Serial Femtosecond Crystallography
Henry N. Chapman\textsuperscript{1,2,3}

\textsuperscript{1}Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany
\textsuperscript{2}Department of Physics, University of Hamburg, Lurperchausee 149, 22607 Hamburg, Germany
\textsuperscript{3}Centre for Ultrafast Imaging, Luruper Chaussee 149, 22607 Hamburg, Germany

X-ray free-electron lasers produce brief flashes of X-rays that are of about a billion times higher peak brightness than achievable from storage ring sources. Such a tremendous jump in X-ray source capabilities, that came in 2009 when the Linac Coherent Light Source begun operations, was unprecedented in the history of X-ray science. Protein structure determination through the method of macromolecular crystallography has consistently benefited from the many increases in source performance from rotating anodes to all generations of synchrotron facilities. But when confronted with the prospects of such bright beams for structural biology, enthusiastic proposals were tempered by a trepidation of the effects of such beams on samples and challenges to record data [1]. A decade after these discussions (and others in USA) on the applications of X-ray FELs for biology, the first experiments took place at LCLS giving results that fulfilled many of the dreams of the early visionaries. In particular, the concept that diffraction representing the pristine object could be recorded before the X-ray pulse completely vaporises the object was validated [2], confirming predictions [3] that established dose limits could be vastly exceeded using femtosecond-duration pulses. The first experiments illuminated a path to achieve room-temperature structures free of radiation damage, from samples too small to provide useful data at synchrotron facilities, as well as providing the means to carry out time-resolved crystallography at femtoseconds to milliseconds. In the five years since, progress has been substantial and rapid, invigorating the field of macromolecular crystallography [4,5]. This phase of development is far from over, but with both the LCLS and the Spring-8 Ångström Compact Free-electron Laser (SACLA) providing facilities for measurements, the benefits of X-ray FELs are already being translated into new biological insights.

One of the main bottlenecks in protein structure determination is the need for large well-diffracting crystals of the protein of interest, which can take months or years of effort to achieve, if at all. Smaller crystals should be easier to obtain, given that showers of small crystals are often observed in crystallisation trials. For a fixed defect density, the likelihood to obtain a portion of well ordered crystals should increase as their volume decreases. The strength of a diffraction pattern of such an ideal crystal is proportional to the number of unit cells, and thus the volume, of the crystal. It is also proportional to the
X-ray fluence (photons per unit area) incident on the sample. However, given a small crystal, one cannot necessarily just expose it for longer to achieve a stronger and measurable signal. A small fraction of the incident X-ray photons scatter from the sample and contribute to the diffraction pattern, but a larger fraction are absorbed. These ionising photons cause changes to the very structure being examined. At cryogenic temperatures, where many X-ray radiolysis products are immobilised, the diffraction of macromolecular crystals drops by about 1 Å resolution for each 10 MGy of dose (i.e. about 30 MGy for 3 Å resolution) [6]. The X-ray dose is the energy absorbed per unit mass of material and, since the X-rays penetrate the entire volume of the crystal, the dose is also proportional to fluence, making diffraction dependent on dose. To give a sense of 1 MGy (= 1 J kg⁻¹), consider a small vial of water that uniformly receives this dose. Water has a heat capacity of 4800 J kg⁻¹ K⁻¹, so assuming that no heat or energetic particles flow out of the vial it would heat up by 10⁶ J kg⁻¹/(4800 J kg⁻¹ K⁻¹) = 208 K. It is remarkable that this amount of energy is transferred when only one in 50,000 carbon atoms in a protein absorbs an X-ray photon. Thirty times fewer atoms than this contribute to the diffraction pattern, which is why so many repeating units are needed in the crystal.

An X-ray FEL pulse of 30 fs duration can reach a power density of more than 10⁻¹⁷ W cm⁻² when focused to micrometer dimensions. No material can withstand this—the pulse turns a sample into a plasma and leaves a hole in a solid sample tens of micrometers larger than the beam size. Indeed, the dose to the sample can exceed 3 GGY [7,8], thousands of times higher than our example of heated water. Nevertheless, when a macromolecular crystal is placed in such a beam, Bragg spots can be observed to a resolution of 2 Å or better, if the perfection of the crystal allows. Since Bragg spots are created when scattered waves from the repeated units of a crystal constructively interfere, they are very sensitive to small average atomic displacements which would otherwise disturb that interference. A 2-Å Bragg spot is considerably reduced in strength by only 0.3 Å average atomic displacement, showing that this degree of order persisted for at least a large fraction of the pulse [8]. Lighter atoms move more quickly than heavier, due to lower inertia, and H atoms have been clocked at moving by bond distances in 12 fs [9], showing that the shorter the pulse the more pristine the diffraction information.

By taking diffraction measurements at over 1,000 times higher fluence than tolerable at conventional sources it is possible to use correspondingly smaller crystal volumes, but smaller volumes than this are possible (or higher resolution diffraction) by averaging weak diffraction over many crystals to build up signal above noise. Diffraction from crystals as small as 6 unit cells across has been observed [2]. This gives hope for one of the big dreams of using FEL pulses to overcome the need for crystals entirely [3]. In fact it was this vision that brought collaborators together to the LCLS in 2009 to explore...
these potentials and challenges in experiments on both virus particles [10] and nanocrystals [2]. Experiments were strongly influenced by single-particle diffraction experiments at the FLASH soft X-ray FEL at DESY, Hamburg, where a key hurdle was to prevent background scattering from swamping the signal from weakly-diffracting samples. An intense beam alone is not enough—both the background and signal increases with beam intensity—and progress was made in measuring diffraction from aerosol beams [11], and ideas to accumulate diffraction from laser-aligned particles in liquid drops [12,13]. For the LCLS experiments a breakthrough in the liquid jet technology, the gas-focusing nozzle, arrived just in time to solve the problem of crystals becoming clogged in small-bore capillaries. Fine micrometer-diameter liquid jets could instead be made by accelerating the liquid emerging from a larger diameter nozzle by the action of a coaxial gas flow [14], known as a gas virtual dynamic nozzle (GVDN). These innovations came together with new detectors and instrumentation [15,16] for crystallography experiments at LCLS where it became clear, only minutes after running the first sample—a suspension of photosystem I nano-crystals in their mother liquor—that this was a powerful new approach for macromolecular crystallography.

The crystal certainly does not rotate during the exposure—all motion is frozen in the flash exposure. And since the sample explodes, each crystal gives a single still diffraction pattern in whatever orientation it happened to have. Although needle-like crystals can flow-align in the jet, the distribution of orientations is usually random and isotropic giving, in aggregate, a complete coverage of three-dimensional reciprocal space as long as enough patterns from enough crystals can be collected. Detector frames are read at the rate of the incoming pulses of 120 per second, but not every X-ray pulse intersects a crystal. In a flowing jet or beam, the probability of striking a crystal (and hence recording diffraction from a crystal) is proportional to the density of crystals and the volume of the overlap of the focused X-ray beam with the jet. Strike proportions of 3 to 30% of pulses are common, so in 5 hours of collection it may be possible to amass 60,000 to 600,000 single-crystal patterns. This mode of data collection became to be known as serial femtosecond crystallography (or SFX).

The low solvent-scattering background of a 2 μm diameter liquid jet is key to measuring diffraction from sub-micrometer crystals but it comes with the drawback of high sample consumption. A typical jet flows at a speed of about 7 m/s or almost 6 cm in the 8.3 ms between pulses at the LCLS. In this time more than 10,000 crystals can flow by, making the efficiency of this method less than 0.01%. To sustain the jet for long enough to acquire 60,000 patterns for a full dataset thus requires more than 3 ml of concentrated solution of crystals. There are several ways to improve the efficiency: pulse the flow, slow down the flow, or to introduce samples into the beam on a movable thin support. Large crystals can be measured by adapting instrumentation used at synchrotron beamlines [17]. There is a considerable difference in the delivery methods depending on
whether the samples are measured in vacuum or at atmosphere. In-vacuum experiments achieve lowest backgrounds and allow for low photon energies. Running liquid jets reliably in vacuum is challenging, and until recently experiments were often plagued by freezing of nozzles and the formation of icicles, dependent on the device used to catch the spent material. Delivery methods also impact how time-resolved measurements can be carried out, with liquid systems allowing for mixing experiments in a continuous flow geometry. Photoexcitation of the sample in a pump-probe measurement requires the spacing of measurements to be larger than the width of the pump light pulse, but long delay times (of milliseconds) must account for flow of the sample.

One way to slow down the sample flow across the beam is to use a more viscous carrier medium. Lipidic cubic phase (LCP) has been successfully used as both the crystallisation and delivery medium for many integral membrane proteins, made possible using a device that extrudes this material from a small nozzle into vacuum [18]. The velocity of the extrusion can be easily varied and be made slow enough to strike everything that flows past. The speed is usually set to move about 100 μm between X-ray pulses to avoid any damage inflicted by the preceding pulse. The extrusion diameter is about 50 μm, so background is higher than liquid jets. This is counteracted by using membrane crystals that are usually larger than 5 μm width, providing strong diffraction above this background. Due to the dynamic range of current detectors, this requires attenuating the beam. LCP extrusion has been used for seven structures solved so far, including the angiotensin receptor [19] and the rhodopsin-arrestin complex [20] which had both resisted structure determination at synchrotron facilities. LCP volumes as little as 25 μl are sufficient to obtain a full dataset. LCP extrusion also works to deliver soluble proteins [21] and other viscous gels or greases are possible [22,23].

Another method to make a very thin liquid jet in air or vacuum is by the process of electrospinning [24]. A drop of conductive liquid emerging from a capillary can be pulled into a fine Taylor cone by an electric field. With high enough viscosity (achieved by adding glycerol or poly ethyl glycol to the buffer solution) the liquid is formed into a jet rather than a spray of droplets. Jets are achieved at flow rates 10 times slower than liquid jets from a GVDN and hence give a correspondingly higher sample efficiency. The method was used for measurements of the photosystem II complex [25]. Acoustic ejection [26] is a possible way to producing droplets in air of about 50 μm diameter, timed to the arrival of the X-ray pulses. A final method in this survey of flowing samples across the beam is to nebulise the suspension to produce an aerosol which are then directed into a beam in vacuum using an aerodynamic lens. This produces the lowest background, albeit at the worst sample efficiency of all techniques, and the degree of hydration and sample temperature has not yet been fully explored.
The most efficient sample delivery approach is to raster an array of crystals prepared on a grid of membranes or open windows in a support [27-29]. When placed on a fast scanning stage it is possible to place individual cells of the array at the beam at a rate of 120 positions per second. Thus if crystals fill every cell, or if their positions are mapped in advance, it should be possible to achieve a 100% strike proportion to collect 10,000 patterns in less than 2 minutes at LCLS. For full pulse fluence this would require a spacing of about 100 μm between samples, or 1 cm^2 area, for these 10,000. Higher sample densities have been used at less than full fluence. In vacuum, samples must be protected, which has been achieved with a layer of sucrose to measure diffraction from two-dimensional crystals [27]. At atmosphere the sample can be kept humid.

Besides the serial diffraction approach of a single crystal per shot, it has been possible to record single-pulse exposures at different places on the one crystal. With the crystal mounted on a goniometer, a known rotation between the exposures provides a quasi-rotation method that helps to achieve accurate integrated Bragg counts with a hundred crystals or so. This requires a balance between X-ray beam area and fluence, exposure separation, and crystal size, with care that prior exposures do not cause radiation damage even if there are no visible effects. While the large crystals needed would be equally suitable for synchrotron measurements, the short FEL pulse still has the advantage of higher time resolution for photo-induced reactions, and the immunity to radiation damage is a boon for radiation sensitive samples, such as metalloproteins, which otherwise suffer from the reduction of metal centres by radiolysis products. In this way it was possible, to obtain the unreduced structures of photosystem II [30] and cytochrome c oxidase [31].

The new method of collecting tens of thousands of still diffraction patterns required new analysis procedures, and software packages are now available to treat the data [32,33]. The approach developed for the first LCLS experiments was to consider the snapshot patterns of randomly-oriented crystals as providing the input for a Monte Carlo integration of Bragg intensities. That is, even though the precise characteristics of each measured Bragg spot are not known (such as the partiality of reflection) one could simply sum up all counts in all spots that share the same Miller index, with a precision increasing as the square root of the number of observations. This requires each pattern to be indexed, and the software CrystFEL [32] was created to manage the indexing and integrate data in predicted spot locations. The approach was designed to allow summing of signals above noise, even when spots are too weak to be observed in individual patterns. Various strategies to improve precision have been tried by the community besides simply collecting more patterns, such as targeting the pixels which contain signal [34], using seeded pulses of the FEL to reduce wavelength fluctuations, or classifying patterns by unit cell dimensions [35]. However the best improvements came
with the ability to refine spot locations, model and refine partiality and scaling, as well as large improvements in detector linearity and noise levels [36,37]. The approach is accurate enough to phase by anomalous diffraction [38] or isomorphic replacement [39]. The current state of the art can be illustrated by a recent analysis by Ginn et al. [37] of diffraction from the cytoplasmic polyhedrosis virus type 17 (CPV17) collected at 1.3 Å wavelength, which revealed the anomalous signal of sulfur even though this represents only 0.42 e per S atom responsible for a 0.5% change in diffraction amplitudes. This was achieved from 7,225 patterns from a set of 65,564.

As of September 2015 there were 66 entries in the protein databank of structures measured at LCLS or SACLA, from 26 different proteins (including 13 structures in the time evolution of myoglobin [40]). Many of these structures have also been solved at synchrotron facilities. It is hard to compare diffraction data collected from still snapshots to the conventional rotation patterns and and the metrics used to quantify XFEL data statistics appear to give an overly pessimistic view, whereas electron densities subjectively appear superior. The XFEL structures generally have improved side-chain densities, better definition of salt bridges [41,42] and the impression of finer details even when collected at the same diffraction resolution as synchrotron data. In the case of the SNARE complex, the improvements in the XFEL structure was essential for structure interpretation [42]. Room temperature measurements may avoid biases due to cryogenic cooling [43]). The precision of the data collection is also apparent in time-resolved structures obtained at LCLS, including time points in the photocycle of photoactive yellow protein [44] where each time point was treated as an independent dataset.

Data collection with XFEL pulses opens up new opportunities in addition to obtaining structures from crystals too small for other sources. Time-resolved structures can be obtained at timescales down to tens of femtoseconds, which is the accuracy to which an optical pump pulse can be timed to the X-ray pulse arrival. Since the sample is constantly replenished, it is possible to study irreversible reactions. Small crystals are beneficial in these measurements since they better match optical absorption depths, resulting in much larger pumped fractions as compared to time-resolved Laue diffraction experiments carried out at synchrotron facilities. Small crystals are also beneficial for time-resolved mixing experiments as the time for a reagent to diffuse through a crystal is shortened as the volume is decreased, giving durations as short as 150 μs for a 6 μm³ crystal [45]. There are several potential benefits for phasing the diffraction of small crystals, including the idea to utilise the coherent diffraction of the entire (finite) crystal to enable mapping out the Fourier transform of the repeating molecule [46]. The increased information content from measurements in between Bragg peaks can be sufficient to iteratively phase the diffraction, without the need for a model. It has also been proposed that the scattering factors of heavy atoms in the sample (including sulphur) may saturate at high intensities, allowing phasing approaches similar to anomalous diffraction or
isomorphic replacement by sorting data according to pulse intensity [47].

Currently the biggest limitation to the adoption and use of XFEL crystallography is the availability of beamtime. Current facilities are acutely aware of this and are doing their utmost to increase the capabilities and availability of instruments, such as offering 6-hour slots for crystal screening. The LCLS is currently constructing a beamline for in-air measurements and has also shown the viability of refocusing the spent beam in a SFX experiment (passing through the hole in the detector) to run a second SFX measurement in parallel [48]. Availability will soon increase with new facilities to be opened, such as the European XFEL, SwissFEL, and PAL in Korea, all of which have plans for crystallography stations. With the SFX User Consortium at the European XFEL the SPB/SFX instrument at that facility is implementing the spent-beam approach and dedicated instrumentation for serial crystallography. Although the facility will be able to produce 27,000 pulses per second, in bursts of 2,700 pulses separated by only 220 ns, the planned AGIPD detector [49] will be able to record 3,520 of those per second. There are significant challenges to flow the sample fast enough, but by increasing the pulse separation to 1.5 μs (by using 1 in 7 pulses, 3,800 pulses per second) the full AGIPD capacity would be used and 50 m/s sample speeds would be needed. This would yield the same number of collected patterns as at LCLS in about 1/30th of the time. The possibility of almost unlimited patterns may impact strategies to best treat the data, although it is more likely the increased collection rate will be used for more recalcitrant samples and larger parameter studies (e.g. more time points). The challenge will be to improve sample delivery, diffraction measurements, and analysis to enable structural and functional information at much improved resolution compared with synchrotron sources, making full use of the tolerance of the sample to short intense pulses and the ability to measure over a range of physical conditions.

Figure caption:

Fig. 1: Crystallography with X-ray FELs can be distinguished by various methods used to deliver the sample to the X-ray beam, usually for a single snapshot exposure per crystal. The liquid micro-jet (a) gives low background, suitable for sub-micrometer crystals but has low sample efficiency at current FELs because of its high speed, moving the sample a long distance between X-ray pulses. (b) Extrusion jets are slower, giving higher sample efficiency, but at the cost of higher background. (c) Aerosol injectors give the lowest background but also lowest efficiency. The flowing systems of (a) - (c) all can operate in vacuum. (d) Raster-scanned arrays can move fast enough to hit every point (giving optimum sample efficiency and data collection rate) or could be used on a random field of crystals. (e) A large crystal mounted on a cryo-loop on a goniometer can be exposed in several places with known angular increments between pulses.
References