, and any common effects due to radiation damage are subtracted out. Higher resolution phased data will be needed to determine the cause of the difference, which may be due to a partial ordering of the microcrystals, followed by disordering, probably due to ferredoxin motion; however, temperature changes in the sample must also be considered. The data may be scaled by minimizing the mean-square difference between the bright and dark data sets at low resolution, or over the entire resolution range. Here the liquid jet diameter was about 4 μm , the x-ray focus spot size about 7 μm squared, and the 10 ns pump laser focus diameter was about 400 μm . The x-ray beam energy of 2 keV limited resolution to about 0.8 nm, and 70 fs pulses were used. Delays between a few microseconds and 30 μs can be accommodated with this arrangement—longer delays may be dealt with by focusing the pump laser inside the liquid delivery tube upstream of the nozzle. Higher resolution data are currently under analysis.

8. Snapshot SAXS

Small- (and wide-) angle x-ray scattering (SAXS and WAXS) from identical, randomly oriented molecules in solution can be expected to provide additional information if the ‘exposure time’ is much less than the rotational diffusion time of the molecules, so that they do not rotate during the exposure. Kam (1977) pointed this out, and showed how SAXS patterns from molecules frozen in either space or time become anisotropic, yielding two-dimensional diffraction rather than the one-dimensional data normally obtained from SAXS. He also showed how *ab initio* inversion of these data to real-space images may be achieved, without modeling. Kam showed

how the use of angular autocorrelation functions can be used to unscramble orientational disorder.

These two-dimensional correlated fluctuations ('speckles in SAX') have been seen in recent experiments. We are not concerned here with coherent interparticle interference, which is assumed to be suppressed. In one case the scattering anisotropy was used to extract hidden symmetries in the sample (Wochner *et al* 2009), in another it was used to reconstruct a two-dimensional real-space image of a typical particle from experimental data collected from many randomly oriented gold nanorods lying on their side on a silicon nitride membrane (Saldin *et al* 2011a). In this sense the method of Kam unscrambles rotational disorder, and this experimental paper also indicates that the method works for mixtures of different types of particles. (Some gold balls were accidentally included in the gold rod solution, and are seen reconstructed correctly.) Detailed simulations showing reconstruction of membrane proteins in solution in two dimensions is given in Saldin (2010), and an application of the method to virus particles in three dimensions with icosahedral symmetry, which greatly reduced computational demands, is given in Saldin *et al* (2011b). Experimental data of this type, from molecules in solution, are currently being collected at LCLS. However, the application of this method to proteins randomly oriented in three dimensions (rather than these inorganic samples), where the refractive index difference between protein and buffer is small, will be extremely challenging.

The Kam theory, in very simplified form, shows that a sum of the angular autocorrelation functions of diffraction patterns from groups of identical randomly oriented particles converges to that of one particle. To recover a real-space image of a typical particle from this angular autocorrelation function of one particle, it is then necessary to solve the phase problem twice, once to recover the diffraction pattern intensity of a particle from the angular autocorrelation function, and again to recover the real-space image. The angular autocorrelation is formed from the autocorrelation function of the intensity around a ring in the diffraction pattern (distance around the ring plays the same role as time delay in the Hanbury–Brown and Twiss experiment). Cross-correlation functions between different rings are also used.

Related, less ambitious approaches, may however be more productive, such as the use of a molecule of known structure as a starting point. Deviations from this structure, such as an enzyme undergoing a reaction, might then be followed using correlated fluctuations in the snapshot SAXS patterns taken from the enzyme interacting with a substrate. If data are collected at different points along the liquid jet, temperature variation provides an additional experimental parameter (Schmidt *et al* 2010). Rather than attempting *ab initio* inversion, this approach, based on modeling, might also be based on a comparison of angular correlation functions of diffraction patterns from many molecules, with a similar quantity computed from models of the protein. The radiation damage mitigation effects of short pulses will apply here, with the resulting potential for higher dose; however, it remains to be seen how the resolution in this mode will compare with that in conventional SAX and WAX experiments, normally limited to perhaps 1 nm or worse.

An important result has recently been demonstrated for studies of this type. Since the wanted standard deviation in the correlated fluctuations is proportional to the number of molecules per shot, while the Poisson shot noise in the x-ray beam has a standard deviation also proportional to the number of molecule, their ratio, the signal-to-noise ratio, is independent of the number of molecules per shot (Kirian *et al* 2011b). Nevertheless, the signal-to-noise ratio increases as the square root of the number of shots. This suggests that this approach may not improve on methods, which use a single molecule per shot. However, an important consideration is the hit rate, which is 100% for any snapshot SAX approach, and, for experimental reasons, very much less for single-particle CDI with an x-ray probe approaching the size of a virus. Finally, it has been suggested (Saldin 2010) that unwanted interparticle interference effects will wash out when angular correlations are taken (see also Kirian (2011b)).

9. Sample preparation

For single-particle diffraction, the particles may be simply aerosolized from a volatile buffer such as ammonium acetate, and starting solutions similar to those used for Cryo-EM single-particle analysis might be used. Purification of the particles to provide a monodispersed solution is crucial, and post-selection of 'good' particles is clearly possible based on the quality of the diffraction patterns and their autocorrelation functions. Virus particles have been shown to remain infectious after injection using the gas-focused injector.

The intentional growth of monodispersed protein nanocrystals, rather than larger single crystals (Mueller *et al* 2007) is a new challenge for SFX and is under rapid development. Microcrystals have been grown and shown to be advantageous for MX at synchrotrons using x-ray microdiffraction (Reikel *et al* 2005). Crystal growers often observe a spontaneous 'shower of microcrystals' during growth, which may be ideal for SFX. A variety of methods have been used to date, including sonication or mechanical grinding of larger crystals (Von Dreele 2007), and cross-linking with glutaraldehyde to terminate growth in lysozyme (Lomb *et al* 2011). For photosystem I membrane proteins, the growth has been developed in mother liquid above a PEG solution, so that settling of microcrystals above a certain size into the PEG terminates growth. A filter is used in the delivery line to set an upper limit on crystal size—for full details of one growth protocol see Hunter *et al* (2011). New techniques, such as SONICC (Gualtieri *et al* 2011) have been developed which allow us to detect crystals at the submicrometer scale, and when this is done it was found recently that 30% of what were previously discarded as amorphous precipitates were actually good crystals. Dynamic light scattering instruments are also now being used to determine the particle size distribution in this particle size range, from molecules up to nanocrystals a few micrometers in width. Nanocrystals have also been grown *in vivo*, using cells infected with a virus containing a suitable genome to fabricate the required protein (Koopmann *et al* 2012). In this work the protein nanocrystals were extracted from the cells before being injected into the x-ray beam—it

remains to be seen if entire cells, containing nanocrystals, can be injected without the creation of excessive x-ray background. If a critical nucleus size exists, and if a filter is used, a size window might also be established. We note that the method of solid-state NMR requires similar nanocrystals, and has developed appropriate methods for their growth (Martin and Zilm 2003). An analysis of the diffuse scattering in the patterns can give information on growth process and strain fields (Robinson and Harder 2009).

Once proteins have been brought into the nucleation zone, the buffer can be diluted with high viscosity solution to slow or inhibit the further growth of the microcrystals. Since control of supersaturation is crucial, one could quench into a supersaturated region by fast temperature change, and after a given time, quench into a region of lower supersaturation where nucleation of new crystals does not occur, but the remaining material is just sufficient to allow growth to the desired size. If crystal size can be monitored (e.g. using dynamic light scattering and SONICC), then growth can be halted by sudden temperature change once the required size is reached. Data mining of growth conditions has also been used to assist understanding the membrane protein crystallization process. In one study (Asure *et al* 2006), supervised learning algorithms were used to model the data space of 480 standard conditions originating from Hampton Research. The results were scored according to whether lamellar phases, amorphous precipitates or crystals resulted. 87% of the samples produce lamellar phases, which may contain 'invisible' crystals. A study of the 'unsuccessful' growth conditions recorded on the web by the protein structure initiative may reveal just the conditions, which produce submicrometer crystallites, unresolved by optical microscopy.

Much depends on the reasons for which protein crystals stop growing. Many explanations for this 'overseeding' and terminated growth have been suggested (Gunton *et al* 2007), including exhaustion of the molecular supply, excessive defect formation, charging effects, or coating of crystallites by denatured proteins from the solution. A review of protein crystal growth mechanisms can be found in Vekilov (2004).

McPherson *et al* (2004) have described the use of atomic force microscopy for detailed studies of the protein crystal growth process with molecular resolution, revealing a fascinating variety of defects such as screw dislocations (which enable heterogeneous nucleation) and other defects which prevent growth of large crystals. Although mostly devoted to much larger crystallites than those used in SFX, their work makes clear that, despite the Ostwald ripening process, the mother liquor contains a high concentration of microcrystals which do not grow into macroscopic crystals for MX. In addition to the 'unsuccessful' growth solutions, new regions and paths within the phase diagram for growth may be sought which will produce a solution of submicrometer crystals from proteins which do not otherwise grow large single crystals at all. This would open up new thermodynamic pathways for the preparation of crystallographic samples.

Other new approaches being evaluated in the laboratory of Professor Fromme at Arizona State University are as follows. For proteins where 'salting out' or removing salt is needed

to induce crystallization, ultrafiltration under gas pressure, followed by fractional sedimentation may be used. Here the settling time is used to sort the crystals by size. A spin column technique has also been used, in which is a syringe filled with size exclusion gel which, when equilibrated with the final condition of salt, rapidly places the protein into the final salt solution. For 'salting in' proteins, where salt or other precipitants need to be added in, a high precipitant interface might be placed underneath the protein solution. This creates an interface where the crystals form. Once these crystals reach a certain size, dependent on precipitant density, they fall fully into the precipitant where they can no longer grow due to lack of protein in the solution.

In summary, the overall efficiency of solving protein structures can be greatly increased from its current value of a few percent, if a means can be found to use the micrometer-sized (or invisible, submicrometer) crystallites present in growth solutions (and seen in AFM studies) which are currently wasted. It remains to be seen whether each nanocrystal consists entirely of proteins in one distinct conformation (different from those in another nanocrystal), and, if so, if these can be sorted during data analysis.

10. Key issues—challenges and solutions

Papers in this Key Issues series aim to identify the crucial issues for a new field. These may be summarized as follows:

10.1. The optimum beam energy and identification of resolution-limiting factors

In general, in order to take advantage of the much stronger interaction of photons with matter at lower energy (proportional to λ^2 for CDI (Howells *et al* 2009)), one uses the lowest beam energy which will provide the required resolution. The number of photons per pulse scales inversely with the beam energy for a SASE mode XFEL. Overall, one therefore expects a λ^3 dependence for scattered intensity.

For nanocrystals, although atomic resolution has been obtained at 9 keV, more detailed comparisons of these results with those obtained by conventional MX remain to be completed, in extension of the analysis done by Barty *et al* (2011), in order to determine the longest pulse duration consistent with atomic resolution. (The number of photons per pulse is theoretically proportional to pulse duration for SASE mode XFEL (Pellegrini 2011).) The optimum pulse duration appears to be very roughly in the vicinity of 40–70 fs at present. For single particles, resolution is currently not limited by the x-ray wavelength or pulse duration, but by structural inhomogeneity in the merged data, by background noise which negates the effect of long data accumulations, and by the limited dynamic range of detectors. It is also limited by the transition to 'incoherent' scattering, which occurs at high angles, for scattering from a large number N of atoms with near-random atomic coordinates, as the intensity changes from an N^2 dependence at low angles to an N dependence at high angles. A partial solution to the problem of weak protein scattering is to use a beam energy between the oxygen and

carbon K-shell ionization edges (the water window), where the difference in scattering strength between water and protein (and between organelles and cytoplasm) is greatest (Falcone *et al* 2011).

The most straightforward methods to improve resolution for single-particle XFEL imaging are therefore to use a more powerful XFEL and a more tightly focused beam. For a terawatt self-seeding machine with short-period superconducting helical undulator, working in the water window, the number of scattered photons may increase over the current number at the LCLS by about a factor of 100 (Geloni 2010; Pellegrini 2011). The number of incident photons expected in a 10 fs pulse at 8 keV is 7×10^{13} , compared with 2×10^{12} photons at present, and the shorter pulses allow more of the pulse to actually be scattered into Bragg peaks. While it is planned to reduce the focused beam diameter at the LCLS down to $0.1 \mu\text{m}$ (which would provide a 100-fold increase in scattered intensity over the present $1 \mu\text{m}$ focus), the resulting reduction in hit-rate due to errors in beam pointing and injector instabilities must be considered. A variety of feedback schemes are under consideration, in which, for example, a pair of photodetectors placed above and below the direct beam within the central gap in the detectors might be used to supply signals, which correct the vertical position of the injected particles. If a high hit rate can be maintained, a total increase by a factor of 10^4 in scattered photons can be expected for a terawatt FEL with $0.1 \mu\text{m}$ focus, in the water window, over the current LCLS with $1 \mu\text{m}$ focus, leading to significant resolution increase for single particles.

It might be argued that, since the smallest nanocrystals which have produced useful diffraction patterns so far have about ten molecules on a side, an increase in XFEL fluence of about 1000 should allow imaging of one molecule. This could be made up by a reduction in beam diameter by a factor of ten, together with an increase of ten in the power of the XFEL. Background reduction will also be needed.

Sample thickness must be limited in the soft-x-ray region. For thicknesses of protein above about $1 \mu\text{m}$, multiple elastic scattering of soft-x-rays starts to create great difficulties for inversion of diffraction patterns to density maps; however, if data are collected at two wavelengths, multiple scattering artifacts can be completely removed in certain cases (Spence 2009). Bergh *et al* (2008) conclude that highest resolution with single particles will be obtained at about 1.2 keV when taking into account the reduced absorption and hence slower damage processes at higher energy; however, additional instrumental factors such as x-ray focusing optics and detector efficiency should be considered. For particles in droplets, soft x-ray reflection from the surface of the droplet has been observed (Shapiro *et al* 2008).

10.2. Sample preparation

While single-particle methods are straightforward, much research along the lines suggested in section 9 above is needed into methods for intentional growth of monodispersed protein nanocrystals. For snapshot SAX experiments or those using a mixing jet, it will be important to monitor the temperature

along the jet. This could perhaps be done using Raman scattering, or by measuring the lattice constants of silicon nanoparticles in the jet with the XFEL radiation.

10.3. Injector development

More efficient injectors are needed which waste less of the precious sample, such as those suitable for delivering protein nanocrystals in the more viscous lipidic cubic or sponge phase. Gas-phase injectors for single particles are needed which are capable of a sustained high hit rate in the tens of percent range. This will require advances in aerodynamic gas-focusing techniques, or the development of new gas nozzles, which allow the x-ray beam to be focused very close to a small nozzle, where the particle density is high. The amount of protein needed for SFX will be reduced both by advances in data analysis software which will require much less data, and by fluid switching which, by shutting off the jet between x-ray shots, will eliminate most protein which otherwise runs to waste between shots. Either this switching rate or detector readout, will then provide the bottleneck for data collection in future. Methods are under development for continuous delivery of disk-shaped membrane proteins between lipid layers, using sheet jets, oriented normal to the beam.

For single-particle diffraction using the gas-phase injector, the most urgent problem apart from hit rate, is the control and reproducibility of the thickness of the remaining water jacket around each particle. Unwanted salt concentration increase in the buffer due to evaporation is also an issue, since this may cause conformational changes and would reduce further the weak x-ray contrast between protein and buffer. The reconstructed images from small virus diffraction patterns sometimes produce reconstructed images of particles which are much too large, suggesting that the liquid buffer shell has not fully evaporated (the particle size can be evaluated directly from the autocorrelation function, i.e. the Fourier transform of the diffraction intensity), while larger particles, such as the Mimi virus, are reconstructed with the correct size. For single-particle imaging in the continuous liquid column of the water jet, the main problem is background water scattering and the diffraction streak from the column. A reduction of the jet diameter from $3 \mu\text{m}$ to $0.3 \mu\text{m}$ (demonstrated in DePonte *et al* (2011)) reduces the water background by a factor of 100.

10.4. Time-resolved nanocrystallography

While pump-probe experiments on single-particles may have to await the advent of more powerful XFELs and engineering advances to improve the hit-rate, the preliminary results shown here for time-resolved nanocrystallography in a liquid jet are a promising first step. Higher resolution data are currently being collected using this method, which, when combined with the post-processing methods outlined in section 7, should have the advantage of reduced radiation damage at larger x-ray dose and should allow the use of nanocrystals for time-resolved crystallography. Alternative geometries and timing sequences may be developed, such as the use of two consecutive x-ray pulses (each below the damage threshold) applied to the same nanocrystal in a slow viscous jet, where the nanocrystal was

laser pumped for the first pulse and not pumped for the second pulse. The coherent convergent beam method, by providing limited angular integration of intensities, might provide some of the benefits of the Laue method. For studies of molecular dynamics, the longest delay between pump and probe is currently limited to the length of the collimated jet divided by the jet velocity, which results to about 30 μs . This might be increased through the use of slower, more viscous jets, or by focusing the pump laser inside the nozzle.

10.5. Software developments and phasing

Software developments will likely be the first and most dramatic improvements to be seen, as more efficient algorithms which require much less data and hence beamtime are developed. They may include the use of models for peak profiles, fast MAD phasing, and direct phasing methods based on iterative phasing of shape transforms.

Most important is the need for more efficient hit-finder routines. Efficient automated indexing algorithms that can deal with simultaneous multiple particle hits are also needed. These algorithms may perhaps be based on compressive sensing methods and may be provided with *a priori* information on cell constants and space groups. In addition, algorithms that merge data from partial reflections from different nanocrystals but avoid the inadvertent application of a twin operation are badly needed.

In single-particle data analysis, there is much room for improvement. An iterative phasing algorithm, which converges to a unique solution for a sample with complex charge density and unknown support, for which data are missing at very low angles (due to the beamstop), does not exist. However, the use of multiple stacked detectors each receiving the scattering from the hole in the center of the previous detector addresses the ‘missing data’ problem, while the Shrinkwrap and Charge-flipping algorithms have proven successful in many cases. Experience shows that the method of ‘fragment completion’ is highly effective—if part of the molecular density is known (such as a gold ball label), the reinforcement of this density during iteration greatly aids convergence, and we have noted the extension of the MAD method to single-particle CXI (Scherz *et al* 2008). (The Ptychography method solves the problems of unknown support and complex density but requires multiple data sets from the same particle in the same orientation.) The list of convex constraints for these algorithms is slowly growing. The orientation-determination problem appears solved in principle; in practice only if enough data of sufficiently high quality can be obtained.

As in the field of single-particle cryo-electron microscopy, it might be thought that the final resolution in three dimensions can be indefinitely improved simply by the accumulation of more two-dimensional data. Experience in that field, as in this, however, has shown that the need to collect sufficient data from each particle to allow accurate orientation determination can be limiting. Even for identical particles, the alignment process then depends on the resolution of individual patterns (Henderson 1995).

10.6. Detectors

The greatest need here is for increased dynamic range, so that intense low order scattering can be recorded along with the crucial high-resolution scattering whose intensity falls off approximately as the inverse fourth power of scattering angle. Reduced background electronic and readout noise is also desirable, as are more pixels. These flat subpanels might be arranged on a spherical surface around the sample, allowing data collection to much higher angles, and hence resolution beyond the wavelength, at least for nanocrystals.

The European XFEL, which will be operational in 2015, will vastly increase the pulse repetition rate, pushing detector technology to extremes. Trains of 2700 pulses, spaced by 200 ns, will be fired 10 times per second, for a potential of 27 000 patterns recorded per second. A new detector is being developed that will have the ability to store about hundreds of patterns in each of the bursts: perhaps, 5000 patterns per second. The detector is designed to allow a just-detected pattern to overwrite one of the storage buffers, meaning that decisions could be made to store the best 5000 patterns in the train. This would require some other veto detector, based on fluorescence signal, for example, since the patterns themselves cannot be evaluated until after they are read out. The advantages of on-chip pre-processing for the snapshot WAXS methods which follow are clear.

10.7. Snapshot solution scattering (fast WAXS)

This field is in its infancy, with the first experimental observations of anisotropic snapshot SAXS patterns only recently published. Direct *ab initio* inversion of these patterns using the Kam theory and its developments is the subject of several papers submitted at the time of this review. (Inversion using the Kam theory for two-dimensional inorganic particles is demonstrated experimentally in Saldin *et al* (2011a)). Difficulties with full three-dimensional reconstruction from two-dimensional data are discussed in (Elser 2011). Signal-to-noise appears to be independent of the number of particles in the beam (Kirian 2011b), in which case the method offers no advantages over single-particle-per-shot methods other than the 100% hit rate, which makes the method attractive for dynamic studies of chemical reactions. The possibility of using a symmetry-adapted set of basis functions helps greatly in the case of icosahedral viruses (Saldin *et al* 2011b). In analogy to the reconstruction of three-dimensional scenes from two-dimensional holograms, one faces a trade-off between dimensionality and resolution. The most exciting immediate prospect then is experiments based on proteins (e.g. an enzyme) of known structure undergoing reactions along a liquid jet. Here changes in the angular correlation function might be tracked, and compared with molecular dynamics simulations.

In summary the main problems for single particle work are the limited number of photons per shot, limited detector dynamic range, missing data at low angles, solvent jacket thickness, and conformational variability of particles. If these problems are solved, the method will produce the higher quality data needed for the reconstruction algorithms. The hit-rate problem for single particles may be addressed through

the use of fast WAXS data, where hit rate is 100%; however, this may come at some cost in resolution—the use of few-particles-per-shot may be useful. For nanocrystals, the main problems are detector dynamic range, the large quantity of protein currently needed, the long time (now months) needed for data analysis, and the need for data analysis methods which, when merging partial reflections, do not impose twinning. (The second and third of these problems will probably be resolved shortly.) For time-resolved diffraction, the main challenges are the accuracy of the Monte Carlo or post-refinement method when applied to nanocrystals, and the limited length of pump–probe delay. None of these problems are fundamental, and all should be amenable to solution by the methods of biochemistry and applied physics.

The atmosphere at a recent bio-XFEL conference held at LBNL in January 2011 (see <https://sites.google.com/a/lbl.gov/biology-with-fels/> for presentations) reminded many attendees of the first STM conferences in the mid nineteen eighties. Breakthroughs in science frequently follow the invention of new sources or detectors, or new arrangements of them. As has often been said, ‘To see something new you must do something new’. The rich proliferation of ideas for new experiments where x-ray lasers are applied to solve structural biology problems, and the highly creative approach of the many collaborating international groups, give all of us great confidence in the future of this exciting new field.

Acknowledgments

We are grateful to Drs Fillipe Maia and S Marchesini for many useful discussions. Work supported by NSF awards MCB-1021557, MCB-1120997 and HFSP 024940

References

- Aquila A *et al* 2012 *Opt. Express* **20** 2706
- Asure S, Raman P, Otey M and Parthasarathy S 2006 *Bioinformatics* **22** 40
- Bartells L S 1986 *Chem. Rev.* **86** 491
- Barty A *et al* 2011 *Nature Photon.* **6** 35
- Beerlink A, Wilbrandt P, Zeigler E, Carbone D, Metzger T and Salditt T 2008 *Langmuir* **24** 4952
- Bergh M, Huld G, Timneanu N, Maia F and Hajdu J 2008 *Quart. Rev. Biophys.* **41** 181
- Bevington P and Robinson D K 2002 *Data Reduction and Error Analysis for the Physical Sciences* 3rd edn (New York: McGraw-Hill)
- Bogan M J, Starodub D, Hampton C Y and Sierra R G 2010 *J. Phys. B: At. Mol. Opt. Phys.* **43** 194013
- Bohren C F and Huffman D R 1983 *Absorption and Scattering of Light by Small Particles* (New York: Wiley)
- Bortel G and Tegze M 2011 *Acta Crystallogr. A* **67** 533
- Boutet S and Williams G 2010 *New J. Phys.* **12** 035024
- Boutet S *et al* 2012 *Science* **337** 362
- Breedlove J and Trammel G 1970 *Science* **170** 1310
- Caleman C, Huld G, Maia F, Ortiz C, Parak F, Hadju J, van der Spoel D, Chapman H and Timneanu N 2011a *ACA Nano* **5** 139
- Caleman C, Bergh M, Scott H A, Spence J C H, Chapman H and Timneanu N 2011b *J. Mod. Opt.* **58** 1486
- Cha W, Song S, Jeong N, Harder R, Yoon K, Robinson I and Kim H 2010 *New J. Phys.* **12** 035022
- Chapman, H N *et al* 2006a *Nature Phys.* **2** 839
- Chapman H N *et al* 2006b *J. Opt. Soc. Am.* **23** 1179
- Chapman H N *et al* 2007 *Nature* **448** 676
- Chapman H N *et al* 2011 *Nature* **470** 73
- Chen J, Millane R and Spence J C H 2012 in preparation
- Coppens P *et al* 2009 *J. Synchrotron Radiat.* **16** 226
- DePonte D 2012 private communication
- Dilanian R, Streltsov V, Quiney H and Nugent K 2012 *Acta Crystallogr. A* submitted
- DePonte D, McKeown J, Weierstall U, Doak R and Spence J C H 2011 *Ultramicroscopy* **111** 824
- DePonte D P, Weierstall U, Schmidt K, Warner J, Starodub D, Spence J C H and Doak R B 2008 *J. Phys. D: Appl. Phys.* **41** 195505
- Doerschuk P C and Johnson J E 2000 *IEEE Trans. Inform. Theory* **46** 1714
- Doniach S 1996 *J. Synchrotron Radiat.* **3** 260
- Elser V 2011 *Ultramicroscopy* **111** 788
- Emma P, Akre R, Arthur J, Bionta R, Bostedt C, Bozek J, Brachmann A, Bucksbaum P, Coffee R and Decker F-J 2010 *Nature Photon.* **4** 641
- Falcone R, Jacobsen C, Kirz J, Marchesini S, Shapiro D and Spence J 2011 *Contemp. Phys.* **52** 293
- Faubel M, Schlemmer S and Toennies J P 1988 *Z. Phys. D* **10** 269
- Fienup J R 1982 *Appl. Opt.* **21** 2758
- Fienup J R 1987 *J. Opt. Soc. Am. A* **4** 118
- Frank J 2006 *Three-dimensional Electron Microscopy of Macromolecular Assemblies* (New York: Oxford University Press)
- Frank J 2011 *Molecular Machines in Biology* (New York: Oxford University Press)
- Fraser J S, van den Bedem H, Samelson A J, Lang P T, Holton J M, Echols N and Alber T 2010 *Proc. Natl Acad. Sci. USA* **108** 16247
- Fromme P and Spence J C H 2011 *Curr. Opin. Struct. Biol.* **21** 509
- Fromme P, Bottin H, Krauss N and Setif A 2002 *Biophys. J.* **83** 1760
- Fung R, Shneerson V, Saldin D and Ourmazd A 2009 *Nature Phys.* **5** 64
- Geloni G, Saldin E and Samoylova L 2010 *New J. Phys.* **12** 035021
- Gerchberg R and Saxton W 1972 *Optik* **35** 237
- Giannakis D, Schwander P and Ourmazd A 2012 *Opt. Express* **20** 12799
- Graber T *et al* 2010 *J. Synchrotron Radiat.* **18** 658
- Gualtieri E, Guo F, Kissick D, Jose J, Kohn R, Jian W and Simpson G 2011 *Biophys. J.* **100** 207
- Gunton J, Shirayev A and Pagan D 2007 *Protein Condensation* (New York: Cambridge University Press)
- Hau-Riege S P 2011 *High-intensity X-ray Interactions with Matter* (New York: Wiley)
- Hau-Riege S P *et al* 2010 *Phys. Rev. Lett.* **104** 064801
- Henderson R 1995 *Quart. Rev. Biophys.* **28** 171
- Howells M *et al* 2009 *J. Electron Spectrosc. Relat. Phenom.* **170** 4
- Huld G, Szoke A and Hajdu J 2003 *J. Struct. Biol.* **144** 219
- Hunter M *et al* 2011 *Biophys. J.* **100** 198
- Hunter M *et al* 2011 *J. Phys. D: Appl. Phys.* **41** 195505
- Johansson L *et al* 2012 *Nature Methods* **9** 265
- Jurek Z, Faigel G and Tegze M 2004 *Eur. Phys. J. D* **29** 217
- Kam Z 1977 *Macromolecules* **10** 927
- Kirian R A, Wang X, Weierstall U, Schmidt K E, Spence J C H, Hunter M, Fromme P, White T, Chapman H N and Holton J 2010 *Opt. Express* **18** 5713
- Kirian R A *et al* 2011a *Acta Crystallogr. A* **67** 131
- Kirian R A, Schmidt K, Wang X, Doak R B and Spence J C H 2011b *Phys. Rev. E* **84** 011921
- Koopmann R *et al* 2012 *Nature Methods* **9** 259
- Leslie A G W 2006 *Acta Crystallogr. D* **62** 48
- Loh N D and Elser V 2009 *Phys. Rev. E* **80** 026705
- Loh N D *et al* 2010 *Phys. Rev. Lett.* **104** 225501
- Lomb L *et al* 2011 *Phys. Rev. B* **84** 214111

- Maia F, Ekeberg T, van der Spoel D and Hajdu J 2010 *J. Appl. Crystallogr.* **43** 1535
- Maia F, Yang C and Marchesini S. 2011 *Ultramicroscopy* **111** 807
- Marchesini S, He H, Chapman H N, Hau-Riege S P, Noy, Howells M, Weierstall U and Spence J C H 2003 *Phys. Rev. B* **68** 140101
- Marchesini S 2007 *Rev. Sci. Instrum.* **78** 011301
- Margaritondo and Ribic 2011 *J. Synchrotron Radiat.* **18** 101
- Martin R W and Zilm K W 2003 *J. Magn. Res.* **165** 162
- McPherson A, Kuznetsov Y, Malkin A and Plomp M 2004 *J. Synchrotron Radiat.* **11** 21
- Miao J, Charalambous C, Kirz J and Sayre D 1999 *Nature* **400** 342
- Millane R J 1990 *Opt. Soc. Am.* **7** 394
- Moths B and Ourmazd A 2011 *Acta Crystallogr. A* **67** 481
- Mueller M, Jenni S and Ban N 2007 *Curr. Opin. Struct. Biol.* **17** 572
- Nave C and Hill M A 2005 *J. Synchrotron Radiat.* **12** 299
- Nelson J, Huang X, Steinbrener J, Shapiro D, Kirz J, Marchesini S, Neiman A, Turner J and Jacobsen C 2010 *Proc. Natl Acad. Sci. USA* **107** 7235
- Neutze R, Wouts R, van der Spoel D, Weckert E and Hajdu J 2000 *Nature* **406** 752
- Owen R L, Rudino-Pinera E and Garmen E F 2006 *Proc. Natl Acad. Sci. USA* **103** 4912
- Park H, Qiu X, Rhoades E, Koriach J, Kwok L, Zipfel W, Webb W and Pollack L 2006 *Anal. Chem.* **78** 4465
- Pellegrini C 2011 Personal communication
- Perutz M 1954 *Proc. R. Soc. Lond. A* **225** 264
- Philipp H T, Koerner L, Hromalik M, Tate M and Gruner S 2010 *IEEE Trans. Nucl. Sci.* **57** 3796
- Philipp H T, Ayyer K, Tate M W, Elser V and Gruner S M 2012 *Opt. Express* **20** 13129
- Ptushenko V V, Cherepanov D, Krishtalik L and Semenov A 2008 *Photosynthesis Res.* **97** 55
- Quiney H M and Nugent K A 2011 *Nature Phys.* **7** 142
- Rasmussen S G F et al 2011 *Nature* **469** 175
- Rayleigh J 1879 *Proc. R. Soc. Lond. A* **29** 71
- Reikel C, Burghammer M and Schertler G 2005 *Curr. Opin. Struct. Biol.* **15** 556
- Robinson I K and Harder R 2009 *Nature Mater.* **8** 291
- Rodenburg J M 2008 *Advances in Imaging and Electron Physics* vol 150 ed P W Hawkes (Amsterdam: Elsevier) p 87
- Rossmann M G and van Beek C G 1999 *Acta Crystallogr. D* **55** 1631
- Rossmann M G, Leslie A G, Sherin S A and Tsukihara T 1979 *J. Appl. Crystallogr.* **12** 570
- Saldin D K 2010 *Phys. Rev. B* **81** 174105
- Saldin D K, Poon H C, Bogsan M J, Marchesini S, Shapiro D A, Kirian R A, Weierstall U and Spence J C H 2011a *Phys. Rev. Lett.* **106** 115501
- Saldin D K, Poon H C, Schwander P, Uddin M and Schmidt M 2011b *Opt. Express* **19** 17318
- Sayer D 1952 *Acta Crystallogr.* **5** 843
- Scherz A, Zhu D, Rick R, Schlotter W, Roy S, Luning J and Stohr J 2008 *Phys. Rev. Lett.* **101** 076101
- Schmidt K, Spence J, Weierstall U, Starodub D, Chapman H, Howells M and Doak B 2008 *Phys. Rev. Lett.* **101** 115507
- Schmidt M, Graber T, Henning R and Srajer V 2010 *Acta Crystallogr. A* **66** 198
- Schwander P, Giannakis D and Yoon C H 2012 *Opt. Express* **20** 12827
- Scott H A 2001 *J. Quant. Spectrosc. Radiat. Transfer* **71** 689
- Seibert M et al 2010 *J. Phys. B: At. Mol. Opt. Phys.* **43** 194015
- Seibert M M et al 2011 *Nature* **470** 78
- Setif P, 2001 *Biochim. Biophys. Acta* **1507** 161
- Shapiro D A, DePonte D, Doak R B, Fromme P, Hembree G, Hunter M, Marchesini S, Schmidt K, Weierstall U and Spence J C H 2008 *J. Synchrotron. Radiat* **15** 593
- Shneerson V L, Ourmazd A and Saldin D 2008 *Acta Crystallogr. A* **64** 303
- Sjores S H W 2010 *Advances in Protein Chemistry and Structural Biology* **81** 89
- Solem J C 1986 *J. Opt. Soc. Am. B* **3** 1551
- Son S-K, Young L and Santra R 2011 *Phys. Rev. A* **83** 033402
- Southworth-Davies R J, Medina M A, Carmichael I and Garmen E F 2007 *Structure* **15** 1531
- Spence J C H 2003 *High Resolution Electron Microscopy* 3rd edn (Oxford: Oxford University Press)
- Spence J C H 2008 *Diffraction (lensless) imaging Science of Microscopy* ed P W Hawkes and J C H Spence (New York: Springer) pp 1196–225 chapter 19
- Spence J C H 2009 *Acta Crystallogr. A* **65** 28
- Spence J C H and Cowley J M 1978 *Optik* **50** 129
- Spence J C H and Doak R B 2004 *Phys. Rev. Lett.* **92** 198102
- Spence J C H, Kirian R A, Wang X, Weierstall U, Schmidt K E, White T, Barty A, Chapman H N, Marchesini S and Holton J 2011 *Opt. Express* **19** 2866
- Spence J C H, Weierstall U and Howells M 2002 *Phil. Trans.* **360** 875
- Starodub D et al 2008 *J. Synchrotron Radiat.* **15** 62
- Struder L et al 2010 *Nucl. Instrum. Methods Phys. Res. A* **614** 483
- Tegze M and Bortel G 2012 *J. Struct. Biol.* **179** 41
- Vekilov P 2004 *Methods Mol. Biol.* **300** 15
- Von Dreele R B 2007 *J. Appl. Cryst.* **40** 133
- Vredenburg E and Luiten J 2011 *Nature Phys.* **7** 747
- Wang X, Kruijs F E and McMurry P H 2005 *Aerosol Sci. Technol.* **39** 611
- Weierstall U, Doak B, Spence J, Starodub D, Shapiro D, Kennedy P, Warner J, Hembree G, Fromme P and Chapman H 2008 *Exp. Fluids* **44** 675
- Weierstall U, Spence J C H and Doak R B 2012 *Rev. Sci. Instrum.* **83** 035108
- White T A, Kirian R, Martin A, Aquila A, Nass K and Chapman H 2012 *J. Appl. Cryst.* **45** 335
- Whitaker E 1953 *Acta Crystallogr.* **6** 222
- Wochner P, Gutt C, Autenreith T, Demmer T, Bugaev V, Ortiz A, Duri A, Zontone F, Greubel G and Dosch H 2009 *Proc. Natl Acad. Sci.* **106** 11511
- Yang C, Wang Z and Marchesini S 2010 *Proc. SPIE* **7800** 78000
- Ziaja B, Castro A, Weckert E and Moller T 2006 *Eur. Phys. J. D* **40** 465