Chapter 10

Serial Synchrotron X-Ray Crystallography (SSX)

Kay Diederichs and Meitian Wang

Abstract

Prompted by methodological advances in measurements with X-ray free electron lasers, it was realized in the last two years that traditional (or conventional) methods for data collection from crystals of macromolecular specimens can be complemented by synchrotron measurements on microcrystals that would individually not suffice for a complete data set. Measuring, processing, and merging many partial data sets of this kind requires new techniques which have since been implemented at several third-generation synchrotron facilities, and are described here. Among these, we particularly focus on the possibility of in situ measurements combined with in meso crystal preparations and data analysis with the XDS package and auxiliary programs.

Key words Serial synchrotron crystallography (SSX), Microcrystal, Lipidic cubic phase (LCP), In meso in situ, Room temperature (RT), Cryogenic temperature, Data collection, Data quality, Merging, XDS, XSCALE

1 Introduction

Macromolecular crystallography (MX) has been constantly evolving since the very first X-ray structure determinations of protein molecules in the 1950s and 1960s. Nowadays X-ray crystal structures of biological macromolecules are determined at an unprecedented speed; this year, about one structure is deposited every hour in the Protein Data Bank (PDB). This is due in large part to developments in molecular biology, crystallization, data collection and processing and structure solution, as well as to advances in synchrotron radiation technology. However, the basic method of X-ray diffraction data collection remains unchanged; almost exclusively, diffraction data are collected on a single crystal entity with the rotation method [1] using a monochromatic X-ray beam. In this experiment, the most important measured quantity is the integrated intensity of a reflection, which is given by Darwin's formula [2, 3]:

$$I_{hkl} = I_0\, r_e^2 \frac{V_{xtal}}{V_{cell}^2} \frac{\lambda^3}{\omega} LPA \left|F_{hkl}\right|^2 \qquad (1)$$

Alexander Wlodawer et al. (eds.), *Protein Crystallography: Methods and Protocols*, Methods in Molecular Biology, vol. 1607, DOI 10.1007/978-1-4939-7000-1_10, © Springer Science+Business Media LLC 2017

where I_{hkl} is the integrated intensity of reflection hkl, and is proportional to the square of the structure factor (F_{hkl}), I_0 is the intensity of the incident X-ray beam, r_e is the classic electron radius, λ is the X-ray wavelength, ω is the angular velocity of the crystal, L is the Lorentz factor, P is the polarization factor, and A is the X-ray transmission. For our purpose, we can leave various correction factors and constants out and assume that the squared structure factor is proportional to the content of the unit cell. Then Eq. 1 can be written as [4]:

$$I_{hkl} \sim I_0 \frac{V_{xtal}}{V_{cell}} \tag{2}$$

In Eq. 2, the measured intensity is proportional to both the intensity of the incident beam (I_0) and the diffraction volume illuminated by X-rays (V_{xtal}) and is inversely proportional to the unit cell volume (V_{cell}). This means, simply, that the intensities of the measured reflections decrease as the crystal size gets smaller, or as the unit cell size gets larger. Another important aspect in diffraction data collection is radiation damage, which limits the maximum obtainable data resolution for a given diffraction volume. This has a significant consequence: there is a lower limit to the size of crystals one can reliably extract diffraction data from, prior to the onset of radiation damage [4].

In the pioneering work of macromolecular crystallography in the 1950s and 1960s, experiments were carried out with well diffracting large crystals at room temperature [5, 6]. Good crystalline order and large diffraction volume allowed acquiring diffraction data with sufficient accuracy at room temperature. In the 1970s and 1980s, the structural study of viruses presented new challenges in diffraction data collection. Because of the very large unit cell (one or two magnitudes larger than for average proteins), the intensity of reflections decreases accordingly (Eq. 2). Only one or two diffraction patterns could be obtained from each crystal and the complete data set had to be assembled from many crystals [7]. This could be considered as the very first serial crystallography (SX) work although the term SX did not exist at that time. From the 1990s to 2000s, synchrotron radiation started to play an important role in modern macromolecular crystallography (http://biosync.sbkb.org). The brightness and energy tunability of synchrotron radiation enabled study of crystals smaller than before and established experimental phasing with anomalous scattering [8]. Around that time, cryogenic cooling methods were also being developed that allowed data collection with greatly reduced radiation damage [9]. Since then, collecting complete data sets from a single crystal became the method of choice in MX (referred to as "conventional crystallography (CX)" here) and it works well for crystals with diffraction volume of around 10,000 μm^3 (about

20 × 20 × 20 μm^3) and larger [10]. From the 2000s, advances in crystallization methods enabled crystallization of challenging targets, such as large multi-protein complexes and membrane proteins, which often only yield micro-crystals with the largest dimension below 20 μm and the smallest dimension below 5 μm. This is about one order of magnitude smaller in diffraction volume compared with crystals used in CX. Therefore, radiation damage prevents collecting complete data sets to high resolution from those crystals. At the third generation synchrotron sources using microfocused X-rays with beam size comparable to the crystal size, partial data sets could be collected from these microcrystals, which are then merged together to form a full data set. This method is referred as microcrystallography [11, 12].

The first hard X-ray free electron laser (XFEL) opened new avenues for crystallographic data collection in 2009 [13]. The extremely high intensity and femtosecond pulses of XFELs are not suited for the rotation method, because the X-ray pulse would destroy the crystal almost immediately. However, during the femtosecond pulse, the diffraction pattern resulting from the pulse can be collected prior to crystal damage, generating a still image—the so-called "diffraction before destruction" method. Hundreds to many thousands of such still images on as many isomorphous crystal entities can be merged together to generate a full data set. This method, termed serial femtosecond crystallography (SFX), has the advantage that even the smallest microcrystals, and potentially nanometer-sized crystals, can give useful diffraction data with femtosecond XFEL pulses. Since then, sample preparation and delivery, data collection and processing methods have been actively developed to facilitate SFX experiments [14] and they have inspired recent development of SX at synchrotron sources. Although the serial data collection approaches have been used in virus crystallography and microcrystallography before, the traditional sample mounting methods are simply unfeasible when thousands of crystal samples need to be measured to get a sufficiently complete data set. The new high-throughput sample delivery methods developed for SFX have enabled screening numerous crystals and collecting their diffraction data with synchrotron radiation at an unprecedented speed. This new data collection method is named serial synchrotron crystallography (SSX), which is emerging as complementary method to CX.

The diffraction principles of both methods are the same but SSX departs from CX in many aspects, ranging from sample preparation and sample delivery to data collection and processing. Below we document the experimental and computational procedures for SSX that were devised and applied at synchrotron facilities (Fig. 1). Although the method is still under extensive development, it has already delivered data with high resolution comparable to CX data collection methods and with quality sufficient for experimental phasing.

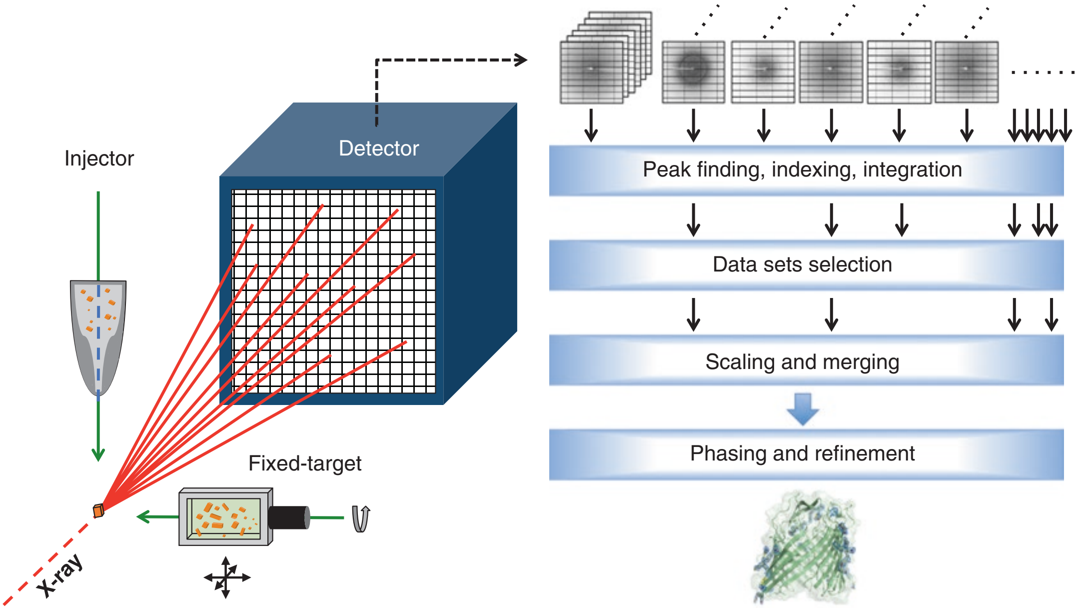

Fig. 1 Serial synchrotron crystallography (SSX): from crystals to structures

2 Sample Preparation and Delivery

In CX, single crystals are treated with cryoprotectant, harvested in loops and snap-cooled in liquid or gaseous nitrogen. Each crystal is then screened for its X-ray diffraction and the best ones are used for the final data collection. This method is effective when individual crystals have enough diffraction volume to yield a complete data set. However, it is time-consuming and cumbersome in microcrystallography and is certainly not compatible with serial crystallography, in which hundreds to thousands of crystals need to be investigated. Therefore SX calls for methods to prepare crystals in a sufficient quantity and deliver them serially into the X-ray beam in a high-throughput manner. In CX crystallization is optimized for the growth of large crystals [15]. In contrast, the demand for large numbers of small crystals in SX requires alternative sample preparation and delivery methods [16]. A variety of sample delivery systems for SX have been developed and tested at synchrotron beamlines in recent years. They can be broadly grouped into two classes—injector methods and fixed-target methods. We review them together with related crystallization developments in this section.

2.1 Injector Methods

The injector methods were originally developed for SFX application at XFELs. They come in two main variants—the gas dynamic virtual nozzle (GDVN) injector and the lipid cubic phase (LCP) injector (and a closely related version called high viscosity extrusion (HVE) injector) [17–19]. The GDVN injector uses gas focusing to generate a liquid stream of a few micrometers in diameter and was used in the first SFX experiments with micron-sized crystals of photosystem I (0.2–2 μm) [13], lysozyme (1 × 1 × 3 μm^3) [20], and cathepsin B (0.9 × 0.9 × 11 μm^3) [16] at the Linear Coherent Light Source (LCLS). The GDVN injector enabled measurement of high-resolution diffraction images from nanometer to micrometer sized crystals with minimum diffraction background while still keeping the crystals fully hydrated. However, sample consumption is very high with GDVN (10–100 mg per data set) and the high flow rate of 10 m/s makes the crystal-X-ray interaction time too short to measure sufficient diffraction data with synchrotron radiation. Based on the GDVN concept, an electrospinning injector was designed with a tenfold reduction in flow rate [21]. However, the electrostatic charging from the electrospinning method may affect crystals. The LCP and HVE injectors extrude a continuous 20–50 μm diameter stream at a much slower velocity of 0.1–0.3 mm/s, which is more suitable for data collection with micro-focused X-rays at synchrotron sources. In addition to LCP and other meso phases with membrane protein crystals, the method has been extended to other high viscosity media such as grease, Vaseline, and agarose, which could be used as carrier media for soluble proteins [19, 22, 23]. The LCP injector reduces sample consumption dramatically (50- to 100-fold) compared to the GDVN injector and has enabled structure determination with SFX methods of several membrane proteins and complexes (β-adrenergic receptor [24], opioid receptor [25], angiotensin receptor [26], and rhodopsin-arrestin complex [27]). In these studies, crystals were grown in LCP and their size varied from sub-ten to a few tens of micrometers and the average protein consumption is less than 0.5 mg per data set.

Serial crystallography with LCP/HVE injectors has been explored in synchrotron sources recently. Using lysozyme as a model system, Botha et al. reported the high-resolution structure and feasibility of experimental phasing from SSX methods using an HVE injector at the PXII beamline at the Swiss Light Source (SLS) [19]. Lysozyme was crystallized with a standard protocol and crystals were introduced into LCP and vaseline by gentle mixing. The average crystal size was 10 × 10 × 30 μm^3 and the diameter of the LCP/Vaseline stream was 50 μm. Around the same time, Nogly et al. obtained the first SSX membrane protein structure with the LCP injector at beamline ID13 at the European Synchrotron Radiation Facility (ESRF) [28]. The test protein was bacteriorhodopsin (bR) and crystals were grown in LCP with average crystal size of 5 × 30 × 30 μm^3. The diameter of the LCP stream was about 50 μm. Typically 10–20 μl samples were loaded into the injector per experiment. The sample consumption is about a few hundreds micrograms of protein per data set.

The injector is essentially a container-free sample delivery method with low diffraction background and the data acquisition is simple and high-throughput. However, the method comes with its own challenges: (1) optimizing the crystallization to generate a large number of relatively homogenous crystals, whose size matches the X-ray beam and whose concentration is optimal with the flow rate of the injector stream; (2) controlling flow dynamics to minimize crystal movement when it passes through the X-ray beam; (3) processing and merging of still diffraction images.

Related to the injector method, thin-walled glass capillaries (10 μm walls and 100 μm inner-diameter) have been used as a container to accommodate a flow of crystal suspension for SSX experiments at beamline P11 at PETRAIII [29]. The combination of low flow velocity (5 mm/s) and scanning of capillary allows SSX data collection with micro-crystals in aqueous media. Lysozyme crystals of 3–6 μm in size were used as test protein. One drawback of the capillary method is the additional background scattering from capillary walls and crystallization solution inside the capillary. Another related approach is acoustic droplet ejection (ADE) [30]. Instead of delivering samples continuously, ADE offers "drop-on-demand" to eject samples only when needed. The sample consumption can be greatly reduced and the data collection can be carried out both at room and cryogenic temperatures [31, 32]. In addition to injecting droplets, acoustic force can levitate droplets in the X-ray interaction region [33]. In this case, the crystals are rotating inside the droplet and a diffraction movie is recorded with a fast frame-rate X-ray detector to trace out the orientation change.

2.2 Fixed-Target Methods (Deposition and In Situ)

Fixed-target methods were developed as an alternative approach to injector methods to improve sample hit-rate and reduce sample consumption in SFX experiments. In principle, the hit-rate could reach 100% and sample consumption could be reduced to micrograms The first demonstration was carried out with rapid encystment protein (REP24) crystals of $5 \times 10 \times 30\ \mu m^3$ in size deposited on silicon nitride (Si_3N_4) membranes (of 50 nm thickness) and protected by Paratone-N for room temperature measurement under vacuum at the LCLS [34]. Soon after, synchrotron goniometer-based data collection methods have been extended to SFX with crystals either positioned inside a grid or mounted with a loop or a mesh [35, 36].

At synchrotron MX beamlines, the fixed-target and related systems offer full control of the data collection and could be made compatible for measurement at both room and cryogenic temperatures. Most established data collection and processing methods for CX are readily adapted. Therefore, SSX with various fixed-target systems has been actively pursued recently at third generation synchrotron facilities as an extension to the established synchrotron micro-crystallography.

One main challenge in the fixed-target methods is to reduce X-ray background scattering from support materials and crystallization media around the crystals. If thin film materials are used to hold crystals, they should ideally be watertight, optical- and UV-transmitting, and non-birefringent. Various low X-ray background materials have been examined, such as polymer films like cyclic olefin copolymer (COC), PDMS, and PMMA in micrometer thickness, and Si_3N_4 and Si membranes in nanometer thickness. As one of the thinnest materials possible, graphene as crystal container has been successfully demonstrated recently [37–39]. Crystals can be either deposited on a chip or a grid or simply a conventional loop or mesh for cryo-crystallography or grown in situ in a crystallization plate designed with low diffraction background.

2.2.1 Deposition Methods

The deposition method is a two-step process. Crystals are grown by conventional crystallization methods first, and then transferred to the fixed-target supports for diffraction measurement. In order to expedite the serial measurement, chips with features promoting self-assembly and self-localization of crystals and grids with tailored hole sizes have been developed to allow automated data collection with predefined crystal locations. A chip with a Si mesh and polyimide film has been designed to position crystals in the prescribed locations with random orientations by exploiting the liquid-pinning potential and surface roughness [40]. With lysozyme (rod-shape, 5–50 μm) and ferritin (block-shape) as model systems, diffraction data were obtained at room temperature at beamline PXII at the SLS. To assist loading microcrystals, a micropatterned chip was developed. The chip features 150 nm thick Si_3N_4 windows to reduce X-ray background scattering and an alternating hydrophobic and hydrophilic surface pattern to assist localizing crystals in defined regions [41]. Still diffraction images were collected with lysozyme crystals of 10–50 μm in size at GM/CA beamline at the Advanced Photon Source (APS). A single-crystalline Si chip with micropores was developed to minimize background scattering [42]. The chip is made from single-crystalline Si with an active area of 1.5×1.5 mm^2 and 10 μm thickness. The size (typically a few microns), shape and pattern of micropores are controllable by micro-fabrication. In principle thousands of crystals can be loaded on a single chip. The idea is that extra crystallization solution, which contributes to background scattering, is wicked away through the micropores, while the crystals that are larger than the pores are retained. The compact format of the chip is cryo-compatible and the whole chip can be snap-cooled in liquid nitrogen. Rotation diffraction data were collected with microcrystals of CPV18 polyhedrin of size 4 μm or smaller at beamline I24 at the Diamond Light Source (DLS). In a different approach to increase hit-rate and automate serial data collection, a microfluidic chip with an array of traps was designed to collect crystals at predefined locations and semi-still diffraction images (0.02°) were collected at beamline 12-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) [43]. To facilitate SX at standard MX beamlines, a compact sample-mounting grid made with polycarbonate plastic with 75 holes of 400, 200, and 125 μm in diameter was designed at the SSRL [35, 44]. The crystals can be loaded either manually or automatically with liquid-handling robots. The grid fits into the magnetic base and puck used in standard cryo-crystallography. The grids can be mounted by an automatic sample changer as for standard loop samples and the diffraction data can be collected at cryogenic temperature with cryojet cooling in the co-axial configuration.

Instead of depositing crystals in predefined locations/regions, Coquelle et al. loaded lysozyme crystals between two Si_3N_4 windows and used the continuous rastering method to collect still diffraction images from the entire Si_3N_4 assembly [45]. The crystal mounting loops and meshes commonly used in cryo-crystallography could be adapted for SSX experiments. In fact, the micro-mesh has been used to mount micro-crystals for collecting diffraction data serially for more than a decade [46–51]. Typically, a suspension containing many microcrystals is loaded on the mesh and extra solution is removed before snap-cooling. The crystals are located either visually with an on-axis microscope at the beamline or with diffraction scanning. Recently this approach has been extended to allow continuous data collection by helical scanning of a loop loaded with numerous micro-crystals. With this method, the structure of cathepsin B to 3.0 Å resolution has been determined from needle-shape crystals with diameter less than 1 μm at beamline P11 at PETRAIII [52].

2.2.2 *In Situ Methods*

The deposition process requires additional manipulation of crystals, which is time-consuming and may damage the crystal. It would be advantageous if X-ray data collection could be carried out directly with crystals in their crystallization compartment. This is the in situ method and it has many variants. Microfluidics is an attractive technology for crystallization because it allows fast screening and optimization of broad crystallization conditions with a minimum amount of protein [53]. With low X-ray scattering materials, crystallization and in situ X-ray data collection can be performed on a single microfluidic chip. Microbatch [54] and counter-diffusion [55, 56] crystallization methods have been implemented and the in situ X-ray diffraction experiments have been conducted with model proteins. The feasibility of experimental phasing with anomalous diffraction has been verified with in situ diffraction data from selenomethionine derivatized proteins, Yb derivatized crystals, and with a native protein [54, 56, 57].

Recently, microfluidic methods have also been successfully applied to in meso crystallization for membrane proteins [58]. To control crystal size and concentration, a kinetically optimized microfluidic chip was developed to crystallize proteins in emulsion droplets with one crystal per drop [59]. The microfluidic chips are suitable for SSX experiments. The room temperature SSX data from phosphonoacetate hydrolase (a soluble protein) and the photosynthetic reaction center (a membrane protein) have been collected at beamline LS-CAT at the APS [57, 58]. The SSX data from glucose isomerase has been measured at beamline F1 at the Cornell High Energy Synchrotron Source (CHESS) [59].

The 96-well crystallization plate with SBS format has been used for in situ diffraction screening at many MX beamlines [60–62]. For this purpose, 96-well crystallization plates with low background scattering were designed (CrystalQuickX and MiTeGen In-Situ-1). In addition to screening, complete data sets were obtained from both soluble and membrane proteins with in situ data collection [62–64]. The standard in situ SBS format plates still have relatively thick films. A thin polymer-film sandwich (TPFS) has been recently demonstrated with films as thin as 10 μm [65]. The whole TPFS plates can be used for in situ data collection at room temperature and each individual well can be cut out from the plate and placed under a cryojet stream for data collection at low temperatures. To bridge the in situ method and the conventional loop harvesting method, the CrystalDirect approach was developed to automate crystal harvesting directly from the crystallization plate through laser photoablation. This method allows controlled selection of crystals and removal of extra crystallization solution before harvesting [66, 67].

For membrane proteins, the lipid cubic phase (or in meso) crystallization is a very effective method [68]. The crystals grown by this method tend to be small and difficult to harvest due to the high viscosity of the mesophase. IMISX (in meso in situ serial crystallography) methods have been developed to allow efficient in situ serial data collection from microcrystals [69, 70]. The IMISX uses a double sandwich plate design where the crystallization takes place in the inner chamber made by two thin (25 μm) COC films separated by a spacer and the outer chamber consists of glass plates to avoid water loss and for easy handling and transportation. The IMISX plate can be set up either manually or robotically. Individual wells can be cut out from the plate and directly used for in situ data collection at room temperature or snap-cooled in liquid nitrogen for cryogenic data collection. The methods have been validated with several membrane proteins including enzymes, transporters, and receptors. High resolution structures have been obtained with crystals as small as 5 μm at beamline PXI at the SLS [70]. The IMISX methods are applicable for soluble proteins as well.

The recent development in IMISX and TPFS methods enabled in situ SSX for both membrane and soluble proteins at both room and cryogenic temperature. The ability to collect data at room temperature allows conformational space and dynamics of macromolecular molecules to be probed [71, 72]. At cryogenic temperatures with significantly reduced radiation damage, complete data sets can be obtained with fewer crystals, thus reducing sample consumption. It has been shown that nanogram to single-digit microgram quantities of protein can yield a high quality SSX data set. More importantly, cryogenic freezing allows us to prepare samples in advance, preserve and store them in their best state, and transport them for diffraction data collection when beamtime becomes available.

3 Data Collection

In the last decade, and primarily prompted by the advent of the PILATUS detector, crystallography has transitioned from the traditional “high dose” strategy (exposing the crystal to obtain as much signal as possible per diffraction image) to a “right dose” strategy, where the maximum attainable signal for a crystalline entity is more carefully determined to minimize the effects of radiation damage. In this section we document the boundary conditions applicable to data collection in general, and the peculiarities of partial data sets from small crystals.

3.1 Diffraction Signal and Noise

The primary goal of diffraction data collection is to obtain a complete data set with both high precision, as characterized by values of $CC_{1/2}$ and I/σ of the merged data overall, but particularly in the highest resolution shell, and high accuracy, i.e., with a minimum deviation of intensities from their true values. The precision of the data is essentially limited by a compromise between dose and the X-ray induced radiation damage, and their accuracy is often limited by the systematic errors in measurement, where—again—radiation damage is a large contributor. The strength of the diffraction signal is determined by both energy and flux of X-rays delivered on the crystal, and the intrinsic diffraction properties of the crystal, as discussed in depth by Holton and Frankel [4]. Briefly, the total diffracted signal is proportional to the diffraction volume and the average diffracted signal per reflection is inversely proportional to the unit cell volume (Eq. 2). Therefore, higher flux in the incident beam, or larger diffraction volume or both are needed for crystals with large unit cells. The background under Bragg peaks is from scattering of any material along the X-ray beam path, which could be a disordered portion of the crystal, material around the crystal (e.g., crystallization solution, lipid phase in in meso crystallization), sample support (e.g., mounting loop, films in fixed-target supports), and air. The background scattering has characteristic maxima around 3.6 and 4.5 Å for water and lipid cubic phase

(LCP), the two most common crystallization media, respectively. Reduction in X-ray background is essential for precise and accurate measurement of weak diffraction signals from micro-crystals.

The most effective approaches to improve signal-to-noise ratio in X-ray diffraction experiment with synchrotron radiation are to match X-ray beam size and crystal size and to reduce extra scattering materials in the X-ray path. Illuminating the whole diffraction volume enhances the diffraction signal with less absorbed X-ray dose, and reducing surrounding materials around the crystal minimizes the background scattering. In SSX with injector methods, one needs to find a good combination of X-ray beam size, crystal size and velocity of the stream to have most of the crystal illuminated by X-rays while it passes through. Ideally one should have a diameter of the injector stream not much bigger than the crystal size. In practice, a certain minimum stream size is needed to avoid clogging. In SSX with fixed-target methods, huge efforts have been made to develop systems with low X-ray scattering materials and less crystallization media around crystals. The goniometer or scanning stage, which holds the fixed-target samples, allows full control of crystal characterization and data collection. Crystals can be located with a grid scan, and rotation data can be collected for each selected crystals subsequently [50, 70]. Alternatively, the whole sample could be scanned either with still images or with oscillation images [45, 52].

3.2 Radiation Damage

Radiation damage limits the amount of diffraction data that can be obtained from a given diffraction volume. A protein crystal can stand doses of a few kGy at room temperature before significant loss of diffraction signal. The exact tolerable dose is case-specific and largely depends on diffusion processes of free radicals, which generate secondary damage [73]. Therefore, the damage is both dose and time (dose-rate) dependent. It has been suggested that at sufficiently high dose-rate, some “undamaged” diffraction data could be obtained before the diffusion processes start destroying crystallinity [45, 74]. At room temperature, the damage can spread well beyond the irradiated area and result in crystal deformation and cracking. In the case of cryogenically cooled crystals, the damage is only dose dependent and does not extend beyond a few μm from the irradiated area. More importantly, cryogenic cooling extends the tolerable dose to 20–30 MGy [75, 76], which is about two orders of magnitude higher compared with the dose limit at room temperature. This allows a useful amount of diffraction data to be measured from micron-sized crystals. For an average protein of several hundred amino acids, a $20 \times 20 \times 20\ \mu m^3$ crystal can yield a complete data set [10]. Therefore, it is safe to say that 3° of usable data to diffraction resolution can be obtained from one $5 \times 5 \times 5\ \mu m^3$ crystal prior to the onset of damage. If the average crystal size is known, the required number of crystals for a complete data set can be estimated. With further beamline optimization in reducing background scattering and improvement of X-ray detectors, more data can be obtained from even smaller crystals. It has been estimated that a crystal as small as 1.2 μm can produce a complete data set to 2 Å resolution under ideal experimental conditions [4].

3.3 X-Ray Beam and Detector for SSX

Recent advances in X-ray optics have made micrometer-sized X-ray beams routinely available at many MX beamlines at third generation synchrotron sources [12]. The next generation synchrotron technology promises even smaller X-ray beam with much reduced divergence and one to two orders of magnitude higher flux density. For challenging cases of weakly diffracting micro-crystals, the X-ray beam size and divergence could be tailored to increase I/σ by maximizing diffraction signal, minimizing background scattering, and reducing spot size on the detector [77]. Modern pixel array detectors are particularly suitable for the SSX experiment: the single photon sensitivity allows reliable measurement of weak diffraction signals; the small pixel size improves characterization of sharp reflections (common in room temperature and in situ measurements); large active area and high dynamic range allows accurate measurements of both strong and weak reflection spots; and high frame-rate with negligible deadtime enables continuous, shutterless data acquisition [78].

For injector methods, the "dynamics" of the interaction of X-rays and crystal can be monitored using X-ray detectors with fast frame-rate. Diffraction images with corrupted diffraction signal, due to crystal either moving out of X-ray beam or being damaged by X-rays, can be excluded in data processing. For fixed-target methods, when combined with X-ray microbeam, high flux and fast scanning stages, a fast detector allows crystal localization, diffraction characterization, and data collection in an automated workflow [50].

3.4 Data Collection Strategy

Various data collection strategies for single-crystal work with the rotation method (CX) have been used over the past decades and most of them were derived from work with image plate and CCD detectors, and influenced by the capabilities of data processing software at the time. The most common method used to be the collection of the minimum needed multiplicity in the minimum angular coverage [79, 80] with an accumulated X-ray dose below the 20 MGy limit. This method has been applied to micro-crystals in SSX as well. It is impossible to collect a complete data set from each crystal to its diffraction resolution due to its small diffraction volume. Typically intense and micro-focused X-rays are used to compensate for limited diffraction volume and to minimize background scattering, and a small wedge of data is collected until the 10–20 MGy limit is reached. The process is repeated for each crystal until the desired completeness and multiplicity are achieved

[50, 62, 63, 69, 70]. The data sets obtained in this way have sufficient quality for most molecular replacement and refinement calculations. However, the data quality may not be good enough for experimental phasing with weak anomalous scatterers, which demands data with higher accuracy and precision [81].

More recently it was realized that pixel array detectors (PADs), such as the PILATUS makes it possible to spread the tolerable X-ray dose over a larger total rotation range than the minimum required by space group symmetry and a particular crystal setting, without the penalty incurred by the readout noise of CCD detectors. This leads to high multiplicity data sets typically covering rotation ranges of 180–360°, and has the advantages of not requiring a strategy calculation, resulting in high completeness, allowing efficient scaling and outlier rejection, reducing systematic errors by averaging over their possible values, and leaving a safety margin for potentially discarding radiation-damaged frames near the end of data collection [82, 83]. Although the diffraction signal is weak in each diffraction image, it can be recorded reliably with modern detectors and extracted accurately with data processing programs. This dose distribution strategy is equally applicable for SSX. With the same amount of total dose, instead of a small wedge of data with high dose per diffraction image, a lower dose can be used to collect data with more angular coverage. This strategy will deliver complete data sets quickly with fewer crystals, which allows full characterization of the unit cell, symmetry, space group, and diffraction properties such as mosaicity, Wilson B-factor, and resolution. The intrinsic diffraction resolution of the crystals under study will be reached when more data are added, because averaging (or accumulating) will yield the same signal-to-noise ratio as that obtained with high exposure.

Another recent development in data collection is the multi-orientation and multi-crystal strategy (instead of the conventional single-crystal and single-orientation). Both multi-orientation data collection (i.e., change orientation of the crystal relative to the spindle [84, 85]) and multi-crystal merging methods [86, 87] are very powerful in reducing systematic errors and producing data with higher accuracy, which leads to better experimental phasing and potentially more accurate structures [88]. The SSX data are essentially collected in a multi-orientation and multi-crystal way with a tolerable X-ray dose per crystal. Therefore, the SSX method should be able to produce data with excellent quality as long as there is a sufficient supply of isomorphous crystals.

For data collection with still images as in injector methods and fast scanning in fixed-target methods, the highest tolerable dose could be used in a single shot aimed to use all the diffraction power that one crystal can provide. When crystal size approaches $1 \times 1 \times 1\ \mu m^3$ and smaller, the diffraction volume may be too small to yield sufficient diffraction signal above background noise for one single high-resolution image within the dose limit. The practical limit of the smallest crystal for synchrotron macromolecular crystallography has yet to be established. However, latest developments in data processing promises the possibility of extracting signals from extremely weak data (i.e., sparse data collected with much lower dose and/or from much smaller crystals.) [89].

Table 1 summarizes the protein systems and experimental conditions used in SSX to date. The smallest crystals are blocks 3–5 μm in each dimension and needles with diameter of 1 μm. The X-ray beam size as small as 150 × 180 nm^2 has been used with hen egg-white lysozyme crystals measuring 20 × 20 × 20 μm^3. It is evident that SSX with sub-ten micrometer crystals are reachable at current synchrotron beamlines.

3.5 Data Completeness and Multiplicity in SSX

The expected completeness of the merged data depends on the symmetry, the number of data sets, their angular range and their mosaicity. The latter influence exists because reflections at the start and end of a data set's angular range are partials which are later omitted from scaling. Typically, the effective angular range of a data set is given by its nominal angular range minus two times its mosaicity.

The formula for the statistically expected distribution of multiplicities in the merged data for the case of a random orientation of crystals and a centrally positioned detector is [69]:

$$B(n * s, p, k) = \binom{n * s}{k} p^k (1 - p)^{n * s - k} \tag{3}$$

The distribution is binomial, with a mean equaling the number of data sets n multiplied by two times (if Friedel's law holds) the number of non-centering symmetry operators (e.g., $s = 4$ in $C2$ and $s = 16$ in $P422$), and multiplied by the effective angular range of each data set expressed as a fraction of 180° (p).

The binomial formulation readily allows to calculate the completeness $c = 1 - B(n^* s, p, 0) = 1 - (1 - p)^{n^* s}$ of the merged data. For an effective angular range of 1° ($p = 1/180$) and space group $P1$ ($s = 2$), $n = 207$ data sets are required for 90% complete merged data, and twice that number for 99% complete data. If the effective angular range is 10°, these numbers are 20 and 40, respectively. The average multiplicity corresponding to 90% and 99% completeness is about $2 * 20 * 10/180 \sim 2.2$ and $2 * 40 * 10/180 \sim 4.4$, respectively, independent of the space group.

The observed multiplicity distribution of acentric reflections in SSX data collected from lysozyme crystals in an IMISX plate is plotted together with the corresponding binomial distribution in Fig. 2a. The merged SSX data consist of a summed 135.6° data recorded from 113 crystals and a rotation range of 1.2°. The

Table 1
Summary of experimental setups for serial synchrotron crystallography[a]

Protein	Delivery method	Wavelength (Å)	Beam size (µm)	Flux (Ph/Sec)	Dose (MGy/crystal)[b]	Temp (K)	Crystal size (µm)	Osc (°)	No. of crystal	S.G.	Resolution (Å)	PDB ID	Beamline/reference
β_2AR-T4 L[c]	Deposition (loop)	1.033	10.6 × 11.6	2.2×10^{11}	10–15	100	4 × 8 × 25	10–15	27	$C2$	2.4	2RH1	APS GM/CA [47]
AcMNPV	Deposition (mesh)	0.9778	8 × 8	n/a	n/a	100	5–10	n/a	17	$I23$	1.84	2WUX	DLS I24 [48]
AcMNPV SeMet	Deposition (mesh)	0.9778	8 × 8	n/a	n/a	100	5–10	n/a	31	$I23$	3.0	n/a	DLS I24 [48]
Lysozyme	Deposition (chip)	1.0	n/a	n/a	n/a	293	5–50	1	~150	$P4_32_12$	2.3	n/a	SLS PXII [40]
BEV2	In situ (plate)	0.96859	20 × 20	1×10^{12}	(0.5)	297	50–60	~0.4	28	$F23$	2.1	n/a	DLS I24 [62]
FcγRIIIA	In situ (plate)	097791	20 × 20	5×10^{11}	(0.1)	297	30 × 30 × 30	~1	44	$P6_122$	2.4	n/a	DLS I24 [62]
PhnA SeMet	In situ (microfluidic)	n/a	n/a	n/a	n/a	295	100–150	10	19	$P4_32_12$	2.11	n/a	APS LS-CAT [57]
CatB in vivo	Deposition (loop)	1.2398	4 × 5	1.2×10^{12}	34	110	0.9 × 0.9 × 11	1	130	$P4_22_12$	3.0	4N4Z	PETRA P11 [52]
CPV18 in cellulo	Deposition (mesh)	0.96859	6 × 6	2×10^{11}	28	100	~5	2	20	$I23$	1.7	4OTV	DLS I24 [49]
CPV18 isolated	Deposition (mesh)	0.96859	6 × 6	2×10^{11}	21	100	~5	3	20	$I23$	1.7	4OTS	DLS I24 [49]
Lysozyme	Capillary	1.27	6 × 9	2.0×10^{12}	0.3	296	3–6	0	40,233	$P4_32_12$	2.09	4O34	PETRA P11 [29]
Glucose isomerase	In situ (microfluidic)	0.9179	100 × 100	5.53×10^{10}	(0.1)	295	30 × 40 × 50	10	72	$I222$	2.09	n/a	CHESS F1 [59]
Photosyn RC	In situ (microfluidic)	0.97950	50	n/a	n/a	298	60–100	5	23	$P4_12_12$	1.45	4TQQ	APS GM/CA [58]
Insulin	Deposition (mesh)	n/a	20 × 20	8×10^{9}	10	n/a	10	6	101	$R3$	1.7	n/a	NSLS X25 [32]
Insulin	Deposition (in situ plate)	n/a	100 × 100	n/a	n/a	294	50	6	88	$R3$	1.8	n/a	NSLS X29 [32]
Insulin	Deposition (belt)	n/a	200 × 200	n/a	n/a	100	50	6	88	$R3$	1.8	n/a	NSLS X12c [32]
Lysozyme	Injector	1.32	10 × 30	2.0×10^{12}	(0.6)	293	10 × 10 × 30	0	11,081	$P4_32_12$	1.9	4RLM	SLS PXII [19]

(continued)

Table 1
(continued)

Protein	Delivery method	Wavelength (Å)	Beam size (μm)	Flux (Ph/Sec)	Dose (MGy/crystal)[b]	Temp (K)	Crystal size (μm)	Osc (°)	No. of crystal	S.G.	Resolution (Å)	PDB ID	Beamline/reference
Lysozyme-Au (MIRAS)	Injector	1.0	10 × 30	1.0×10^{12}	(0.2)	293	10 × 10 × 30	0	11,915	$P4_32_12$	2.5	n/a	SLS PXII [19]
Lysozyme-I (MIRAS)	Injector	1.9	10 × 30	1.0×10^{12}	(0.6)	293	10 × 10 × 30	0	42,115	$P4_32_12$	2.5	n/a	SLS, PXII [19]
Lysozyme-S	Injector	2.07	10 × 30	1.0×10^{1}	(0.7)	293	15 × 15 × 60	0	106,737	$P4_32_12$	2.8	n/a	SLS, PXII [19]
Bacteriorho dopsin	Injector	0.954	2 × 3	9.1×10^{11}	0.7	294	5 × 30 × 30	0	5,691	$P6_3$	2.4	4X31	ESRF ID13 [28]
Lysozyme	Deposition (chip)	1.0	1.5 × 2.5	1×10^{11}	3.2	293	20 × 20 × 20	0	5,966	$P4_32_12$	1.95	4WL7	ESRF ID13 [45]
Lysozyme	Deposition (chip)	1.0	0.15 × 0.18	1.7×10^{10}	29.1	293	20 × 20 × 20	0	46,516	$P4_32_12$	1.85	4WL6	ESRF ID13 [45]
Lysozyme	Deposition (microfluidic)	0.98	10 × 10	2×10^{12}	2.4	294	10 × 10 × 15	0.02	232	$P4_32_12$	2.5	4WMG	SSRL BL12-2 [43]
Lysozyme	Deposition (microfluidic)	1.03318	10 × 10	n/a	n/a	293	10–15	0	324	$P4_32_12$	1.55	4Z98	APS GM/CA [41]
HiTehA	In situ (plate)	1.0	10–50 match xtal	2×10^{11}	(~0.2)	293	10–75	6–10	56	$H3$	2.3	4YCR	DLS I24 [63]
Lysozyme	In situ (IMISX)	1.0332	10 × 18	3×10^{11}	0.27	293	10 × 10 × 20	1.2	113	$P4_32_12$	1.8	4XJB	SLS PXII [69]
Lysozyme Br-SAD	In situ (IMISX)	0.9205	10 × 18	1.5×10^{10}	0.17	293	10 × 20 × 30	2	239	$P4_32_12$	1.8	4XJF	SLS PXII [69]
Lysozyme S-SAD	In situ (IMISX)	1.7	10 × 30	9×10^{9}	0.08	293	10 × 10 × 20	2	992	$P4_32_12$	2.0	4XJH	SLS PXII [69]
PepTst	In situ (IMISX)	1.0332	10 × 18	3×10^{11} 7.5×10^{10}	0.13	293	10 × 10 × 20	0.6	237 335	$C222_1$	2.8	4XNI	SLS PXII [69]
AlgE	In situ (IMISX)	1.0332	10 × 10	1.5×10^{11}	0.2	293	5 × 5 × 20	1	175	$P4_32_12$	2.0	4XNK	SLS PXII [69]
CPV18	Deposition (chip)	0.96859	7 × 7	2×10^{11}	28.5	100	<4	4	22	$I23$	1.5	4X35	DLS I24 [42]

Lysozyme	Deposition (chip)	0.96859	7 × 7	2×10^{11}	20.2	100	<4	3	73	$P4_32_12$	2.1	4X3B	DLS I24 [42]
Bacteriorhodopsin	Deposition (mesh)	0.976	10 × 10	1.5×10^{11}	6.8	100	2 × 5 × 5	10	10	$P6_3$	2.54	5A45	ESRF ID29 [50]
Maelstrom SeMet	Deposition (mesh)	0.979	10 × 10	9.5×10^{10}	4.5	100	20–50	10	45	$H32$	3.46	n/a	ESRF ID23-1 [50]
β_2AR	In situ (IMISX)	1.0332	10 × 18	3.2×10^{11}	8.4	100	5 × 10 × 30	3	104	$C2$	2.5	5D5A	SLS PXII [70]
DgkA	In situ (IMISX)	1.0	10 × 18	1.8×10^{11}	~10–20	100	20 × 20 × 50	20–40	12	$P2_12_12_1$	2.8	5D56	SLS PXII [70]
CPV1 in cellulo	Deposition (mesh)	1.0	10 × 10	$\sim 3.6 \times 10^{11}$	(~90)	100	5–15	1	9–18	$I23$	1.55	5EXY	Australian MX2 [51]

[a] Only SSX experiments with number of crystals larger than 10 are listed. The table is sorted by publication year

[b] Dose in bracket is estimated based on Eq. 5 in [90]

[c] Data on many GPCRs were collected in a similar way

observed multiplicity agrees very well with the binomial distribution, which confirms the random orientation of crystals. From the observed average multiplicity of the merged SSX data, the effective rotation range is estimated to be 0.86°. That is the average amount of data per crystal contributed to the final data set. For comparison, the multiplicity of a lysozyme data set collected from a single crystal with total rotation range equal to the summed total rotation range in SSX is also plotted (Fig. 2a). The distribution of multiplicity is broader and the average multiplicity is lower in SSX data because of the reduced effective rotation coverage.

The above considerations hold if crystal orientations are random. If, however, the crystals have preferred orientations due to their morphology, the completeness will generally be lower, and the distribution of multiplicities will differ from a binomial. The distorted multiplicity distribution features one peak and one shoulder at low and high multiplicities, respectively. This signature can be used to identify preferential orientation during SSX experiments. One such example is given in Fig. 2b. Here, plate-like crystals of the β_2-adrenoreceptor (β_2AR) were grown by the IMISX method and with a tendency to lie with their flat face parallel to the surface of the IMISX plate. The final merged data contains 104 crystals and a rotation range of 3° per crystal. With the estimated effective rotation range of 2.49°, 259° of data in total should give a completeness of 99.7% $(1-(1-2.5/180)^{104*4})$ for the monoclinic space group $C2$. However, the observed overall completeness was only 95% due to the preferred orientation effect. In practice, this issue may be at least partially compensated for by increasing the number of partial data sets, by increasing the tilt

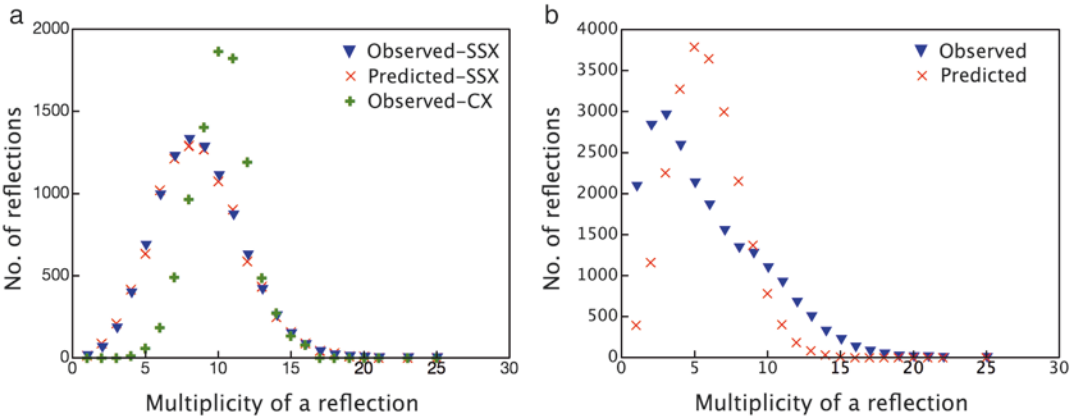

Fig. 2 Distributions of multiplicity in SSX. Figures are adapted from publications [69, 70]. (**a**) Lysozyme SSX data recorded with 113 crystals and a rotation range of 1.2° (effective rotation of 0.86°) and lysozyme CX data recorded with one crystal and a rotation range of 135°. (**b**) β_2AR SSX data recorded with 104 crystals and a rotation range of 3° (effective rotation of 2.49°)

angle of the sample support for data collection, by employing the additional degree of freedom of a kappa goniometer, or by using different mounting methods.

4 Data Processing and Merging

The next paragraphs outline the steps and considerations in data processing of partial data sets collected with the rotation method. The data processing with still images is reviewed in another chapter of this book.

4.1 Processing Individual Data Sets

Any of the commonly used data processing programs, XDS [91, 92], MOSFLM [93], or HKL [94] can in principle be used with partial data sets. However, in practice the processing of hundreds of data sets requires an automated, streamlined procedure that avoids manual intervention. XDS was chosen by us and others for this purpose since it can easily be scripted, and its operation is highly robust.

Scripts have been developed for the processing and merging of data sets collected with PILATUS and EIGER detectors at the Swiss Light Source. One script extracts header information and generates a standard XDS.INP file which is then used to process each partial data set in turn. Owing to the small number of frames in each data set, the running time of the script per data set is short, and data sets can be processed concurrently because the processing directories are uniquely assigned to each data set.

The script creates an XDS.INP file with parameters which differ from the default detector templates distributed with XDS in the following ways:

- for spot finding, the minimum number of pixels in a spot is set to 2, because most of the crystals are smaller than the detector pixels, the beam at beamline X06SA has low divergence, and the point spread function of the PILATUS and EIGER is negligible.
- approximate unit cell parameters and space group are specified if known; the symmetry information only needs to represent the correct Bravais type. Constraints on cell parameters, like equality of axes or fixed angles, increase the accuracy of spot prediction during the integration by reducing the number of degrees of freedom. In principle, space group determination may be carried out after processing all data sets in $P1$, the default if the space group is unknown. Knowledge about the correct Bravais lattice may be obtained from a single weakly exposed low-resolution data set for which the tolerable X-ray dose was spread over a wide rotation range.

- the detector distance as well as the direct beam position on the detector are given as accurately as possible, e.g., as determined with data from a good test crystal.
- the minimum fraction of indexed reflections is set below the default of 0.5 in order to index and integrate as many data sets as possible.

The indexing is usually successful for more than 90% of all partial data sets if enough reflections (50 or more; the minimum in XDS is 25) are found; if multiple adjacent or overlapping crystals contribute to a data set, the indexing usually picks up the strongest lattice.

When processing individual data sets, a reference data set (if available) can be used to resolve indexing ambiguities, which occur in some space groups, and may exist in any space group for specific relationships between cell parameters.

After the final processing, some data sets representing cracked crystals or compromised by the existence of multiple lattices, or reflections from salt, mesophase or ice, or mis-indexing can be identified and discarded. A simple way of doing this is to select those few partial data sets which have at least one cell parameter deviating by more than 3 (or 4) standard deviations from the average; this rule would falsely discard only one out of 370 (or 15,788) data sets if the cell parameters follow a Gaussian distribution. Obviously, it is prudent to start the procedure from a generous cutoff (4 standard deviations, for example), and to iterate the following steps: (a) discard the worst outliers, (b) recalculate the average cell parameters, and (c) tighten the cutoff. Histograms of cell parameter values for SSX data sets from β_2AR are shown in Fig. 3, which indeed possess an approximate Gaussian shape.

4.2 Scaling and Merging the Data Sets

After data integration, the resulting XDS_ASCII.HKL reflection files are scaled and merged in a first XSCALE run. XSCALE has undergone extensive development for serial crystallography; versions released since March 2015 support the efficient scaling and merging of thousands of partial data sets. These tasks require the existence, in each partial data set, of reflections whose unique indices also occur in other data sets. In a situation where the total rotation range of all data sets taken together is much higher than the minimum rotation range for the given space group, almost all reflections of each data set have such counterparts in other data sets, and the resulting network of scaling relationships uniquely determines the scale factor of each partial data set (except for a common arbitrary overall scale factor). However, if the number of partial data sets is so small that their total rotation range approaches the minimum rotation range, some partial data sets may have no unique reflections in common with other data sets, and therefore cannot be scaled. This situation is detected and reported by XSCALE, and those “non-overlapping” data sets have to be discarded after the first XSCALE run.

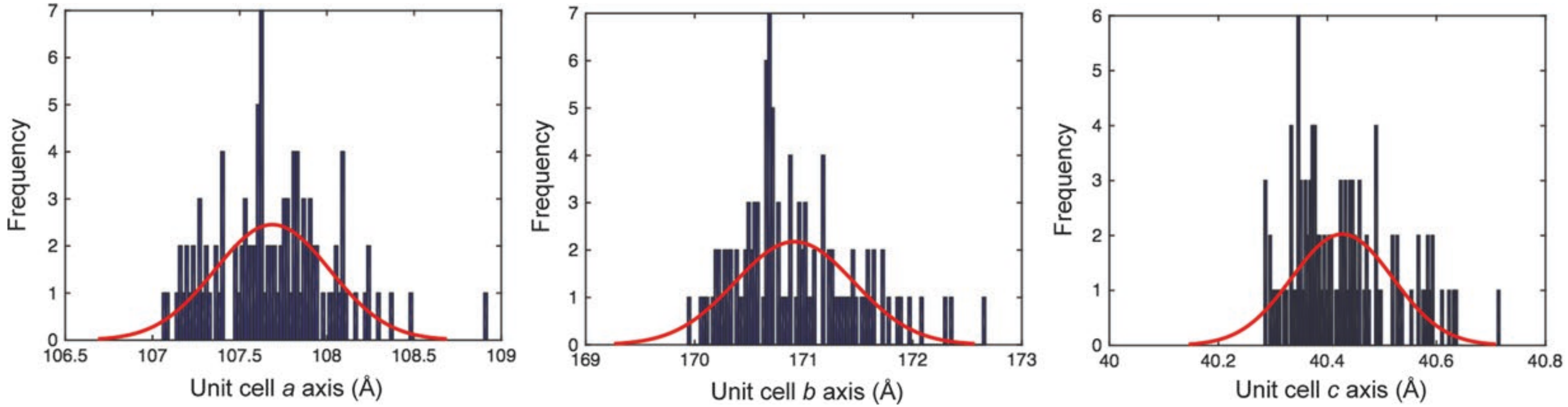

Fig. 3 Histograms of unit cell parameters in SSX data sets recorded with 111 β_2AR crystals (a 10σ peak not shown)

The XSCALE run yields statistics for the completeness and precision of the merged data. These statistics are meaningless if the individual data sets are not consistently indexed, as discussed above. However, situations may arise in which no reference data set is available during processing. In this case, one may use the method of Brehm and Diederichs [95] to identify groups of data sets indexed in the same way, and re-index all except one group to achieve an indexing setting which is consistent across all partial data sets. A program xscale_isocluster (http://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/SSX) is available for this purpose. As a result, all data sets can be merged with meaningful statistics.

The second and further runs of XSCALE are devoted to finding and removing "intensity outlier" data sets. To understand the principles of this procedure, the next section first discusses important aspects of data quality indicators.

5 Assessing and Improving the Precision of Merged Data

5.1 Data Quality Indicators

X-ray crystallography has a history of several decades. Many different kinds of statistical indicators have been defined and applied during this time; some have been adopted by the community, others not. It is remarkable that the most commonly used crystallographic statistic, R_{merge} (also called R_{sym}; [96]), defined as:

$$R_{merge} = \frac{\sum_{hkl}\sum_{i}^{n}\left|I_i - \bar{I}\right|}{\sum_{hkl}\sum_{i}^{n} I_i} \quad (4)$$

where n is the number (multiplicity) of symmetry-related reflections with intensities I_i, has no counterpart in other quantitative sciences. In so far, crystallography has separated itself from mainstream statistical techniques, with some unfortunate consequences for the understanding and interpretation of its data quality indicators. R_{merge} essentially measures the mean fractional deviation of symmetry-related reflection intensities from their average, but is based on absolute differences instead of the statistically better understood and more robust square root of averaged squared differences, like those found in the PCV (percentage coefficient of variation). As with any other absolute difference based residuals, this makes it difficult to perform certain types of calculations with R_{merge}, since there exists, for example, no closed analytical formula for its derivative with respect to its arguments. Another disadvantage is that R_{merge} has no upper limit value; the denominator may become smaller than the numerator in weak high-resolution shells, and large values result, that are difficult to interpret.

Furthermore, R_{merge} calculated from a sample is a biased estimator of the population R_{merge}, in the same sense as the sample variance, when defined as $\frac{1}{n}\sum\left(I_i - \bar{I}\right)^2$, is a biased estimator of the population variance. As with sample variance, which needs to be redefined as $\frac{1}{n-1}\sum\left(I_i - \bar{I}\right)^2$ to be unbiased, R_{merge} needs to be redefined as $R_{meas} = \frac{\sum_{hkl}\sqrt{\frac{n}{n-1}}\sum_i^n\left|I_i - \bar{I}\right|}{\sum_{hkl}\sum_i^n I_i}$, an insight that was published two decades ago [97]. However, even now R_{meas} has not replaced R_{merge}, which is still being used in decision making, where its bias favors low-multiplicity over high-multiplicity data collection.

Finally, the community has not fully realized the fact that both R_{meas} and R_{merge} measure the precision of the individual measurements I_i, rather than the precision of the merged $\bar{I}$. The precision of the merged $\bar{I}$ depends on the sum of the number of photons in each of its n contributing I_i, and thus there are many different experimental strategies that result in the same precision of the merged $\bar{I}$, but yield very different values of R_{meas} and R_{merge}. This fact offers the experimenter an important degree of freedom for optimizing the experiment; favoring those experimental strategies that result in low R_{meas} or R_{merge} biases the experiment toward early radiation damage and the minimal rotation range. For a long time, this has been an unfortunate practice in CX.

There are three indicators that measure the precision of the merged data $\bar{I}$: R_{pim}, which is another variant of R_{merge} in which the factor $\sqrt{n/(n-1)}$ in the numerator of R_{meas}, is replaced by $\sqrt{1/(n-1)}$, thus accounting for the increase in precision by $\sqrt{n}$ when merging n independent observations. R_{pim} shares with R_{merge} the property that its value is unbounded and difficult to interpret.

Second, there is the average signal-to-noise ratio $\langle \bar{I}/\sigma(\bar{I})\rangle$. This indicator suffers from the fact that there are different ways and procedures to estimate $\sigma(\bar{I})$, as is reflected by the fact that different data processing programs yield quite different values for $\langle \bar{I}/\sigma(\bar{I})\rangle$ even if their estimates of the $\bar{I}$ values closely agree [98]. Furthermore, it offers no simple way to identify data sets that degrade the merged signal, because $\langle \bar{I}/\sigma(\bar{I})\rangle$ will always rise when including more observations even if the additional data are non-isomorphous.

Third, there is a correlation-coefficient based quantity called $CC_{1/2}$ which was introduced a few years ago [99], and has gained acceptance in the community because its values allow statistically well-founded decisions [100] particularly about the signal present in weak high-resolution data. Its numerical value lies in the range from −1 to 1 (but in practice only the range 0 to 1 is important) which is easily interpretable, and an analytical relationship with $\langle \overline{I} / \sigma(\overline{I}) \rangle$ exists under well-defined circumstances [101].

5.2 Identifying Outlier Data Sets

In principle, each data set in SSX has both random and systematic differences relative to all other data sets. The random component is an unavoidable consequence of the photon-counting experiment; the systematic difference is usually referred to as non-isomorphism and its size is a priori unknown. Unfortunately, there is no simple way to separate these two types of differences. This is desirable since data sets that are weak (with large random error) should not be discarded, whereas data sets that are non-isomorphous (with large systematic error) should be. It may be noted that the evaluation of unit cell parameters, as explained above, is a first filter for highly non-isomorphous data sets, but since the cell parameters of partial data sets are not very precise, the efficiency of the filter is low.

Several strategies for identifying outlier data sets have been devised and employed by us. Our first attempt [69] used the asymptotic $\langle I/\sigma(I) \rangle$ ratio (ISa), as determined by XSCALE. The ISa value is calculated from the product of the parameters a and b of the error model which XSCALE establishes for each individual data set by fitting its $\sigma(I_i)$ values to the root-mean-square difference between its intensities I_i and $\overline{I}$. One such analysis for β_2AR data is presented in Fig. 4a, and data sets with low ISa values are indicated. The problem with this approach seems to be the fact that the number of reflections in each data set is low and the spread of differences between I_i and $\overline{I}$ may be large, so that although the parameters a, b may result in reasonable $\sigma(I_i)$ estimates, their product may not be very robust.

Our second strategy is straightforward [70]. For each data set, we calculate the average of its intensity correlation coefficients ($CC_{dataset}$) against all other data sets. Finally, we discard those data sets, which display the lowest average correlation. This procedure is robust and does not depend on the $\sigma(I_i)$, but since it cannot differentiate between random and systematic error, it may discard weak data sets and may not discard non-isomorphous ones. Of course, discarding weak data sets does not compromise the merged data much; not discarding non-isomorphous data sets, however, is an issue not solved by this algorithm. Apart from its simplicity, an advantage is that this procedure, which we call “cherry-picking,” can be performed before the XSCALE runs, since it does not require scaled intensities. One such correlation analysis with the β_2AR data is presented in Fig. 4b. Three data sets have $CC_{dataset}$ less than 0.9 and these are also the ones with low ISa values. The methodology of comparing CC values could also be employed in selecting the frame/dose cutoff beyond which radiation damage of a

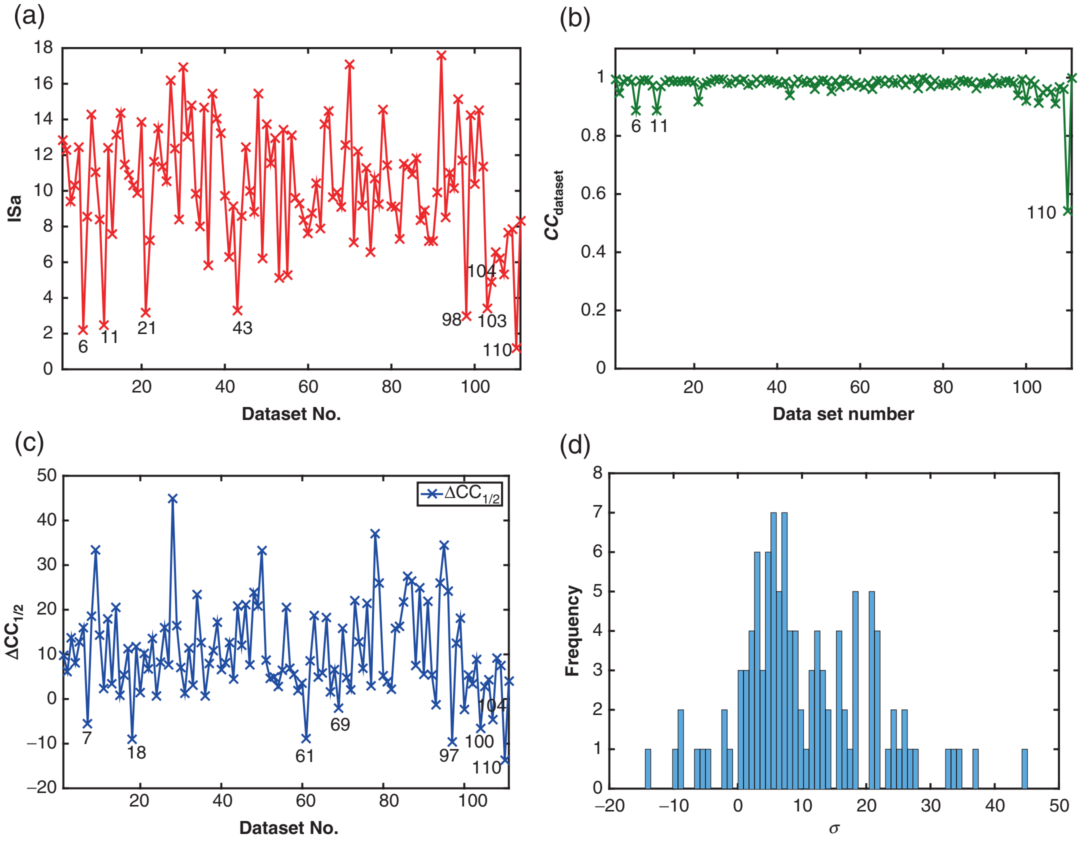

Fig. 4 Data set selection methods applied to β_2AR SSX data (111 data sets). (**a**) ISa method. Data sets with ISa value less than 5 are labeled. (**b**) $CC_{dataset}$ method. Data sets with CC_{datase} less than 0.9 are labeled. (**c**) $\Delta CC_{1/2}$ method. Data sets with negative $\Delta CC_{1/2}$ values are labeled. (**d**) Histogram of $\Delta CC_{1/2}$

partial data set is deemed problematic. As an example, we show room temperature data for the alginate transport protein, AlgE, where it is clear that after seven frames, the correlation (CC_{frame}) with the less damaged data diminishes (Fig. 5a). For these data, we chose to only merge frames 1–5 [69]. When a microbeam is used for microcrystals embedded in sample delivery media, the difficulties in crystal centering in the X-ray beam direction can result in moving part of the crystal out of X-ray beam during the rotation data collection. Such diffraction images can easily be excluded with CC_{frame}. The CC_{frame} can also be used to analyze data integration. For example, the average of all frame-based CC from 111 crystals of β_2AR displays a top-hat profile with a low CC_{frame} for reflections from the first and last frames of the rotation range (Fig. 5b). Their number is low, since most of the reflections in these frames have a partiality below the default acceptance threshold (75%), consistent with a rotation range per frame of 0.1° and average mosaicity of

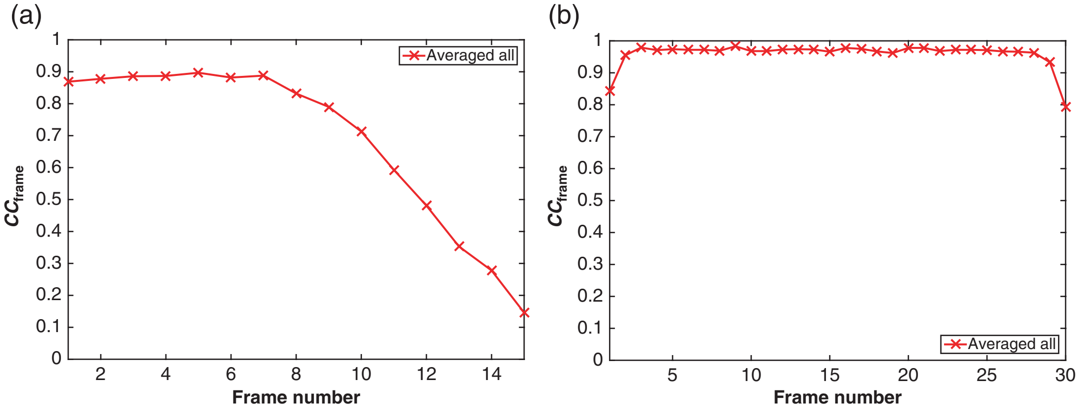

Fig. 5 Average CC_{frame} in SSX data recorded with AlgE and β_2AR crystals. (**a**) AlgE data recorded with 175 crystals and a rotation range of 3° (0.2° per frame). (**b**) β_2AR data recorded with 104 crystals and a rotation range of 3° (0.1° per frame)

0.12°—another illustration of the effective rotation range discussed in Subheading 3.5. The fact that the CC_{frame} is low is mainly due to errors in their partiality estimate, which arise because the geometry refinement of a partial data set (3° of rotation for β_2AR crystals) is less well determined than for CX data sets. To reduce this type of error, we are investigating the use of a higher partiality threshold than the default.

Recently, we showed that a "leave-one-out" calculation of $\Delta CC_{1/2,i} = CC_{1/2,\mathrm{all}} - CC_{1/2,\mathrm{without}\ i}$ can unequivocally identify non-isomorphous data sets [102]. In effect, discarding data sets with strongly negative $\Delta CC_{1/2,i}$ optimizes the target function, $CC_{1/2}$. The distinction between random and systematic error is achieved due to the fact that weak data sets (high random error) should still result in (small) positive $\Delta CC_{1/2,i}$ values, whereas non-isomorphous data sets produce negative $\Delta CC_{1/2,i}$ values. As discussed in the original work, it may be difficult to identify weak non-isomorphous data sets since their $\Delta CC_{1/2,i}$ may be indistinguishable from those of weak isomorphous data sets within the range of observed $\Delta CC_{1/2,i}$ values. In other words, a particular $\Delta CC_{1/2,i}$ value may not necessarily be statistically significant. This consideration suggests that data set outlier detection by this "$\Delta CC_{1/2}$ method" is effective only for sufficiently strong data sets; the dose and crystal size which allows this is under investigation. We applied the $\Delta CC_{1/2}$ method to the β_2AR data and the result is presented in Fig. 4c, d.

The β_2AR example illustrates the challenges in scaling and merging SSX data. Three data set selection methods are in agreement regarding the identity of the worst data set (number 110 in Fig. 4), but not beyond that. Actually, this result is not surprising since, as explained in Subheading 5.2, the three methods differ in their theoretical foundation. Rejecting unjustified outliers will increase the precision, but decrease the accuracy of the merged data. From our recent work [102], we expect that the $\Delta CC_{1/2}$ values can give a more useful ranking of non-isomorphism than the other indicators.

Automation is indispensable in SSX data processing because it is simply not practical to analyze hundreds to thousands of data sets manually. The aforementioned data processing, selection, scaling, merging, and analysis methods are robust and can be easily scripted and incorporated in data analysis pipelines at synchrotron beamlines. Fully automated pipelines from SSX data collection to the final merged data have been implemented recently at the ESRF ("MeshAndCollect" [50]), SPring-8 ("Zoo system," private communication), and the SLS.

6 Summary

In this chapter, we present a variety of methods for exposing microcrystals to the X-ray beam that are far from being voluminous enough to individually yield complete data sets, and an established and proven way for collecting, processing, and merging such data.

A new crystallographic method must be judged by its feasibility and ability to solve and refine new structures. As discussed above, the random orientation of crystals together with a modest oversampling of orientation space ensures good completeness: a coverage of the minimal rotation range (Table 1 in [79]) with about 98–99% completeness requires on average fourfold multiplicity. Since 99% completeness and an average multiplicity of 4 can likewise be considered as reasonable goals when planning single-crystal data sets, it is apparent that SSX from crystals in random orientations is an efficient means for covering reciprocal space. Additionally, SSX has the advantage over single-crystal CX, which is always limited by radiation damage, that using data from additional partial data sets will reliably and significantly improve the merged data, because the scaling is better determined, outlier intensities can be identified and rejected more efficiently, and the higher multiplicity not only results in more precise, but also more accurate data.

After following the processing steps outlined in the previous section and obtaining the merged data, the subsequent procedures for experimental phasing or molecular replacement and refinement against SSX data are, in our experience, the same as those for data collected in CX. In practice, the quality of the resulting data has enabled us to phase bromide soaks and native-SAD measurements [69, 70] with standard procedures, e.g., substructure determination with SHELXD [103], and to refine with phenix.refine [104]. Being able to phase from the anomalous signal thus attests to the high quality diffraction data and high degree of isomorphism attainable with the crystals used.

Undoubtedly, the SSX methods will continue evolving. However, current methods are mature enough for routine use. Any new crystallographic method must be compared with the existing or other developing (or alternative) methods. In this respect, SSX overcomes the limitations of single-crystal work with respect to the availability of large crystals, and the radiation damage that a single crystal tolerates. Of course there is a middle ground; in the traditional approach, several data sets from single crystals can be combined. In CX, single crystals must be harvested and mounted, one at a time, followed by X-ray diffraction screening, which makes the screening of all available crystals impractical. SSX offers attractive alternatives with innovative and automated sample delivery and serial data collection methods.

Data collection in CX provides an extreme example of the "cherry-picking" method. Screening a number of crystals and collecting a final data set from the "best crystal" is common practice, and thousands of structures have been solved this way. However, this is not necessarily best practice when it comes to SSX, because the particular diffraction geometry, chosen rotation range, and crystal peculiarity may lead to systematic measurement errors. It has been demonstrated convincingly that merging data from statistically equivalent crystals can improve both precision and accuracy of the merged data [86].

In this respect, a key assumption of the SSX method is that most crystals under investigation are statistically equivalent (isomorphous). This may not hold for systems where slight differences in molecular packing and/or composition results in crystals with significant differences in their unit cell parameters and/or reflection intensities. If these variations fall into distinct classes, clustering analysis may remedy the problem by sorting crystals into different classes and merging them separately. According to the mosaic block theory, the outcome of a CX experiment is an average structure of all mosaic blocks. The SSX experiment adds another level of averaging across all merged crystals with different levels of non-isomorphism. Based on data to date, the difference between CX and SSX structures would appear to be minor. On one hand, the individuality of each crystal can get averaged out, which can result in a lower number of observed solvent molecules in SSX structures. On the other hand, averaging can enhance common features of crystals, such as alternative side-chain conformations [24, 28, 69].

However, methods to investigate isomorphism (or rather, the lack thereof) are still in their infancy, and there are compelling scientific reasons to develop them, because the lack of dynamic information is one of the shortcomings of X-ray crystallography, which would partly be overcome by detection and analysis of groups of commonly occurring variations in macromolecular crystals.

Current methods are one-dimensional. Specifically, the ISa-based selection, the “cherry-picking” and the $\Delta CC_{1/2}$ methods all lead to a ranking of data sets relative to the average of all other data sets, rather than to a clustering of variants.

Both SSX and CX are bound by the radiation damage limit. The XFEL method has a dose advantage compared to SSX, since the femtosecond pulse can deliver a much higher dose per shot before primary radiation damage processes set in, and may result in a good signal-to-noise ratio at high resolution and ultimately “damage free” structures. For SSX, the number of photons contributing to a merged unique reflection can ultimately only be increased by exposing more crystals.

On the other hand, compared to SFX, which is limited to still images, SSX can accurately sample reflection profiles during rotation. In the case of stills, if the crystals only have a small number of mosaic blocks, the “rocking curve” of a reflection consists of the superposition (addition) of each block’s individual rocking curve, which may be shifted relative to each other. If the number of mosaic blocks were large, their superposition would be (according to the Central Limit Theorem) Gaussian in shape; for small numbers however, each reflection will have a different profile, and may have several maxima and appear jagged. That means that any estimate of full intensity, which is based on a partiality estimate and the sampled portion of the jagged profile, will be in error even if the partiality estimate is correct. We believe that this effect reduces the attainable precision of XFEL data that can be compensated for only by collecting more data. Thus, while the XFEL method has the dose advantage, it suffers the disadvantage of sampling “uneven” reflection profiles, which may lower its usefulness for small crystals. Furthermore, typical protein crystals are far from ideal and their reflection profiles may exhibit non-Gaussian behavior, which also makes profile sampling with still images less efficient.

In summary, SSX has emerged as a complementary method to CX. The technologies developed for SSX and the next generation synchrotron sources make possible the acquisition of better data from smaller crystals, which was either impossible or very tedious and time consuming previously. It should be feasible to obtain high resolution structures with micrometer or even nanometer sized crystals. The serial nature of the SSX experiment makes automation indispensable, which calls for further development in workflows from crystallization, sample delivery to data collection, processing and merging. SSX is also important for screening initial hits in crystallization and for pre-characterizing samples for SFX experiments. Together with CX, SSX and SFX will broaden the horizon for X-ray based structural biology in the coming decades.

Acknowledgments

We thank Greta Assmann, Wolfgang Brehm, Martin Caffrey, Chia-Ying Huang, Vincent Olieric, Ezequiel Panepucci, Rangana Warshamanage, and all other members of the groups at the Swiss Light Source (Paul-Scherrer-Institute, Villigen, Switzerland), Trinity College (Dublin, Ireland) and University of Konstanz (Konstanz, Germany) for discussions and their contributions toward developing the methodology. We also thank Aaron Finke and Martin Caffrey for proofreading the manuscript and Rangana Warshamanage and Chia-Ying Huang for preparing the figures.

References

1. Arndt UW, Wonacott AJ (1977) The rotation method in crystallography. North-Holland Publishing Company, Amsterdam
2. Darwin CG (1914) XXXIV. The theory of X-ray reflexion. Philos Mag Ser 6 27:315–333
3. Warren BE (1969) X-ray diffraction. Addison-Wesley Pub. Co., Reading, MA
4. Holton JM, Frankel KA (2010) The minimum crystal size needed for a complete diffraction data set. Acta Crystallogr D Biol Crystallogr 66:393–408
5. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature 181:662–666
6. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G (1960) Structure of hæmoglobin: a three-dimensional fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. Nature 185:416–422
7. Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G (1978) Tomato bushy stunt virus at 2.9 A resolution. Nature 276:368–373
8. Hendrickson WA (2000) Synchrotron crystallography. Trends Biochem Sci 25: 637–643
9. Hope H (1988) Cryocrystallography of biological macromolecules: a generally applicable method. Acta Crystallogr B 44:22–26
10. Sliz P, Harrison SC, Rosenbaum G (2003) How does radiation damage in protein crystals depend on X-ray dose? Structure 11:13–19
11. Cusack S, Belrhali H, Bram A, Burghammer M, Perrakis A, Riekel C (1998) Small is beautiful: protein micro-crystallography. Nat Struct Biol 5(Suppl):634–637
12. Smith JL, Fischetti RF, Yamamoto M (2012) Micro-crystallography comes of age. Curr Opin Struct Biol 22:602–612
13. Chapman HN, Fromme P, Barty A et al (2011) Femtosecond X-ray protein nanocrystallography. Nature 470:73–77
14. Schlichting I (2015) Serial femtosecond crystallography: the first five years. IUCrJ 2:246–255
15. Gavira JA (2015) Current trends in protein crystallization. Arch Biochem Biophys 602:3–11
16. Liu W, Ishchenko A, Cherezov V (2014) Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. Nat Protoc 9:2123–2134
17. DePonte DP, Weierstall U, Schmidt K, Warner J, Starodub D, Spence JCH, Doak RB (2008) Gas dynamic virtual nozzle for generation of microscopic droplet streams. J Phys D Appl Phys 41:195505
18. Weierstall U, James D, Wang C et al (2014) Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. Nat Commun 5:3309
19. Botha S, Nass K, Barends TRM et al (2015) Room-temperature serial crystallography at synchrotron X-ray sources using slowly flowing free-standing high-viscosity microstreams. Acta Crystallogr D Biol Crystallogr 71:387–397
20. Boutet S, Lomb L, Williams GJ et al (2012) High-resolution protein structure determination by serial femtosecond crystallography. Science 337:362–364
21. Sierra RG, Laksmono H, Kern J et al (2012) Nanoflow electrospinning serial femtosecond crystallography. Acta Crystallogr D Biol Crystallogr 68:1584–1587

22. Sugahara M, Mizohata E, Nango E et al (2015) Grease matrix as a versatile carrier of proteins for serial crystallography. Nat Methods 12:61–63
23. Conrad CE, Basu S, James D et al (2015) A novel inert crystal delivery medium for serial femtosecond crystallography. IUCrJ 2:421–430
24. Liu W, Wacker D, Gati C et al (2013) Serial femtosecond crystallography of G protein-coupled receptors. Science 342:1521–1524
25. Fenalti G, Zatsepin NA, Betti C et al (2015) Structural basis for bifunctional peptide recognition at human δ-opioid receptor. Nat Struct Mol Biol 22:265–268
26. Zhang H, Unal H, Gati C et al (2015) Structure of the angiotensin receptor revealed by serial femtosecond crystallography. Cell 161:833–844
27. Kang Y, Zhou XE, Gao X et al (2015) Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. Nature 523:561–567
28. Nogly P, James D, Wang D, White TA, Shilova A, Nelson G, Liu H, Johansson L (2015) Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. IUCrJ 2:168–176
29. Stellato F, Oberthür D, Liang M et al (2014) Room-temperature macromolecular serial crystallography using synchrotron radiation. IUCrJ 1:204–212
30. Roessler CG, Agarwal R, Allaire M et al (2016) Acoustic injectors for drop-on-demand serial femtosecond crystallography. Structure 24:631–640
31. Roessler CG, Kuczewski A, Stearns R, Ellson R, Olechno J, Orville AM, Allaire M, Soares AS, Héroux A (2013) Acoustic methods for high-throughput protein crystal mounting at next-generation macromolecular crystallographic beamlines. J Synchrotron Radiat 20:805–808
32. Soares AS, Mullen JD, Parekh RM, McCarthy GS, Roessler CG, Jackimowicz R, Skinner JM, Orville AM, Allaire M, Sweet RM (2014) Solvent minimization induces preferential orientation and crystal clustering in serial micro-crystallography on micro-meshes, in situ plates and on a movable crystal conveyor belt. J Synchrotron Radiat 21:1231–1239
33. Tsujino S, Tomizaki T (2016) Ultrasonic acoustic levitation for fast frame rate X-ray protein crystallography at room temperature. Sci Rep 6:25558
34. Hunter MS, Segelke B, Messerschmidt M et al (2014) Fixed-target protein serial microcrystallography with an X-ray free electron laser. Sci Rep 4:6026
35. Cohen AE, Soltis SM, González A et al (2014) Goniometer-based femtosecond crystallography with X-ray free electron lasers. Proc Natl Acad Sci U S A 111:17122–17127
36. Hirata K, Shinzawa-Itoh K, Yano N et al (2014) Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL. Nat Methods 11:734–736
37. Wierman JL, Alden JS, Kim CU, McEuen PL, Gruner SM (2013) Graphene as a protein crystal mounting material to reduce background scatter. J Appl Crystallogr 46:1501–1507
38. Warren AJ, Crawshaw AD, Trincao J, Aller P, Alcock S, Nistea I, Salgado PS, Evans G (2015) In vacuo X-ray data collection from graphene-wrapped protein crystals. Acta Crystallogr D Biol Crystallogr 71:2079–2088
39. Sui S, Wang Y, Kolewe KW, Srajer V, Henning R, Schiffman JD, Dimitrakopoulos C, Perry SL (2016) Graphene-based microfluidics for serial crystallography. Lab Chip. Advance article. doi:10.1039/C6LC00451B
40. Zarrine-Afsar A, Barends TRM, Müller C, Fuchs MR, Lomb L, Schlichting I, Miller RJD (2012) Crystallography on a chip. Acta Crystallogr D Biol Crystallogr 68:321–323
41. Murray TD, Lyubimov AY, Ogata CM, Vo H, Uervirojnangkoorn M, Brunger AT, Berger JM (2015) A high-transparency, micro-patternable chip for X-ray diffraction analysis of microcrystals under native growth conditions. Acta Crystallogr D Biol Crystallogr 71:1987–1997
42. Roedig P, Vartiainen I, Duman R et al (2015) A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering. Sci Rep 5:10451
43. Lyubimov AY, Murray TD, Koehl A et al (2015) Capture and X-ray diffraction studies of protein microcrystals in a microfluidic trap array. Acta Crystallogr D Biol Crystallogr 71:928–940
44. Baxter EL, Aguila L, Alonso-Mori R et al (2016) High-density grids for efficient data collection from multiple crystals. Acta Crystallogr D Biol Crystallogr 72:2–11
45. Coquelle N, Brewster AS, Kapp U, Shilova A, Weinhausen B, Burghammer M, Colletier JP (2015) Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. Acta Crystallogr D Biol Crystallogr 71:1184–1196

46. Coulibaly F, Chiu E, Ikeda K, Gutmann S, Haebel PW, Schulze-Briese C, Mori H, Metcalf P (2007) The molecular organization of cypovirus polyhedra. Nature 446:97–101
47. Cherezov V, Hanson MA, Griffith MT, Hilgart MC, Sanishvili R, Nagarajan V, Stepanov S, Fischetti RF, Kuhn P, Stevens RC (2009) Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 microm size X-ray synchrotron beam. J R Soc Interface 6(Suppl 5):S587–S597
48. Ji X, Sutton G, Evans G, Axford D, Owen R, Stuart DI (2010) How baculovirus polyhedra fit square pegs into round holes to robustly package viruses. EMBO J 29:505–514
49. Axford D, Ji X, Stuart DI, Sutton G (2014) In cellulo structure determination of a novel cypovirus polyhedrin. Acta Crystallogr D Biol Crystallogr 70:1435–1441
50. Zander U, Bourenkov G, Popov AN, de Sanctis D, Svensson O, AA MC, Round E, Gordeliy V, Mueller-Dieckmann C, Leonard GA (2015) *MeshAndCollect*: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines. Acta Crystallogr D Biol Crystallogr 71:2328–2343
51. Boudes M, Garriga D, Fryga A, Caradoc-Davies T, Coulibaly F (2016) A pipeline for structure determination of ıt in vivo-grown crystals using ıt in cellulo diffraction. Acta Crystallogr D Biol Crystallogr 72:576–585
52. Gati C, Bourenkov G, Klinge M et al (2014) Serial crystallography on in vivo grown microcrystals using synchrotron radiation. IUCrJ 1:87–94
53. Li L, Ismagilov RF (2010) Protein crystallization using microfluidic technologies based on valves, droplets, and SlipChip. Annu Rev Biophys 39:139–158
54. Kisselman G, Qiu W, Romanov V, Thompson CM, Lam R, Battaile KP, Pai EF, Chirgadze NY (2011) X-CHIP: an integrated platform for high-throughput protein crystallization and on-the-chip X-ray diffraction data collection. Acta Crystallogr D Biol Crystallogr 67:533–539
55. Dhouib K, Khan Malek C, Pfleging W et al (2009) Microfluidic chips for the crystallization of biomacromolecules by counter-diffusion and on-chip crystal X-ray analysis. Lab Chip 9:1412–1421
56. Pinker F, Brun M, Morin P et al (2013) ChipX: a novel microfluidic chip for counter-diffusion crystallization of biomolecules and in situ crystal analysis at room temperature. Cryst Growth Des 13:3333–3340
57. Perry SL, Guha S, Pawate AS, Bhaskarla A, Agarwal V, Nair SK, Kenis PJA (2013) A microfluidic approach for protein structure determination at room temperature via on-chip anomalous diffraction. Lab Chip 13:3183–3187
58. Khvostichenko DS, Schieferstein JM, Pawate AS, Laible PD, Kenis PJA (2014) X-ray transparent microfluidic chip for mesophase-based crystallization of membrane proteins and on-chip structure determination. Cryst Growth Des 14:4886–4890
59. Heymann M, Opthalage A, Wierman JL, Akella S, Szebenyi DME, Gruner SM, Fraden S (2014) Room-temperature serial crystallography using a kinetically optimized microfluidic device for protein crystallization and on-chip X-ray diffraction. IUCrJ 1:349–360
60. Jacquamet L, Ohana J, Joly J et al (2004) Automated analysis of vapor diffusion crystallization drops with an X-ray beam. Structure 12:1219–1225
61. Bingel-Erlenmeyer R, Olieric V, Grimshaw JPA et al (2011) SLS crystallization platform at beamline X06DA—a fully automated pipeline enabling in situ X-ray diffraction screening. Cryst Growth Des 11:916–923
62. Axford D, Owen RL, Aishima J et al (2012) In situ macromolecular crystallography using microbeams. Acta Crystallogr D Biol Crystallogr 68:592–600
63. Axford D, Foadi J, Hu N-J, Choudhury HG, Iwata S, Beis K, Evans G, Alguel Y (2015) Structure determination of an integral membrane protein at room temperature from crystals *in situ*. Acta Crystallogr D Biol Crystallogr 71:1228–1237
64. Gelin M, Delfosse V, Allemand F, Hoh F, Sallaz-Damaz Y, Pirocchi M, Bourguet W, Ferrer JL, Labesse G, Guichou JF (2015) Combining “dry” co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography. Acta Crystallogr D Biol Crystallogr 71:1777–1787
65. Axford D, Aller P, Sanchez-Weatherby J, Sandy J (2016) Applications of thin-film sandwich crystallization platforms. Acta Crystallogr F Struct Biol Commun 72:313–319
66. Cipriani F, Röwer M, Landret C, Zander U, Felisaz F, Márquez JA (2012) CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. Acta Crystallogr D Biol Crystallogr 68:1393–1399
67. Zander U, Hoffmann G, Cornaciu I et al (2016) Automated harvesting and processing of protein crystals through laser photoabla-

tion. Acta Crystallogr D Biol Crystallogr 72:454–466

68. Caffrey M (2015) A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallogr F Struct Biol Commun 71:3–18
69. Huang CY, Olieric V, Ma P, Panepucci E, Diederichs K, Wang M, Caffrey M (2015) In meso in situ serial X-ray crystallography of soluble and membrane proteins. Acta Crystallogr D Biol Crystallogr 71:1238–1256
70. Huang CY, Olieric V, Ma P et al (2016) In meso in situ serial X-ray crystallography of soluble and membrane proteins at cryogenic temperatures. Acta Crystallogr D Biol Crystallogr 72:93–112
71. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. Proc Natl Acad Sci U S A 108:16247–16252
72. Keedy DA, Kenner LR, Warkentin M et al (2015) Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. Elife 4:e07574
73. Leal RMF, Bourenkov G, Russi S, Popov AN (2013) A survey of global radiation damage to 15 different protein crystal types at room temperature: a new decay model. J Synchrotron Radiat 20:14–22
74. Owen RL, Paterson N, Axford D, Aishima J, Schulze-Briese C, Ren J, Fry EE, Stuart DI, Evans G (2014) Exploiting fast detectors to enter a new dimension in room-temperature crystallography. Acta Crystallogr D Biol Crystallogr 70:1248–1256
75. Henderson R (1990) Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. Proc R Soc Lond B 241:6–8
76. Owen RL, Rudiño-Piñera E, Garman EF (2006) Experimental determination of the radiation dose limit for cryocooled protein crystals. Proc Natl Acad Sci U S A 103:4912–4917
77. Evans G, Axford D, Owen RL (2011) The design of macromolecular crystallography diffraction experiments. Acta Crystallogr D Biol Crystallogr 67:261–270
78. Mueller M, Wang M, Schulze-Briese C (2012) Optimal fine φ-slicing for single-photon-counting pixel detectors. Acta Crystallogr D Biol Crystallogr 68:42–56
79. Dauter Z (1999) Data-collection strategies. Acta Crystallogr D Biol Crystallogr 55:1703–1717
80. Bourenkov GP, Popov AN (2006) A quantitative approach to data-collection strategies. Acta Crystallogr D Biol Crystallogr 62:58–64
81. Borek D, Minor W, Otwinowski Z (2003) Measurement errors and their consequences in protein crystallography. Acta Crystallogr D Biol Crystallogr 59:2031–2038
82. Liu ZJ, Chen L, Wu D, Ding W, Zhang H, Zhou W, Fu ZQ, Wang BC (2011) A multidataset data-collection strategy produces better diffraction data. Acta Crystallogr A 67:544–549
83. Weinert T, Olieric V, Waltersperger S et al (2015) Fast native-SAD phasing for routine macromolecular structure determination. Nat Methods 12:131–133
84. Brockhauser S, White KI, AA MC, RBG R (2011) Translation calibration of inverse-kappa goniometers in macromolecular crystallography. Acta Crystallogr A 67:219–228
85. Waltersperger S, Olieric V, Pradervand C et al (2015) PRIGo: a new multi-axis goniometer for macromolecular crystallography. J Synchrotron Radiat 22:895–900
86. Liu Q, Dahmane T, Zhang Z, Assur Z, Brasch J, Shapiro L, Mancia F, Hendrickson WA (2012) Structures from anomalous diffraction of native biological macromolecules. Science 336:1033
87. Olieric V, Weinert T, Finke AD et al (2016) Data-collection strategy for challenging native SAD phasing. Acta Crystallogr D Biol Crystallogr 72:421–429
88. Liu Q, Hendrickson WA (2015) Crystallographic phasing from weak anomalous signals. Curr Opin Struct Biol 34:99–107
89. Ayyer K, Philipp HT, Tate MW, Wierman JL, Elser V, Gruner SM (2015) Determination of crystallographic intensities from sparse data. IUCrJ 2:29–34
90. Holton JM (2009) A beginner's guide to radiation damage. J Synchrotron Radiat 16:133–142
91. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66:125–132
92. Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66:133–144
93. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr 67:271–281

94. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326
95. Brehm W, Diederichs K (2013) Breaking the indexing ambiguity in serial crystallography. Acta Crystallogr D Biol Crystallogr 70: 101–109
96. Arndt UW, Crowther RA, Mallett JF (1968) A computer-linked cathode-ray tube microdensitometer for X-ray crystallography. J Sci Instrum 1:510–516
97. Diederichs K, Karplus A (1997) Improved R-factors. Nat Struct Biol 4:269–275
98. Krojer T, von Delft F (2011) Assessment of radiation damage behaviour in a large collection of empirically optimized datasets highlights the importance of unmeasured complicating effects. J Synchrotron Radiat 18:387–397
99. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. Science 336:1030–1033
100. Diederichs K, Karplus PA (2013) Better models by discarding data? Acta Crystallogr D Biol Crystallogr 69:1215–1222
101. Karplus PA, Diederichs K (2015) Assessing and maximizing data quality in macromolecular crystallography. Curr Opin Struct Biol 34:60–68
102. Assmann G, Brehm W, Diederichs K (2016) Identification of rogue datasets in serial crystallography. J Appl Crystallogr 49:1021–1028
103. Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D Biol Crystallogr 66:479–485
104. Adams PD, Afonine PV, Bunkóczi G et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66:213–221