

Chapter 8

Phasing Serial Crystallography Data



Richard A. Kirian, Joe P. J. Chen, and John C. H. Spence

8.1 Introduction

The development of serial femtosecond crystallography (SFX) at X-ray free electron lasers (X-ray FELs) allows for the use of tiny protein crystals down to just a few unit cells along an edge, measured at physiological temperatures, and with a time resolution far better than can be achieved with synchrotrons or electron microscopes. The unique properties of the X-ray FEL source has furthermore resulted in the appearance of entirely new ideas for solving the crystallographic phase problem. At the same time, in combination with work on phasing single-particle data (with one bioparticle per shot), SFX has stimulated research into new phasing methods for serial crystallography (SC) at synchrotrons, and protein crystallography in general. In the sense that these new phasing methods depend on the application of constraints, they might be considered developments of traditional “direct methods” such as density modification approaches.

It was the injection of new ideas from the signal processing and optics communities in a formative review article by Millane [1] that initiated the modern revival of interest in numerical iterative phasing methods for both single particles and crystals. These ideas go back, at least, to the paper by Sayre [2], who first pointed out that if scattering could be detected between Bragg reflections it would assist phasing (since intensity zero-crossings could then be identified), and to work by Gerchberg and Saxton [3] on iterative phasing for non-periodic samples in electron microscopy. The first successful algorithm based on this approach added feedback to the Gerchberg and Saxton algorithm and was described as the Hybrid Input-Output (HIO) algorithm by Fienup [4]. These algorithms, which are reviewed

R. A. Kirian (✉) · J. P. J. Chen · J. C. H. Spence
Department of Physics, Arizona State University, Tempe, AZ, USA
e-mail: rkirian@asu.edu; jpchen1@mainex1.asu.edu; spence@asu.edu

further in Chaps. 9 and 14, iterate between real and reciprocal space while imposing known constraints at each step, such as the boundary of the molecule, the sign of the scattering medium, and the measured scattering intensity. Many variants of these “iterated projection” algorithms are the subject is reviewed by Marchesini [5], Millane and Lo [6] and Spence [7]. In most cases, scattering that was more finely sampled in angle than twice the Bragg angle (“oversampling”) was required. However, it is now understood that for many high-solvent crystals, where the molecule fills only a portion of the unit cell, one has access to sufficient information to solve the phase problem using only Bragg intensities, as we describe below in more detail. The control of hydration was also the basis for some early phasing efforts [8] aimed at sampling the molecular transform at several points by changing the unit cell volume. In Chap. 9, the production of diffuse scattering between Bragg reflections is described, which can also be used in this way. As we detail below, a means of determining if a unique solution can be expected for both single particles and crystals has now emerged through the introduction of a metric called the constraint ratio Ω [9]. At the same time that new algorithms were being considered, it was natural to try existing phasing methods with SFX data.

Below we begin with a brief overview of the phase problem and how conventional crystallographic phasing methods have been applied to SFX data (see also [10]). We then describe novel phasing techniques enabled by the unique properties of X-ray FELs.

8.2 The Phase Problem

Under the Born approximation, the far-field diffraction of an arbitrary object by X-rays is related to the object’s structure by the Fourier transform

$$F(\mathbf{q}) = \int_{-\infty}^{\infty} f(\mathbf{x}) e^{i\mathbf{q}\cdot\mathbf{x}} d\mathbf{x} \quad (8.1)$$

where \mathbf{x} and \mathbf{q} are coordinates in real and Fourier (reciprocal) space, respectively, i is the square root of -1 , $f(\mathbf{x})$ is the complex scattering density of the object, and $F(\mathbf{q})$ is the complex amplitude of the diffracted wavefield. To a first approximation, the scattering density of the object is proportional to its electron density, with the exception being cases where the X-ray energy is close to an electronic resonance (i.e., transition energy.) Resonance leads to a significant imaginary component of the scattering density and may be treated accurately with an explicit atomic model in which the total scattering factor is composed of a summation over atomic scattering factors.

The term “complex amplitude” means that $F(\mathbf{q})$ is a complex function, having a magnitude, $|F(\mathbf{q})|$, and a phase, $\varphi(\mathbf{q})$, and is expressible as:

$$F(\mathbf{q}) = |F(\mathbf{q})| e^{i\varphi(\mathbf{q})} \quad (8.2)$$

Knowing $F(\mathbf{q})$ allows one to obtain the scattering density $f(\mathbf{x})$ through the inverse Fourier transform:

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} F(\mathbf{q}) e^{-i\mathbf{x}\cdot\mathbf{q}} d\mathbf{q} \quad (8.3)$$

The lack of suitable materials to act as a lens for efficiently focusing scattered X-rays to form an image means that what is accessible in a diffraction experiment is just the intensity of the diffracted wavefront, which is the square of the magnitude of $F(\mathbf{q})$, that is, $|F(\mathbf{q})|^2$, and hence the phase function $\varphi(\mathbf{q})$ is not measured. This constitutes the so-called phase problem. In addition, individual diffraction patterns record scattering, which is constrained by the elastic scattering kinematics to lie on the Ewald sphere, so that these separate recordings must be indexed (oriented with respect to the lab frame), merged and assembled into the three-dimensional diffraction volume before these equations can be applied.

In crystallography, one is faced with an even more severe restriction than the absence of measured phases. The periodic nature of a crystal means that the diffracted intensity of the molecule is modulated by a periodic multiplicative term that peaks at the reciprocal lattice points and is of much lower values elsewhere in reciprocal space. For fully coherent illumination, this multiplicative function is often called the shape transform (see Sect. 8.4.1) since it depends strongly on the overall shape of the crystal. When the number of unit-cell repetitions is large, the shape-transform peaks become sharp and point-like, giving rise to the familiar concept of Bragg peaks, and effectively results in the diffracted intensity being sampled at only the reciprocal lattice points, which we sometimes refer to as “Bragg sampling.”

Without any phase information, it can be shown that this Bragg sampling of the diffracted intensity is below the minimum amount required by Shannon’s sampling theorem to uniquely determine the autocorrelation of the unit cell, which indicates that the Bragg sampling is also not sufficient to uniquely determine the unit cell itself (see [1, 11] for a more in-depth discussion.) This is illustrated in Fig. 8.1, which compares three diffraction scenarios: (1) an isolated non-crystalline molecule, (2) a crystal of finite size under coherent FEL-like illumination, and (3) an infinite crystal under similar coherent illumination. The intensity undersampling problem may be understood by first recalling that the Fourier transform of the diffraction intensities is equal to the autocorrelation of the real-space scattering density, and secondly by noting that only half of the Hermitian centrosymmetric autocorrelation function contains independent values. For the case of an infinite crystal, the essential problem is that the periodic autocorrelation function that results from Bragg-sampled intensities yields only half as many independent measurements as there are independent scattering densities to be determined. To solve the “phase problem” in conventional crystallography, one must therefore overcome the “undersampled-intensity problem.” The solution to this conundrum usually requires additional experimental measurements in which the X-ray wavelength is varied, or the atomic structure is modified in a controlled way, or by making some assumptions about the unknown structure (e.g., that it is similar to a molecule whose structure is already

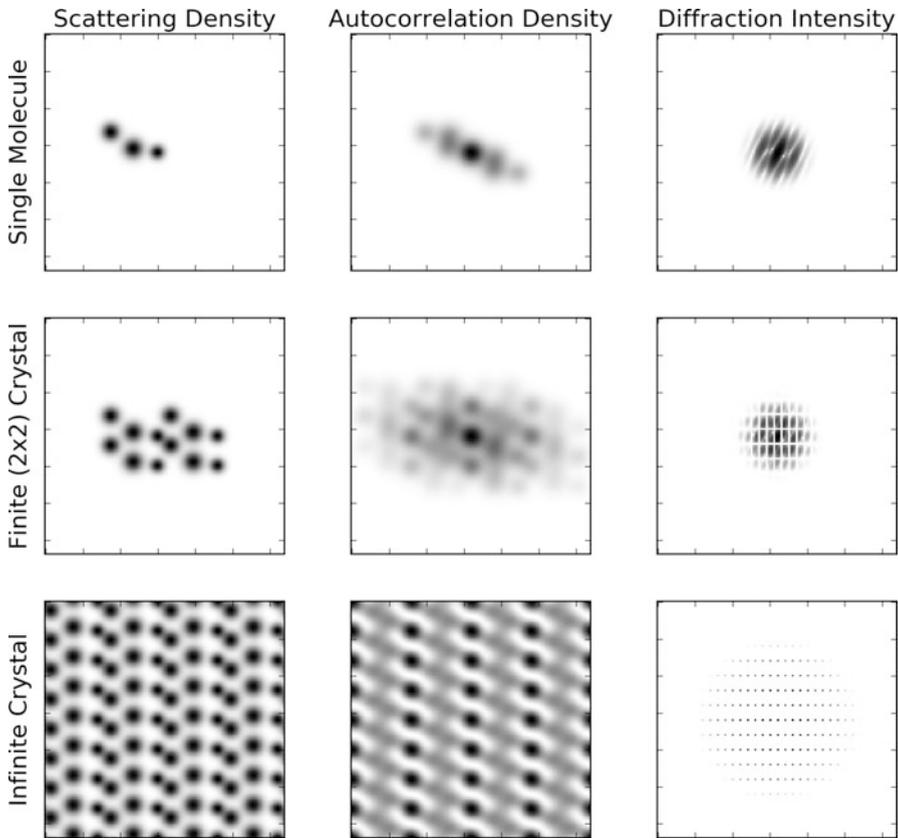


Fig. 8.1 Top row: scattering density of a single generic molecule (left), its autocorrelation function (middle), and diffraction intensities (right). The autocorrelation function is equal to the Fourier transform of the diffraction intensities, and hence is an equally valid representation of the data as the intensities, but more clearly reveals the number of independent measurements. Note the centrosymmetry of the autocorrelation function, in which only half of the non-zero area contain independent measurements. Middle row: the same molecule as in the top row, but arranged into a 2×2 crystal. In principle, a finite crystal may be treated much like any other object, if it were possible to measure appropriately sampled diffraction intensities at high signal-to-noise ratio. Bottom row: the same molecule arranged into an infinite crystal. In this case, the autocorrelation function has the same periodicity as the crystal itself, but due to the centrosymmetry there are only half as many independent measurements per unit cell as there are unknown scattering densities per unit cell; this is what gives rise to the crystallographic phase problem that suffers from a major deficit in the number of measurements

known). However, as to be discussed in Sect. 8.4, X-ray FELs offer some novel solutions to the phase problem in certain situations.

8.3 Conventional Methods

The standard crystallographic approaches to phasing are described in Rupp [12] where further details and references can be found. A brief summary of recent developments in de novo phasing in the context of X-ray FELs can be found in Schlichting [10]. Broadly speaking, all phasing techniques rely on increasing the number of unique measurements (beyond the Bragg-sampled intensities), and/or utilizing prior-known information about the sample. These include Single and Multiple Anomalous Diffraction (SAD and MAD), in which small differences due to anomalous scatterers are utilized; Single and Multiple Isomorphous Replacement (SIR and MIR), in which heavy atoms are placed in a protein crystal and thereby augment the diffraction intensities; Direct Methods, which use the likely zero sum of the phases around loops in reciprocal space; and Molecular Replacement (MR), which is based on modeling against macromolecules with similar sequence or fold in the Protein Data Bank (www.wwPDB.org). Of these methods, MR and SAD are the most popular de-novo technique over 70% of de novo structures deposited in the PDB in 2013 done by SAD [13], but the vast majority of all structures are now done by MR. The traditional Direct Methods require very high resolution data (atomic-resolution), and succeed typically only for small proteins.

It was not at all obvious that conventional phasing methods could be applied to “diffraction-before-destruction” SFX data, since the onset of radiation damage after a few tens of femtoseconds could mask the very small differences in structure factors that must be measured, for example, in the SAD method. In addition, the original Monte Carlo intensity merging method (without post-refinement procedures), described in detail in Chap. 7 of this book, has limited accuracy. Monte Carlo intensity merging requires a 100 times more data to obtain a single order of magnitude improvement in accuracy, as it averages partial reflection intensities over large numbers of crystals of different sizes. Shot-to-shot variations in the X-ray FEL beam intensity and wavelength are additional important stochastic variables to be accommodated in the Monte Carlo method. Despite issues with the stochastic nature of SFX measurements, it emerged that for micron-sized crystals, data collected using X-ray FEL pulses often showed higher resolution than synchrotron data from small crystals. The general trend seems to be that for microcrystals, radiation damage at synchrotrons results in lower resolution data than from X-ray FELs. For large crystals, the resolution achieved with synchrotrons may be better given sufficiently high-quality crystals (see [14] for a review of serial crystallography.)

The first report of SAD phasing with X-ray FEL data was carried out by Barends et al. [15], who recovered the structure of a heavy-atom derivative of lysozyme at 0.21 nm resolution. In this case, gadolinium atoms were used, which produced a relatively large anomalous difference signal of roughly 10%. Pulses of 50 fs duration, 2.6 mJ average power, and 8.5 keV photon energy were used to obtain 60,000 indexed diffraction patterns that yielded an anomalous correlation coefficient (CC_{ano}) of 0.48, with R_{split} about 5% at 0.3 nm resolution. The number of required patterns was later reduced to ~ 7000 upon improvements to the data processing, as

discussed in Nass et al. [16]. A series of de novo phasing demonstrations have since been reported: isomorphous replacement was demonstrated on a mercury derivative [17] followed by the more recent application of direct SAD phasing to the same target [18]. The replacement of methionine residues by seleno-methionine, known as the “magic bullet” of structural biology, is a convenient means of SAD phasing (at 12.65 keV) and has now been demonstrated on SFX data [18, 19] using both the LCLS and SACLA facilities. The first previously unknown structure to be solved was the mosquito larvicide BinAB, which was solved using multiple isomorphous replacement with mercury, gadolinium, and iodine on in vivo grown nanocrystals [20].

Recent research has also focused on SAD phasing using native sulfur atoms, which constitute about 1% of non-hydrogen atoms in proteins. Anomalous signal from sulfur atoms was observed quite early in the development of SFX by Barends et al. [21], but at 7.3 keV the anomalous difference signal from sulfur is small and phase retrieval was not possible at that time, partly due to software limitations. This has since proven possible in more recent work by Nass et al. [16]. Native SAD phasing using sulfur and chlorine is also demonstrated in Nakane et al. [22] for lysozyme, and by Batyuk et al. [23] for the G-protein coupled receptor (GPCR) adenosine receptor A_{2a}. These studies, in which the anomalous difference signal is small (about 1%), demonstrate the increasing accuracy of SFX data analysis due to steady improvements in the SFX data analysis algorithms and detector metrology. In addition to algorithm developments, further improvements can be expected from the FEL sources; for example, the application of a new two-color mode of operation at X-ray FELs has very recently been exploited to gain further improvements in MAD phasing [24].

8.4 Novel Approaches

8.4.1 *Finite Crystal Methods*

From the observation of the first SFX diffraction patterns in 2009, it was clear that new data analysis methods would be needed. This was a consequence of the new experimental arrangements—protein crystals as small as a few dozen unit cells on a side (or bioparticles) sprayed across a pulsed beam, as in the first serial crystallography variants [25, 26]. Many of the patterns showed scattering in the form of interference fringes between Bragg reflections, which, as suggested by Sayre [2], could assist with phasing. The theoretical basis for describing protein nanocrystal diffraction in serial crystallography, including these “shape transform” effects, was provided by Kirian et al. [27], and the many developments of this theory and improved algorithms that have followed are described in Chaps. 7 and 9. For the purposes of phasing, the new opportunities offered by X-ray FEL data can best be understood through the history of the development of the field

of Coherent Diffractive Imaging (CDI), which has been reviewed by Marchesini [5] and Spence [7]. In CDI, the electron density of an object is recovered from its Fourier intensities through iterative algorithms. These iterative solutions to the non-crystallographic phase problem work well for single-particle data, but, because they require scattering to be sampled at sub-Bragg intervals, were originally thought not to be useful for phasing scattering from crystals. It is now understood there are important exceptions to this conclusion, as we discuss here for the case of nanocrystals.

As alluded to in Sect. 8.2, measuring diffraction intensities between reciprocal lattice points can give enough information to uniquely determine the autocorrelation function of the crystal unit cell. Work by Perutz [28] aimed to determine the sufficiently sampled transform of the unit cell by modifying the solvent contents, and hence the unit cell size, of hemoglobin crystals. The term “sufficiently sampled” in this case means at least twice as fine as the Bragg-sampling. Perutz’s work was met with some success but his particular technique fell out of favor with the invention of isomorphous replacement soon after.

The new experimental arrangements provided by X-ray FELs yielded diffraction patterns that showed measureable intensities between the Bragg reflections, as shown in Fig. 8.2. For an idealized nanocrystal immersed in a wide coherent beam, one finds $(N-2)$ interference fringe maxima, for a crystal containing N planes normal to direction \mathbf{g} , running between Bragg reflections in the direction \mathbf{g} . This is akin to the $(N-2)$ subsidiary maxima seen between the principle maxima in the optical transmission diffraction pattern from a grating of N slits. These fringes, running in several directions, therefore give the size of the crystal between facets. The interference fringes occur because the coherence width of the beam is larger than the microcrystals, meaning that the entire crystal is coherently illuminated. In contrast, conventional Bragg diffraction requires only that the coherence width exceed the size of the unit cell. The above situation can be modeled as follows. Write the electron density of the n th crystal, $f_n(\mathbf{x})$, as

$$f_n(\mathbf{x}) = \sum_{j=1}^{N_n} f(\mathbf{x} - \mathbf{r}_{nj}) \quad (8.4)$$

where $f(\mathbf{x})$ is the electron density of the unit cell, \mathbf{r}_{nj} is the spatial shift for the j th unit cell required to construct the n th crystal which has N_n number of unit cells. Taking the Fourier transform, the far-field complex-valued diffraction amplitude of the crystal can be written as

$$F_n(\mathbf{q}) = F(\mathbf{q}) \sum_{j=1}^{N_n} e^{i\mathbf{q}\cdot\mathbf{r}_{nj}} \quad (8.5)$$

The quantity that is measured in the experiment is the intensity of the complex-valued diffracted amplitude, which is given by

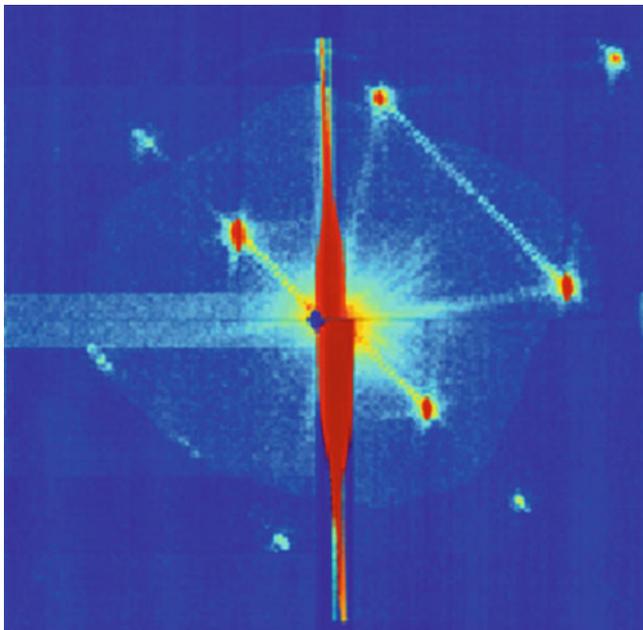


Fig. 8.2 A diffraction pattern obtained at LCLS from a submicrometer crystal of Photosystem I during the first serial femtosecond X-ray crystallography experiment [8]. The interference fringes between Bragg reflections provide the “oversampling” needed in principle to solve the phase problem. The red streak running vertically through the center of the pattern is the scattering of X-rays from the edge of the water jet that carried the crystals into the X-ray beam

$$I_n(\mathbf{q}) = |F(\mathbf{q})|^2 S_n(\mathbf{q}) \quad (8.6)$$

where

$$S_n(\mathbf{q}) = \left| \sum_{j=1}^{N_n} e^{i\mathbf{q} \cdot \mathbf{r}_{nj}} \right|^2 \quad (8.7)$$

is the so-called “shape transform.” Note that this simple product between the modulus-squared unit-cell transform and the shape transform results from the assumption that all unit cells are identical, which may not be the case, as discussed later.

Consider now the average over many diffraction patterns, measured from an ensemble of crystals of different sizes and shapes, which is the result of the data merging step in an SFX experiment. This averaged intensity is given by

$$\langle I_n(\mathbf{q}) \rangle_n = |F(\mathbf{q})|^2 \langle S_n(\mathbf{q}) \rangle_n \quad (8.8)$$

where $\langle \cdot \rangle_n$ denotes the average over the entire ensemble of crystals. Spence et al. [30] suggested that the averaged shape transform $\langle S_n(\mathbf{q}) \rangle_n$ can be determined from $\langle I_n(\mathbf{q}) \rangle_n$ directly by averaging over all translations centered around the reciprocal lattice point \mathbf{g}_h from the diffraction of all crystals. The vector \mathbf{h} denotes a 3-tuple containing the Miller indices. By assuming that the molecular transform and the shape transform are uncorrelated, and that a sufficient number of reciprocal lattice points and crystals are used in forming this average, we can write

$$\langle S_n(\mathbf{q}) \rangle_n = \langle \langle I_n(\mathbf{q} - \mathbf{g}_h) \rangle_n \rangle_h \quad (8.9)$$

In other words, the operation described by Eq. (8.9) is the average of the diffracted intensities over all Wigner–Seitz cells (i.e., the smallest primitive unit cell that can be constructed in reciprocal space); such an operation produces one period of the averaged shape transform. The averaged shape transform over all reciprocal space can then be obtained by replicating the averaged period throughout reciprocal space.

The molecular transform can in principle be obtained via a simple division, needing only the merged experimental intensity, that is,

$$|F(\mathbf{q})|^2 = \frac{\langle I_n(\mathbf{q}) \rangle_n}{\langle \langle I_n(\mathbf{q} - \mathbf{g}_h) \rangle_n \rangle_h} \quad (8.10)$$

This resulting molecular transform is more finely sampled than the Bragg diffraction from conventional crystallography, thus compensating for the data deficiency impeding the solution of the phase problem from Bragg reflections alone. Having recovered a sufficiently sampled molecular transform, it can be phased in ways analogous to reconstructions in CDI by computational iterative methods, such as the HIO algorithm described in Sect. 8.1.

Simulations applying this method to ideal nanocrystals with P1 symmetry can be found in Spence et al. [30]. For all other space groups there exists more than one molecule in the unit cell, which needs to be taken into account. Unlike in conventional crystallography, the inter-Bragg intensities from finite crystals depend crucially on the electron density of the whole crystal rather than the electron density of the unit cell alone. Even for an idealized finite crystal, the way in which molecules terminate on the crystal surface determines which repeating unit cell the crystal is composed of. The usual point-group symmetries associated with Bragg reflections do not generally carry over to the inter-Bragg intensities, and the very notion of the unit cell breaks down because partial unit cells, with incomplete molecular occupancies, are very likely to occur at the crystal surface. With the above in mind, the diffraction intensity clearly cannot be written in terms of a simple product between a shape transform and a molecular transform. This problem, which requires important modifications to current phase-retrieval algorithms, is the subject of active research [31–34].

An experimental test of shape-transform phasing is shown in Fig. 8.3 (from [35]). In these experiments conducted at the FERMI facility at 32.5 nm wavelength, 2D crystals were formed from Pt islands deposited using a focused ion beam

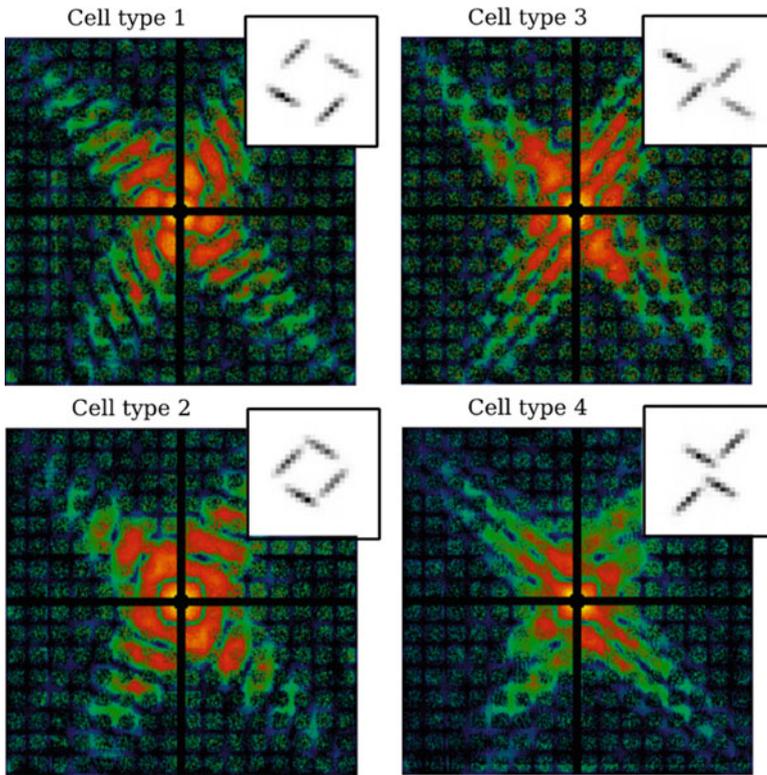


Fig. 8.3 Molecular transforms and real-space reconstructions (upper-right insets) corresponding to four different types of synthetic micro-crystals patterned with a focused ion beam. Each of the four crystal types differed only in their unit-cell configurations (i.e., only the surface truncations differed). The molecular transforms were recovered after averaging the diffraction intensities from multiple crystals with differing shapes and sizes, followed by a procedure to de-couple the crystal lattice transforms as described in the main text. Real-space reconstructions followed conventional CDI methods while utilizing intensities sampled between Bragg reflections [29]

(FIB) on a thin substrate. The same motif was used with four different edge terminations in order to clearly demonstrate the importance of the crystal surface. These experiments vindicate how the procedure of averaging over many shots (a couple dozen in this case), followed by division of the resulting average pattern by the periodically averaged Wigner–Seitz cell, reveals the underlying molecular transform of the crystal. It was also demonstrated that the X-ray beam need not be highly uniform in phase or amplitude over the entire crystal; it is sufficient if the X-ray beam is reasonably uniform over length scales corresponding to a few unit cells since the aim is to reconstruct the unit cell rather than the whole crystal. Some regions between Bragg reflections had very poor signal-to-noise ratio and were left unconstrained during the reconstruction, which is tolerable since, roughly speaking, it is only necessary to double the number of intensity measurements as compared to

Bragg sampling. The main limitation of this experiment is that the crystals had well-defined unit cells, whereas inter-Bragg diffraction observed from protein crystals in space groups other than P1, thus far, appear to arise from crystals that have surface truncations that are not consistent with a common unit cell throughout. As mentioned above, the surface terms remain an important challenge for practical implementations of shape-transform phasing.

For a diffraction-limited coherent beam of nanometer dimensions, the situation is analogous to that in the fully coherent scanning transmission electron microscope (STEM). If the beam divergence angle is larger than the Bragg angle, these coherent diffraction orders overlap at the detector, producing interference fringes that depend on the absolute position of the beam with respect to the crystal lattice, and may be analyzed according to the theory of ptychography for hard X-rays [36].

8.4.2 Intensity Variation Methods

Many phasing methods rely on an ability to record diffraction patterns before and after changing the scattering strength of just one species in a crystal, at a known site. In SIR/MIR, these are the heavy atom replacements; in MAD, the scattering factor is changed by varying the X-ray photon energy to alter the scattering potential. A related technique known as “radiation-induced-phasing” (RIP) [37] utilizes site-specific alterations to protein structures that may be induced by photoabsorption of UV light (for example). For the X-ray FEL, it is possible to create dramatic changes in scattering factors through the ionization of heavy atoms in proteins, which are virtually stationary during femtosecond pulses, and to use this “high-intensity MAD” effect for phasing. In this case, the X-ray diffraction is described by a time integral of the diffracted intensities throughout the duration of the exposure, during which many stochastic ionizations take place. The diffraction of a single shot takes the approximate form

$$I(\mathbf{q}) = \int dt I_0(t) \left| \sum_n f_n(t) e^{i\mathbf{q}\cdot\mathbf{r}_n} \right|^2 \quad (8.11)$$

where $I_0(t)$ is the time-dependent incident intensity, and $f_n(t)$ is the time-dependent atomic scattering factor of atom n located at position \mathbf{r}_n . The single-shot diffraction must in turn be averaged probabilistically over all of the many configurations that are possible as a result of sequential ionizations, which leads to a unique form of partially coherent diffraction. A detailed theoretical treatment is presented by Son et al. [38], which provides a generalization of the Karle–Hendrickson equations for the high-intensity “diffraction-during-ionization” regime. Son et al. provide detailed calculations for the case in which a single atomic species (Fe) is independently ionized and the remainder of the atoms are assumed to diffract normally without ionization. The four coefficients of the generalized Karle–Hendrickson equations are shown as a function of photon energy and pulse fluence nearby the Fe absorption

edge, which reveals how the contrast in these coefficients can be enhanced with increasing fluence.

Although high-intensity MAD phasing has not yet been demonstrated, a selective reduction in scattering power of sulfur atoms (with 2.47 keV K-edge) in protein was demonstrated by Galli et al. [39] by recording diffraction patterns with low fluence and high fluence (to fully ionize the sulfur and so reduce its scattering power.) This was done using 6 keV X-rays, allowing for high-resolution data to be collected, unlike a SAD study at 2.47 keV, where resolution may be limited by the 0.5 nm X-ray wavelength. Applications to a Gd derivative of lysozyme and to Cathepsin B, utilizing local electronic damage, are described in Galli et al. [40] and [41] respectively.

8.4.3 *Constraint Ratios and High-Solvent Crystals*

It was appreciated at an early stage from Pauling's work on the mineral bixbyite ($(\text{Mn,Fe})_2\text{O}_3$) that the solution of the phase problem in crystallography is not unique, as for the class of homometric structures [42]. These are different crystal structures with the same Patterson (autocorrelation) function that therefore produce the same diffracted Bragg intensities. In practice, unique solutions can be expected if the number of independent Fourier equations relating measured Bragg intensities to crystal density is at least equal to the number of unknown phases [43–45]. This can be achieved by surrounding the molecule by an equal volume of known density. The subject is reviewed by Millane and Lo [6], who also include the effects of non-crystallographic symmetry. Elser and Millane [9] considered continuous diffraction from an isolated molecule and defined a constraint ratio Ω as the ratio of the number of independent diffraction intensity samples to the number of real space electron density samples. As alluded to earlier in Sect. 8.2, the number of independent diffracted intensity samples is limited by the Shannon sampling theorem, which gives the critical spacing between intensity samples in Fourier space. At larger spacings, the autocorrelation of the object cannot be uniquely recovered, yet finer spacings do not yield further information. Elser and Millane [9] showed that the constraint ratio can be defined as

$$\Omega = A/(2U) \tag{8.12}$$

where U is the volume of the support (the region in real space occupied by the molecule) and A is the volume of the autocorrelation function of the molecule (the 3D Fourier transform of the diffracted intensities), thus Ω depends on molecular shape, since the region occupied by the autocorrelation function depends on the shape of the molecule [46]. The factor of two in the denominator in the definition of Ω comes about because the autocorrelation function for a molecule is centrosymmetric, as described in Sect. 8.2. A unique solution to the phase problem requires $\Omega > 1$, being a necessary but not sufficient condition. The lower bound for Ω in

3D is 4, arising from support regions that are convex and centrosymmetric. Hence we see that the phase problem is highly over-constrained for continuous (diffuse) scattering that has been merged into a three-dimensional diffraction volume. These results can be related to the required angular sampling interval in diffraction experiments by recalling that, along one dimension, Shannon's theorem requires this continuous scattering to be sampled at intervals of $\Delta\theta = \lambda/L$ (in the small angle scattering approximation) for complete recovery of the density, where L , the width of the autocorrelation function, is twice the width of the molecule. Data is collected as 2D diffraction patterns, possibly affected by the curvature of the Ewald sphere so that they do not correspond to projections. However this data can be merged into a 3D volume to avoid issues associated with the curved sphere.

The analysis of the constraint ratio has been adapted to crystals by Millane and Arnal [47], where in the case of crystals, the autocorrelation function, about twice the size of the molecule in a unit cell, may overlap, so that A must be replaced by the volume V of the unit cell (the unique volume in the periodic Patterson map is half of this volume). The number of independent diffraction data and the number of electron density samples must also be reduced by the order of the space group. The previous equation then becomes

$$\Omega = R/(2f) \quad (8.13)$$

where $f = U/V$ is the fraction of the unit cell occupied by the molecule, V is the volume of the unit cell, and the crystal shows R -fold non-crystallographic symmetry. The constraint ratio for a crystal now depends only on the volume of the molecule, and not its shape. The phase problem may thus in principle be uniquely solved via Bragg intensities alone in cases where the unit cell contains more than 50% solvent in the absence of any non-crystallographic symmetry. In the presence of noise, values of Ω larger than unity are desirable.

Recent experiments using this method at 0.2 nm resolution suggest that a solvent content of greater than 65% is needed [48], however a lower fraction of solvent may be possible in crystals with larger values of R . By applying the Fienup HIO algorithm to crystals with high solvent contents, He and Su [49] have successfully phased a number of protein crystals. He et al. [50] have combined this approach with the Molecular Replacement method and applied it to three datasets.

8.4.4 Two-Dimensional Crystals

The weak scattering power of organic monolayers and two-dimensional crystals makes it extremely difficult to record diffraction patterns in transmission from these structures at synchrotrons, despite the importance of membranes in structural biology. They have been extensively studied by transmission electron diffraction, where imaging solves the phase problem. Here compact support along the beam direction can provide a useful constraint for iterative phasing [51]. Because a much

higher dose can be applied without damage affecting the measured diffraction when using femtosecond X-ray FEL pulses, experimental patterns have now been published from 2D crystals by this method [52]. These patterns, from streptavidin and bacteriorhodopsin, were obtained without cryogenic cooling, and extend to about 0.8 nm resolution (although the data are measured to the edge of the detector). A full diffraction dataset requires the collection of a tilt series (crystal rotated about the normal to the beam), since reciprocal space for a 2D crystal consists of sharp rods running normal to the monolayer at each 2D lattice site. The phase problem in this case has been analyzed in terms of the constraint ratio formalism by Arnal and Millane [1] and Arnal et al. [53], who find that a smaller solvent content is needed than for 3D crystals in order to achieve uniqueness. They also discuss the helpful phasing effect of pores in a membrane, such as aquaporin 1 (AQP1) since the pores reduce the number of unknowns in real space while the volume of the autocorrelation of the unit cell is unchanged, keeping the amount of known Fourier space data constant. Once again, additional constraints, such as non-crystallographic symmetry and histogram matching of density map grey-levels, can assist.

8.4.5 Charge-Flipping and Atomic Resolution Data

A new iterative phasing algorithm for crystals appeared in 2004, in which the real-space operation consists of reversing the sign of the charge density at any pixel where it has a value less than some threshold δ (the only adjustable parameter in the algorithm) [54]. The algorithm starts with random phases satisfying Friedel's law. Experimental Fourier magnitudes are imposed in reciprocal space, and the phases are retained. It has been applied to many experimental datasets and modified for use with powder diffraction data, where it assists in resolving overlapping peaks and provides composition and space-group information [55]. The algorithm has been shown to be equivalent to an Output-Output algorithm in the Fienup scheme, with feedback parameter $\beta = -2$. For a review of the charge-flipping algorithm, see Oszlanyi and Suto [56].

The method requires atomic-resolution data so that, like Direct Methods, the space around atoms can be used as a support, since crystals consist mostly of vacuum. This once again relates to the idea that the phase problem is soluble from the Fourier intensities alone if the constraint ratio is greater than unity, which in this case arises from the diffraction data being of high enough resolution. As the quality and resolution of data from X-ray FELs continues to improve, we can expect this algorithm, and several others, which require atomic-resolution data, to become increasingly useful.

8.5 Conclusions

Imaging techniques based on X-ray FELs will continue to complement other techniques well into the future. X-ray FELs offer the advantage over synchrotrons and cryo-electron microscopy (cryo-EM) that samples can be studied under physiologically relevant temperatures without appreciable damage. These techniques may be extended to time-resolved “pump–probe” variants that enable biomolecular dynamics to be studied with an explicit time delay between the “pump” mechanism and the X-ray probe, with virtually no time-delay limit imposed by the X-ray source. This pump–probe method differs from cryo-EM studies of dynamics that are based on rapidly quenched equilibrium ensembles of molecules, whose images may subsequently be sorted by similarity, and which can provide an energy landscape [57] rather than time-resolved imaging. Recently we have seen the first such studies of gene expression from a virus by the single-particle X-ray FEL method [58].

Unlike the closely related technique of coherent diffractive imaging, conventional crystallographic imaging suffers from undersampled diffraction intensities. Conventional de novo phasing techniques have been applied to X-ray FEL data and the use of these techniques will continue to increase as analysis algorithms and software improve. Additionally, new opportunities for phasing X-ray diffraction data from both single particles and crystals have been enabled by the X-ray FEL. These new ideas, discussed above, show promise but most have not yet been applied to the solution of novel protein structures. This is partly because considerable experimental and theoretical obstacles still need to be overcome, but perhaps also because new techniques simply take time to catch on. With the increasing appearance of atomic-resolution XFEL data, the charge-flipping algorithm and its developments are expected to become more popular for de novo phasing since it does not require the use of a model or chemical modification to the sample. It will be interesting to see what kind of improvements may result from the use of large datasets from nanocrystals that are smaller than a single mosaic block, using coherent X-ray beams of a similar size.

Acknowledgments Supported by NSF STC BioXFEL award 1231306.

References

1. Millane, R. P. (1990). Phase retrieval in crystallography and optics. *Journal of the Optical Society of America A*, 7(3), 394–411.
2. Sayre, D. (1952). Some implications of a theorem due to Shannon. *Acta Crystallographica*, 5(6), 843–843.
3. Gerchberg, R. W., & Saxton, W. O. (1972). A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik (Stuttgart)*, 35, 237–246.
4. Fienup, J. R. (1982). Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15), 2758–2769.

5. Marchesini, S. (2007). A unified evaluation of iterative projection algorithms for phase retrieval. *The Review of Scientific Instruments*, 78, 011301.
6. Millane, R. P., & Lo, V. L. (2013). Iterative projection algorithms in protein crystallography. *Acta Crystallographica Section A: Foundations and Advances*, A69, 517.
7. Spence, J. C. H. (2017b). In P. Hawkes & J. C. H. Spence (Eds.), *Science of microscopy* (pp. 1196–1227). New York: Springer.
8. Bragg, L., & Perutz, M. F. (1952). The structure of haemoglobin. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 213(1115), 425–435 The Royal Society.
9. Elser, V., & Millane, R. P. (2008). Reconstruction of an object from its symmetry-averaged diffraction pattern. *Acta Crystallographica Section A: Foundations of Crystallography*, 64(2), 273–279.
10. Schlichting, I. (2017). Experimental phasing of serial femtosecond crystallography data. *IUCrJ*, 4, 517–517.
11. Thibault, P., & Elser, V. (2010). X-ray diffraction microscopy. *Annual Review of Condensed Matter Physics*, 1, 237–255.
12. Rupp, B. (2010). *Biomolecular crystallography: Principles, practice, and application to structural biology*. New York: Garland Science.
13. Hendrickson, W. A. (2014). Anomalous diffraction in crystallographic phase evaluation. *Quarterly Reviews of Biophysics*, 47(1), 49–93. <https://doi.org/10.1017/S0033583514000018>.
14. Spence, J. C. H. (2017). XFELs for structure and dynamics in biology. *IUCrJ*, 4, 322–339.
15. Barends, T. R., Foucar, L., Shoeman, R. L., Bari, S., Epp, S. W., Hartmann, R., et al. (2013a). Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser. *Acta Crystallographica Section A: Foundations and Advances*, D69, 838–842.
16. Nass, K., Meinhardt, A., Barends, T. R., Foucar, L., Gorel, A., Aquila, A., et al. (2016). Protein structure determination by single-wavelength anomalous diffraction phasing of X-ray free-electron laser data. *IUCrJ*, 3(3), 180–191.
17. Yamashita, K., Pan, D., Okuda, T., Sugahara, M., Kodan, A., Yamaguchi, T., et al. (2015). An isomorphous replacement method for efficient de novo phasing for serial femtosecond crystallography. *Scientific Reports*, 5, 14017.
18. Yamashita, K., Kuwabara, N., Nakane, T., et al. (2017). Experimental phase determination with selenomethionine or mercury-derivatization in serial femtosecond crystallography. *IUCrJ*, 4, 639–647.
19. Hunter, M. S., Yoon, C. H., DeMirci, H., Sierra, R. G., Dao, E. H., Ahmadi, R., et al. (2016). Selenium single-wavelength anomalous diffraction de novo phasing using an X-ray-free electron laser. *Nature Communications*, 7, 13388.
20. Colletier, J.-P., Sawaya, M. R., Gingery, M., et al. (2016). De novo phasing with X-ray laser reveals mosquito larvicide BinAB structure. *Nature*, 539, 43–47. <https://doi.org/10.1038/nature19825>.
21. Barends, T. R. M., et al. (2013b). De novo protein crystal structure determination from X-ray free-electron laser data. *Nature*, 505, 244–247.
22. Nakane, T., Song, C., Suzuki, M., Nango, E., Kobayashi, J., Masuda, T., et al. (2015). Native sulfur/chlorine SAD phasing for serial femtosecond crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 71(12), 2519–2525.
23. Batyuk, A., Galli, L., Ishchenko, A., Han, G. W., Gati, C., Popov, P. A., et al. (2016). Native phasing of x-ray free-electron laser data for a G protein–coupled receptor. *Science Advances*, 2(9), e1600292.
24. Gorel, A., Motomura, K., Fukuzawa, H., Doak, R. B., Grünbein, M. L., Hilpert, M., et al. (2017). Multi-wavelength anomalous diffraction de novo phasing using a two-colour X-ray free-electron laser with wide tunability. *Nature Communications*, 8, 1170. <https://doi.org/10.1038/s41467-017-00754-7>.
25. Shapiro, D. A., Chapman, H. N., DePonte, D., Doak, R. B., Fromme, P., Hembree, G., et al. (2008). Powder diffraction from a continuous microjet of submicrometer protein crystals. *Journal of Synchrotron Radiation*, 15(6), 593–599.

26. Spence, J. C., & Doak, R. B. (2004). Single molecule diffraction. *Physical Review Letters*, 92(19), 198102.
27. Kirian, R. A., Wang, X., Weierstall, U., et al. (2010). Femtosecond protein nanocrystallography - data analysis methods. *Optics Express*, 18(6), 5713–5723.
28. Perutz, M. F. (1954). The structure of haemoglobin. III. Direct determination of the molecular transform. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 225(1161), 264–286 The Royal Society.
29. Chapman, H. N., Fromme, P., Barty, A., et al. (2011). Femtosecond X-ray protein nanocrystallography. *Nature*, 470, 73–77.
30. Spence, J. C. H., Kirian, R. A., Wang, X. Y., Weierstall, U., Schmidt, K. E., White, T., et al. (2011). Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals. *Optics Express*, 19(4), 2866–2873. <https://doi.org/10.1364/Oe.19.002866>.
31. Chen, J. P. J., Arnal, R. D., Morgan, A. J., Bean, R. J., Beyerlein, K. R., Chapman, H. N., et al. (2016). Reconstruction of an object from diffraction intensities averaged over multiple object clusters. *Journal of Optics*, 18(11), 114003.
32. Chen, J. P. J., & Millane, R. P. (2014). Diffraction by nanocrystals II. *Journal of the Optical Society of America A*, 31(8), 1730–1737.
33. Elser, V. (2013). Direct phasing of nanocrystal diffraction. *Acta Crystallographica Section A: Foundations of Crystallography*, 69(6), 559–569.
34. Kirian, R. A., Bean, R. J., Beyerlein, K. R., Yefanov, O. M., White, T. A., Barty, A., et al. (2014). Phasing coherently illuminated nanocrystals bounded by partial unit cells. *Philosophical Transactions of the Royal Society B*, 369(1647), 20130331.
35. Kirian, R. A., Bean, R. J., Beyerlein, K. R., Barthelmeß, M., Yoon, C. H., Wang, F. L., et al. (2015). Direct phasing of finite crystals illuminated with a free-electron laser. *Physical Review X*, 5(1), 011015.
36. Spence, J. C. H., Zatsepin, N. A., & Li, C. (2014). Coherent convergent-beam time-resolved X-ray diffraction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1647), 20130325.
37. Ravelli, R., Leiros, H., Pan, B. C., Caffrey, M., & McSweeney, S. (2003). Specific radiation damage can be used to solve macromolecular crystal structures. *Structure/Folding and Design*, 11(2), 217–224.
38. Son, S. K., Chapman, H. N., & Santra, R. (2011). Multiwavelength anomalous diffraction at high X-ray intensity. *Physical Review Letters*, 107(21), 218102.
39. Galli, L., Son, S. K., Klinge, M., Bajt, S., Barty, A., Bean, R., et al. (2015a). Electronic damage in S atoms in a native protein crystal induced by an intense X-ray free-electron laser pulse. *Structural Dynamics*, 2(4), 041703.
40. Galli, L., Son, S. K., Barends, T. R., White, T. A., Barty, A., Botha, S., et al. (2015b). Towards phasing using high X-ray intensity. *IUCrJ*, 2(6), 627–634.
41. Galli, L., Son, S. K., White, T. A., Santra, R., Chapman, H. N., & Nanao, M. H. (2015c). Towards RIP using free-electron laser SFX data. *Journal of Synchrotron Radiation*, 22(2), 249–255.
42. Hosemann, R., & Bagchi, S. N. (1954). On homometric structures. *Acta Cryst*, 7, 237–241.
43. Bricogne, G. (1974). Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Crystallographica Section A: Foundations and Advances*, A30, 395–405.
44. Crowther, R. A. (1969). The use of non-crystallographic symmetry for phase determination. *Acta Cryst*, B25, 2571–2580.
45. Miao, J., Sayre, D., & Chapman, H. N. (1998). Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *Journal of the Optical Society of America A*, 15(6), 1662–1669.
46. Thibault, P. (2007). Algorithmic methods in diffraction microscopy. PhD Thesis, Cornell University.
47. Millane, R. P., & Arnal, R. D. (2015). Uniqueness of the macromolecular crystallographic phase problem. *Acta Crystallographica Section A: Foundations and Advances*, A71, 592–598.

48. Liu, Z.-C., Xu, R., & Dong, Y.-H. (2012). Phase retrieval in protein crystallography. *Acta Crystallographica Section A: Foundations of Crystallography*, *A68*, 256–265.
49. He, H., & Su, W. P. (2015). Direct phasing of protein crystals with high solvent content. *Acta Crystallographica Section A, Foundations and Advances*, *71*(Pt 1), 92–98.
50. He, H., Fang, H., Miller, M. D., Phillips Jr., G. N., & Su, W. P. (2016). Improving the efficiency of molecular replacement by utilizing a new iterative transform phasing algorithm. *Acta Crystallographica Section A: Foundations and Advances*, *72*(5), 539–547.
51. Spence, J. C. H., Weierstall, U., Fricke, T. T., Glaeser, R. M., & Downing, K. (2003). *Journal of Structural Biology*, *144*, 209–218.
52. Frank, M., Carlson, D. B., Hunter, M. S., Williams, G. J., Messerschmidt, M., Zatsepin, N. A., et al. (2014). Femtosecond X-ray diffraction from two-dimensional protein crystals. *IUCrJ*, *1*(2), 95–100.
53. Arnal, R. D., & Millane, R. P. (2017). The phase problem for two-dimensional crystals. I. Theory. *Acta Crystallographica Section A: Foundations and Advances*, *73*, 438–448.
54. Oszlanyi, G., & Suto, A. (2004). Ab initio structure solution by charge flipping. *Acta Crystallographica Section A: Foundations and Advances*, *A60*, 134.
55. Wu, J., Leinenweber, K., Spence, J. C., & O’Keeffe, M. (2006). Ab initio phasing of X-ray powder diffraction patterns by charge flipping. *Nature Materials*, *5*(8), 647–652.
56. Oszlanyi, G., & Suto, A. (2008). The charge flipping algorithm. *Acta Crystallographica Section A: Foundations of Crystallography*, *A64*, 123–134.
57. Dashti, A., Schwander, P., Langlois, R., Fung, R., Li, W., Hosseinizadeh, A., et al. (2014). Trajectories of the ribosome as a Brownian nanomachine. *PNAS*, *111*(49), 17492–17497.
58. Hosseinizadeh, A., Mashayekhi, G., Copperman, J., et al. (2017). Conformational landscape of a virus by single-particle X-ray scattering. *Nature Methods*, *14*, 877–881.