Research

Royal Society Publishing

Phasing coherently illuminated nanocrystals bounded by partial unit cells

Richard A. Kirian, Richard J. Bean, Kenneth R. Beyerlein, Oleksandr M. Yefanov, Thomas A. White, Anton Barty and Henry N. Chapman

Center for Free-Electron Laser Science, Deutsches Elektronen-Synchrotron, Notkestrasse 85, Hamburg 22607, Germany

With the use of highly coherent femtosecond X-ray pulses from a free-electron laser, it is possible to record protein nanocrystal diffraction patterns with far more information than is present in conventional crystallographic diffraction data. It has been suggested that diffraction phases may be retrieved from such data via iterative algorithms, without the use of *a priori* information and without restrictions on resolution. Here, we investigate the extension of this approach to nanocrystals with edge terminations that produce partial unit cells, and hence cannot be described by a common repeating unit cell. In this situation, the phase problem described in previous work must be reformulated. We demonstrate an approximate solution to this phase problem for crystals with random edge terminations.

1. Introduction

The availability of brief, intense and coherent X-ray pulses produced by X-ray free-electron lasers (XFELs) has created the potential for major advancements in macromolecular crystallography [1,2]. Serial femtosecond crystallography (SFX) [3] is among the most successful new paradigms to emerge, which involves directing a stream of randomly oriented protein crystals across the focus of the XFEL beam. SFX data often consist of hundreds of thousands of diffraction patterns, which can be collected in a matter of hours at current pulse-repetition rates. These diffraction patterns are largely free from radiation damage [4] because the timescales of the relevant damage mechanisms are longer than the exposure time [5,6]. This creates, for example, new possibilities to study irreversible dynamic systems [7], radiation-sensitive targets [8] and macromolecules that only form small (approx. 1 μm) crystals and therefore pose difficulties for conventional synchrotron facilities [9].

As with conventional crystallography, the well-known 'phase problem' must be solved in order to reconstruct a real-space electron density map from the measured SFX intensities. Following data processing and reduction [10], some conventional macromolecular crystallography techniques can be readily applied to SFX data. For example, initial phase estimates have been obtained from SFX intensities by molecular replacement, which is effective when a similar known structure is available. On the other hand, *ab initio* phasing methods such as (multiple) isomorphous replacement or multi/single-wavelength anomalous dispersion are under development for XFEL sources [11].

Although a range of effective crystallographic phasing methods exist, there remains a general need for *ab initio* methods that apply to large molecules at medium or low resolution and do not require similar known structures, isomorphous derivatives or resonant diffraction [12]. It was recently suggested that coherently illuminated nanocrystals can provide sufficient information for *ab initio* phasing [13]. The method likely requires diffraction patterns from hundreds of thousands of individual sub-micrometre crystals, which may be obtained via the SFX technique. In the following sections, we consider complications with this approach that arise when the crystals considered in the ensemble do not

terminate exactly at the nominal unit cell boundaries, a considerable problem that was not addressed in previous work. Whereas crystal size and shape distributions are of relatively little consequence, we show that the presence of molecular vacancies at the crystal boundaries obscure the notion of the crystal unit cell and necessitates a reformulation of the problem. In this manuscript, we suggest an approximate means of solving this problem, which we demonstrate through simulations.

2. Phasing methods for coherently illuminated nanocrystals

Among the most striking observations made during the first SFX experiments in 2009 were the distinct intensity distributions observed around Bragg peaks, attributed to the finite lattice of the crystal. These so-called finite-lattice transforms arise when the coherence length of the illumination spans the full width of a finite crystal [3]. The presence of finite-lattice transforms suggests the application of a key idea by Sayre, who suggested that diffraction patterns can be phased *ab initio* if the intensities are sampled *between* Bragg reflections, according to Shannon's sampling theorem [14]. Attempts to extract continuous intensity maps ('molecular transforms') without a highly coherent source were first made by Perutz and others near the time of Sayre's paper [15], though this approach, which requires physical modification of the crystal, has not seen significant use.

The problem of phasing diffraction data from coherently illuminated nanocrystals with varying size and shape has been considered recently, in light of the possibility to collect relevant diffraction data from XFEL sources [13,16–20]. These investigations, which aim to determine the contents of the crystal's unit cell, differ from related work aimed at mesoscopic observations such as strain fields [21–23]. They also differ from previous work that has focused on careful analysis of diffraction data from individual crystals with high signal-to-noise ratio (e.g. [24,25]).

The ensembles of crystals considered by Spence *et al.* were assumed to be constructed by repeated translations of a common unit cell electron density. Under that assumption, the averaged diffracted intensity $I(q)$ of many crystals is proportional to the product of a squared unit-cell transform $|F(q)|^2$, and a mean squared finite-lattice transform $\langle |S_n(q)|^2 \rangle$

$$I(q) \propto |F(q)|^2 \langle |S_n(q)|^2 \rangle_n, \tag{2.1}$$

where q is the momentum transfer vector and n denotes the pattern number. As $\langle |S_n(q)|^2 \rangle$ is a periodic function, it may be determined by averaging the diffraction intensity profiles within all Wigner–Seitz cells, which 'averages out' the unit-cell transform and effectively decouples the two terms in equation (2.1). Dividing equation (2.1) by the finite lattice transform reveals the transform of the common physical unit cell that repeats throughout the entire crystal, which may be phased via numerous iterative procedures employed in coherent diffractive imaging [26].

Challenges associated with noise, particularly in regions far from the Bragg condition where measured intensity is likely to be low, have been considered [19]. Possible solutions include noise filtering [13], selective sampling [16] or iterative algorithms that directly use only Bragg peak intensities and their associated intensity gradients [18]. Internal crystal disorder has also been considered, in which case the incorporation of partial coherence models into phasing algorithms can effectively improve resolution [17].

Whereas previous work has assumed a common molecular arrangement for every unit cell, a different situation may arise for space groups other than P1, in which a nominal unit cell contains multiple symmetry-related molecules. It is likely that many such crystals do not assemble in whole-unit-cell increments; some of the symmetry-related molecules that make up a nominal unit cell may be absent near the boundary of the crystal. The phasing methods considered previously do not apply directly to this situation because a unit cell that is common throughout the entire crystal does not exist, and hence equation (2.1) cannot be applied. Below, we show that in some cases, the data reduction scheme proposed by Spence *et al.* may still be used to recover the electron density of the asymmetric unit that composes the crystal, provided that the phasing algorithm is modified appropriately.

3. Electron densities of finite crystals

Consider a finite crystal in which the nominal unit cell contains one molecule and one symmetry-related copy. We define the electron density of the molecular asymmetric unit as $A(r)$, and its symmetry mate as $B(r)$. The symmetry mate is related to the asymmetric unit by a rotation $\mathbf{R}$ and a translation t

$$B(r) = A(\mathbf{R}r - t). \tag{3.1}$$

In order to define a complete finite crystal, we define two finite sub-lattices $\alpha(r)$ and $\beta(r)$ for the asymmetric unit and its symmetry mate, respectively. We may express the sub-lattices as

$$\alpha_n(r) = \sum_j o_{nj}^{\alpha} \delta(r - r_j) \tag{3.2}$$

and

$$\beta_n(r) = \sum_j o_{nj}^{\beta} \delta(r - r_j), \tag{3.3}$$

which differ only by the molecular occupancies $o_{kj}^{\alpha}, o_{kj}^{\beta} \in 0, 1$ for each lattice point r_j. We can build the electron density of the nth crystal by convolving the sub-lattices with their respective molecular electron densities

$$\rho_n(r) = A(r) \otimes \alpha_n(r) + B(r) \otimes \beta_n(r), \tag{3.4}$$

where $\otimes$ denotes a convolution.

In this framework, we note that one can choose a different *nominal* unit cell by translating $B(r)$ by integer multiples of the three crystal lattice vectors a, b and c. In other words, we are free to replace $B(r)$ with $B(r + d)$, where $d = n_a a + n_b b + n_c c$ and n_a are integers. This has no consequence at the Bragg condition since translations d produce integer multiples of 2π phase shifts at those points in reciprocal space. However, when considering intensities from finite crystals that do not lie at the Bragg condition, care must be taken to make corresponding changes to the occupancies o_{kj}^{β} whenever the nominal unit cell (i.e. the vector d) is re-defined. The *physical* unit cell of the crystal, if one exists, should be distinguished from this nominal unit cell. A physical unit cell can be defined only if one can choose a vector d such that $o_{nj}^{\alpha} = o_{nj}^{\beta}$ for all j. In other words, a physical unit cell can be assigned to a finite crystal *only* if the crystal can be assembled purely from a common unit cell that spans the entire volume of the crystal.

4. The average finite-crystal diffraction intensity

Under the Born approximation, the far-field diffraction amplitude is proportional to the Fourier transform ($\mathcal{F}$) of $\rho_n(r)$

$$\mathcal{F}\{\rho_n(r)\} = \tilde{\rho}_n(q) = \tilde{A}(q)\tilde{\alpha}_n(q) + \tilde{B}(q)\tilde{\beta}_n(q), \quad (4.1)$$

where

$$\tilde{B}(q) = \tilde{A}(\mathbf{R}^{\mathrm{T}}q)e^{-iq\cdot t}, \quad (4.2)$$

$$\tilde{\alpha}_n(q) = \sum_j o_{nj}^{\alpha} e^{iq\cdot r_j} \quad (4.3)$$

and

$$\tilde{\beta}_n(q) = \sum_j o_{nj}^{\beta} e^{iq\cdot r_j}. \quad (4.4)$$

The diffracted intensity is proportional to

$$I_n(q) = \tilde{\rho}_n(q)\tilde{\rho}_n^*(q) \quad (4.5)$$

$$= |\tilde{A}(q)|^2|\tilde{\alpha}_n(q)|^2 + |\tilde{B}(q)|^2|\tilde{\beta}_n(q)|^2$$

$$+ 2\mathrm{Re}\{\tilde{A}(q)\tilde{B}^*(q)\tilde{\alpha}_n(q)\tilde{\beta}_n^*(q)\}, \quad (4.6)$$

and the average diffracted intensity, arising from a distribution of many crystals with variable occupancies o_{nj}^{α} and o_{nj}^{β}, is

$$\langle I_n(q)\rangle_n = |\tilde{A}(q)|^2\langle|\tilde{\alpha}_n(q)|^2\rangle_n + |\tilde{B}(q)|^2\langle|\tilde{\beta}_n(q)|^2\rangle_n$$

$$+ 2\mathrm{Re}\{\tilde{A}(q)\tilde{B}^*(q)\langle\tilde{\alpha}_n(q)\tilde{\beta}_n^*(q)\rangle_n\}. \quad (4.7)$$

In the special case $o_{nj}^{\alpha} = o_{nj}^{\beta}$, in which the crystal has an unambiguous physical unit cell, equation (4.7) may be simplified to the form described in previous work

$$\langle I_n(q)\rangle_n = |\tilde{A}(q) + \tilde{B}(q)|^2\langle|\tilde{\alpha}_n(q)|^2\rangle_n. \quad (4.8)$$

A more realistic model is one in which the crystals are fully occupied up to some boundary, outside of which the occupancies o_{nj}^{α} and o_{nj}^{β} are equal to zero. The statistical properties of o_{nj}^{α} and o_{nj}^{β}, in particular the correlations between them, should be in correspondence with the nature of the crystal growth kinetics, which may vary considerably between different crystal types. In the following sections, we will only consider a simple case in which the free energy associated with the adhesion of a molecule to the crystal surface is identical for all surface molecules.

5. Finite one-dimensional crystals

For illustrative purposes, consider a one-dimensional crystal with two molecules per unit cell, and a unit cell length of a. We will consider an equal mixture of crystals having all possible edge terminations, including those that leave incomplete unit cells at the crystal boundaries. Schematically, the four possible crystal types are shown in figure 1.

We can see that types (i) and (ii) are ideal crystals that differ only by their physical unit cell; we may produce type (ii) from type (i) by re-defining the symmetry mate as

$$B'(r) = B(r + a). \quad (5.1)$$

For brevity, we drop the explicit q dependence and write the average diffraction intensity from these two types as

$$\langle I_n\rangle_n^{\mathrm{i,ii}} = \frac{1}{2}[\langle I_n\rangle_n^{\mathrm{i}} + \langle I_n\rangle_n^{\mathrm{ii}}], \quad (5.2)$$

where $\langle\cdot\rangle^{\mathrm{i}}$ represents the average over crystals with unit cell (i), and $\langle\cdot\rangle^{\mathrm{i,ii}}$ represents the average over crystals with both unit cell definitions (i) and (ii). Thus, if the method prescribed by Spence et al. is applied to the averaged intensity arising
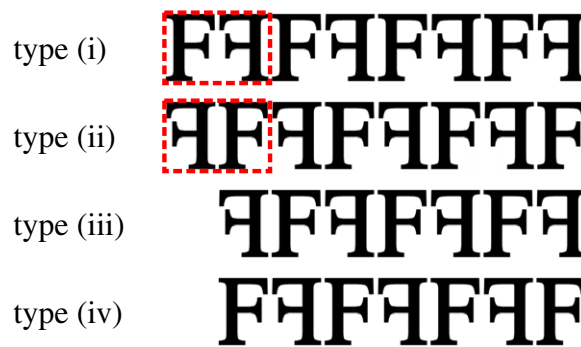
Figure 1. Illustration of four unique types of truncated one-dimensional crystals with two molecules per unit cell. The two unit cell conventions are indicated by the dashed red boxes. Crystal type (ii) is generated from type (i) by switching the first molecule with the second. Types (iii) and (iv) are generated from types (i) and (ii), respectively, by removing the first molecule. (Online version in colour.)

from these two types of ideal crystals, the average over the two *different* unit-cell transforms will be obtained.

Types (iii) and (iv) do not have a well-defined unit cell. They can be understood as follows. Firstly, we see that type (iii) may be formed by removing the first molecule from type (i). Mathematically, this operation may be performed by making the change[1] $\tilde{\alpha}_n \to \tilde{\alpha}_n - 1$ in equation (4.7), which produces the average diffraction intensity

$$\langle I_n\rangle_n^{\mathrm{iii}} = \langle I_n\rangle_n^{\mathrm{i}} - p \quad (5.3)$$

$$\text{and}\quad p = 2\mathrm{Re}\{|\tilde{A}|^2\langle\tilde{\alpha}_n\rangle_n + \tilde{A}\tilde{B}^*\langle\tilde{\alpha}_n^*\rangle_n\} - |\tilde{A}|^2. \quad (5.4)$$

Similarly, the average crystal type (iv) can be produced from type (ii) by making the change $\tilde{\beta}_n \to \tilde{\beta}_n - 1$

$$\langle I_n\rangle_n^{\mathrm{iv}} = \langle I_n\rangle_n^{\mathrm{ii}} - p' \quad (5.5)$$

$$\text{and}\quad p' = 2\mathrm{Re}\{|\tilde{B}|^2\langle\tilde{\alpha}_n\rangle_n + \tilde{A}\tilde{B}'^*\langle\tilde{\alpha}_n\rangle_n\} - |\tilde{B}|^2. \quad (5.6)$$

Finally, the combined average over all four crystal types is

$$\langle I_n\rangle_n = \langle I_n\rangle_n^{\mathrm{i,ii}} - \frac{1}{2}[p + p']. \quad (5.7)$$

We can see that the average crystal diffraction contains a prominent term equal to the average over two possible unit cell definitions. The additional terms p and p' arise from the interference of idealized crystals with a single 'missing' molecule at the boundary of the crystal. In the following sections, we consider the feasibility of phasing the averaged diffraction intensities from an ensemble of randomly terminated crystals based on the conjecture that these interference terms may be neglected.

6. Finite two-dimensional crystals

As the enumeration of two-dimensional crystal types is considerably more complicated than for one-dimensional crystals, we will proceed with simulations. For simplicity, we will assume a square lattice defined by the vectors $a = [a, 0]$ and $b = [0, a]$. We consider the plane group cm [27] with the symmetry mate defined by the rotation (a reflection in two-dimensional projection) and translation

$$\mathbf{R} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (6.1)$$

and

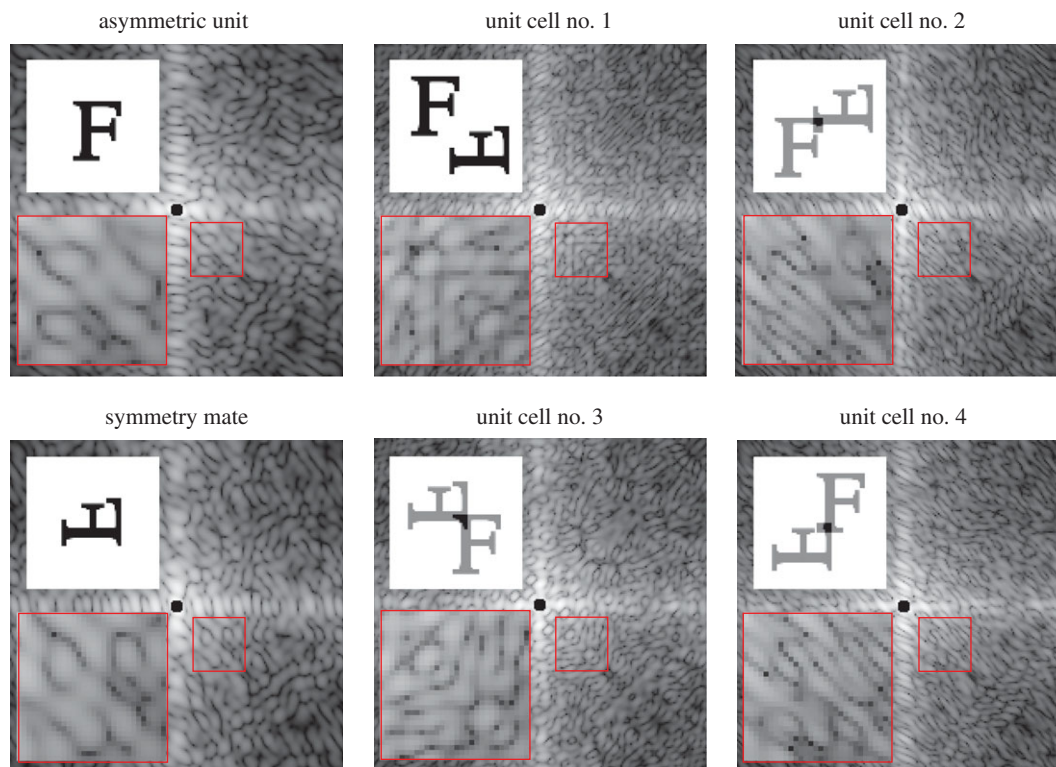$$t = [0.5, 0.5] \cdot a. \quad (6.2)$$

Figure 2. The real-space density (upper insets) and diffraction from the asymmetric unit, symmetry mate, and four compact nominal unit cell definitions. Lower insets show enlarged regions indicated by red boxes. (Online version in colour.)

The real-space densities and diffraction intensities from the asymmetric unit, symmetry mate and four different nominal unit cells are shown in figure 2.

We generated 500 diffraction patterns from random circle-shaped crystals with a flat distribution of radii equal to $4a \pm 0.5a$. We ensured randomized edge terminations by randomly shifting the origin of each crystal lattice and retained only the lattice points for molecules that fell within the circular boundary with a fixed origin. The intensities were sampled such that three measurements lie between adjacent Bragg reflections, and the maximum Miller index extended out to $h, k = 25$. We added random, uniform jitter to the scattering vectors corresponding to each two-dimensional pixel. The addition of jitter emulates a physical detector, which averages over a finite solid angle, and avoids aliasing that would otherwise result from the fine lattice fringes and comparably coarse pixel spacings. We then divided the average crystal intensity by its average Wigner–Seitz cell to obtain what we refer to as the demodulated crystal intensities. The results of this procedure are shown in figure 3. Remarkably, the demodulated crystal intensities closely correspond to those of the incoherent sum of the four compact nominal unit cell transforms shown in figure 2. The resemblance may be quantified by an R-factor defined as

$$R = \frac{\sum_m |\eta I(q_m) - I_{\text{ideal}}(q_m)|}{\sum_m |I_{\text{ideal}}(q_m)|}, \qquad (6.3)$$

where η is a scale factor chosen to minimize R. The intensities $I(q_m)$ are the demodulated crystal intensities, $I_{\text{ideal}}(q_m)$ are equal to the incoherent sum over the four nominal unit cells shown in figure 2. In our case, we find convergence to a value of $R = 0.07$ after 500 patterns. This convergence was largely insensitive to the size and shape of the crystals, but was somewhat slower when the width of the size distribution was increased.

7. Phasing intensities from randomly truncated crystals

An approach to phasing symmetry-averaged diffraction data has been described previously by Elser & Millane [28]. Here, we apply a similar approach with a modified intensity constraint appropriate to our problem. Specifically, in our case we assume that the demodulated crystal intensities $I(q)$ are related to the molecular transform of the asymmetric unit $\tilde{\rho}(q)$ by the approximation

$$I(q) = \frac{1}{4} \sum_{n=1}^{4} |\tilde{\rho}(q) + \tilde{\rho}(\mathbf{R}^{\mathsf{T}}q) e^{-iq \cdot t_n}|^2, \qquad (7.1)$$

where

$$t_1 = [0.5, 0.5] \cdot a, \qquad (7.2)$$

$$t_2 = t_1 - a, \qquad (7.3)$$

$$t_3 = t_1 - b \qquad (7.4)$$

$$\text{and} \quad t_4 = t_1 - a - b. \qquad (7.5)$$

We define two projection operators. The intensity projection $\widetilde{P}_{\text{I}}$ has the action of bringing the magnitudes of the current estimate of the molecular transform, $\tilde{\rho}_i(q)$, into correspondence with the measured intensities

$$\widetilde{P}_{\text{I}} \tilde{\rho}_i(q) = \tilde{\rho}_i(q) \sqrt{\frac{I(q)}{\frac{1}{4} \sum_{n=1}^{4} |\tilde{\rho}_i(q) + \tilde{\rho}_i(\mathbf{R}^{\mathsf{T}}q) e^{-iq \cdot t_n}|^2}}. \qquad (7.6)$$

The support projection $\widetilde{P}_{\text{S}}$ sets the real-space densities to zero in the regions outside of the support S

$$\widetilde{P}_{\text{S}} \tilde{\rho}_i(q) = \mathcal{F} P_{\text{S}} \mathcal{F}^{-1} \tilde{\rho}_i(q), \qquad (7.7)$$
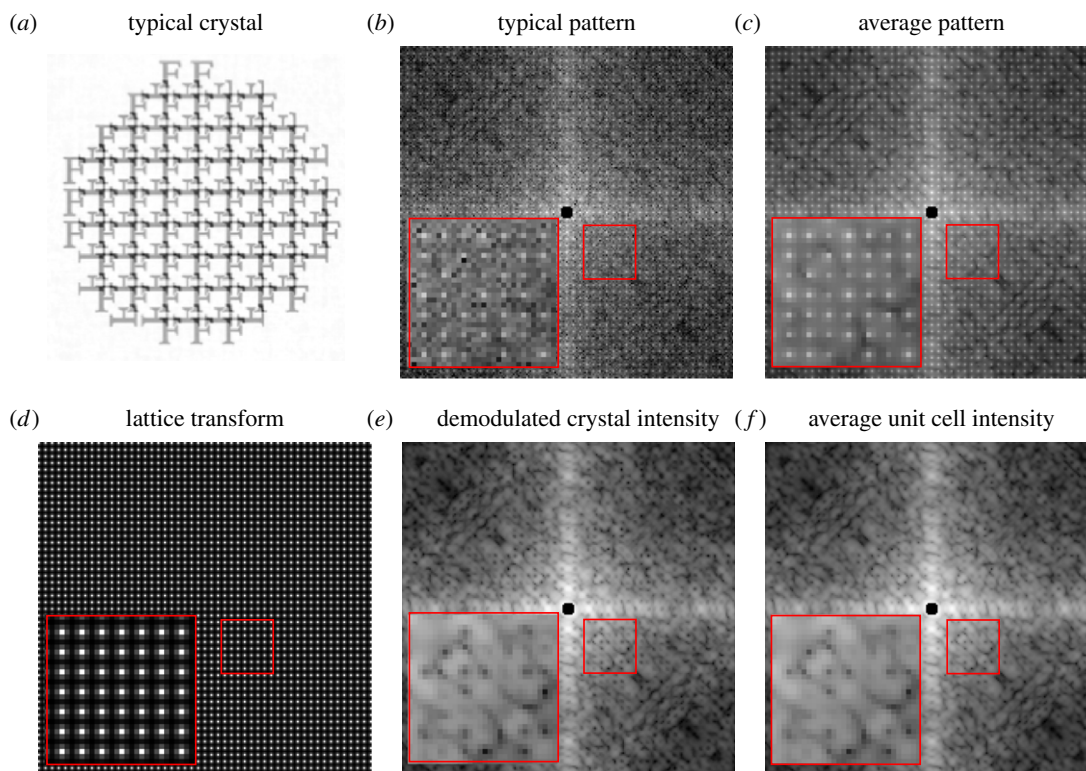
Figure 3. (a) A typical crystal, (b) typical diffraction pattern and (c) the average over 500 diffraction patterns. (d) The average Wigner–Setz cell is divided to obtain (e) the demodulated crystal intensity map which may be compared with (f) the incoherent average over the four compact nominal unit cell definitions shown in figure 2. Lower insets show enlarged regions indicated by red boxes. (Online version in colour.)

where $\mathcal{F}^{-1}$ is the inverse Fourier transform and

$$P_S \rho_i(q) = \begin{cases} \rho_i(r) & \text{if } r \in S \\ 0 & \text{otherwise.} \end{cases} \qquad (7.8)$$

With the two projection operations $\tilde{P}_I$ and $\tilde{P}_S$, we then apply the difference-map (DM) algorithm (assuming $\beta = 1$) [29]

$$\tilde{\rho}_{i+1}(q) = \tilde{\rho}_i(q) + \tilde{P}_S (2 \tilde{P}_I \tilde{\rho}_i(q) - \tilde{\rho}_i(q)) - \tilde{P}_I \tilde{\rho}_i(q), \qquad (7.9)$$

in combination with the error-reduction (ER) algorithm [30]

$$\tilde{\rho}_{i+1}(q) = \tilde{P}_S \tilde{P}_I \tilde{\rho}_i(q). \qquad (7.10)$$

The algorithm switched between the ER and DM algorithms every 10 iterations, beginning with the DM algorithm. The support was updated via the shrinkwrap algorithm [31] every 100 iterations. A successful result, after 2000 iterations in total, is shown in figure 4. The accuracy of this result was quantified by an R-factor that compares the demodulated crystal intensities against the average nominal unit cell transform generated from the electron density estimate $\rho_i(r)$. Specifically, this R-factor was defined as

$$R_{ph} = \frac{\sum_m |\eta I_{comp}(q_m) - I(q_m)|}{\sum_m |I(q_m)|}, \qquad (7.11)$$

where

$$I_{comp}(q_m) = \sum_{n=1}^{4} |\tilde{\rho}_i(q_m) + \tilde{\rho}_i(\mathbf{R}^T q_m) e^{-iq_m \cdot t_n}|^2. \qquad (7.12)$$

This R-factor reduces to a value of $R_{ph} = 0.07$, nearly the same value as R mentioned in §6. The near equivalence of R and R_{ph} suggests that the quality of the reconstruction is limited by the approximate nature of our intensity constraint.

A more rigorous algorithm than the one demonstrated here would ideally use equation (4.7) directly in the intensity constraint, though this would require the introduction of a more sophisticated model for the average sub-lattice cross terms $\langle \tilde{\alpha}_n(q) \tilde{\beta}_n^*(q) \rangle_n$, the parameters of which must be refined in parallel with the phases if the they cannot be determined by other means. Alternatively, it may be possible to better approximate a more general crystal by introducing independent weights on each term of the sum in equation (7.1). These weights must be solved in parallel with the phases as they must be assumed unknown at the first iteration. In our demonstration presented here, we have used the assumption that molecules at the edge of the crystal are randomly truncated, which ensures equal weights in our approximation.

Importantly, we note that there are circumstances where just the integrated Bragg peak intensities are sufficient for unique phasing, particularly for cases in which the solvent fraction (fraction of densities that are uniformly constant) exceeds 50%. Indeed, the simulated crystals shown in figure 3 had a solvent fraction of 62% and can likely be phased without intensities that lie away from the Bragg condition (we include these figures here only for clarity). However, the same algorithm applied to cases where the solvent fractions were 39 and 16% also converged to values of $R_{ph} = 0.07$ or lower, though the greatest fraction of phasing trials succeeded in the cases of higher solvent fractions (about 2/3 at 62%, 1/3 at 39% and 1/10 at 16%). We made no attempt to optimize the algorithm in the cases of higher solvent fractions.

We have not considered the robustness of our algorithm in the presence of measurement errors. Statistical errors in our simulations are induced only by the randomized sampling of crystal sizes and shapes, as well as pixel position jitter, and are smaller than the systematic errors induced by
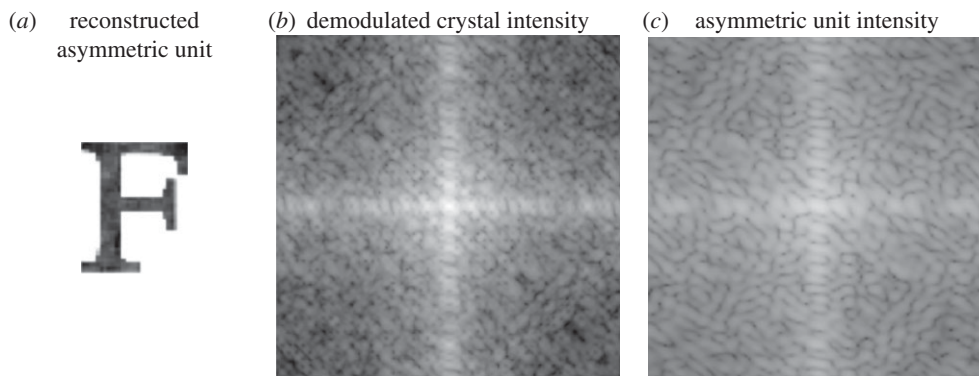
(a) reconstructed asymmetric unit	(b) demodulated crystal intensity	(c) asymmetric unit intensity

Figure 4. (*a*) A successful reconstruction of the asymmetric unit after 2000 iterations. (*c*) The constrained asymmetric unit intensities were projected onto (*b*) the demodulated crystal intensity constraint according to equation (7.6).

our approximate diffraction model. Assessing the impact of errors should be carried out by implementing the modified intensity constraint within a noise-tolerant phasing algorithm (e.g. [32]), which is beyond the scope of this paper.

8. Conclusion

We have raised the point that finite crystals may not have a well-defined physical unit cell. While this realization has no consequence in 'conventional' serial femtosecond nanocrystallography, it is of great importance to the development of phasing algorithms that use intensities that are not restricted to the Bragg condition. With some approximation, we have demonstrated that it is possible to determine the structure of an asymmetric unit via iterative phasing techniques for crystals with random edge terminations, and only two molecules per nominal unit cell.

The consequences of our approximations, which neglect interference terms and assume relatively uncorrelated crystal edge terminations, is in need of further consideration. Moreover, the extension to space groups with more than two symmetry-related molecules may give rise to further complications.

Endnote

[1]This expression is correct when the first term of the summation in equation (4.3) is equal to 1. This may always be accomplished by choosing an appropriate origin of the entire crystal, which never affects the diffraction intensity.

References

1. Spence JCH, Weierstall U, Chapman HN. 2012 X-ray lasers for structural and dynamic biology. *Rep. Prog. Phys.* **75**, 102601. (doi:10.1088/0034-4885/75/10/102601)

2. Schlichting I, Miao J. 2012 Emerging opportunities in structural biology with X-ray free-electron lasers. *Curr. Opin. Struct. Biol.* **22**, 613–626. (doi:10.1016/j.sbi.2012.07.015)

3. Chapman HN *et al*. 2011 Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–77. (doi:10.1038/nature09750)

4. Boutet S *et al*. 2012 High-resolution protein structure determination by serial femtosecond crystallography. *Science* **337**, 362–364. (doi:10.1126/science.1217737)

5. Neutze R, Wouts R, van der Spoel D, Weckert E, Hajdu J. 2000 Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757. (doi:10.1038/35021099)

6. Barty A *et al*. 2012 Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photonics* **6** 35–40. (doi:10.1038/nphoton.2011.297)

7. Aquila A *et al*. 2012 Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt. Expr.* **20** 2706–2716. (doi:10.1364/OE.20.002706)

8. Kern J *et al*. 2013 Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* **340**, 491–495. (doi:10.1126/science.1234273)

9. Redecke L *et al*. 2012 Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227–230. (doi:10.1126/science.1229663)

10. White TA, Kirian RA, Martin AV, Aquila A, Nass K, Barty A, Chapman HN. 2012 CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **45**, 335–341. (doi:10.1107/S0021889812002312)

11. Son S-K, Chapman HN, Santra R. 2011 Multiwavelength anomalous diffraction at high X-Ray intensity. *Phys. Rev. Lett.* **107**, 218102. (doi:10.1103/PhysRevLett.107.218102)

12. Millane RP, Lo VL. 2013 Iterative projection algorithms in protein crystallography. I. Theory. *Acta Crystallogr.* **A69**, 517–527. (doi:10.1107/S0108767313015249)

13. Spence JCH *et al*. 2011 Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals. *Opt. Expr.* **19**, 2866. (doi:10.1364/OE.19.002866)

14. Sayre D. 1952 Some implications of a theorem due to Shannon. *Acta Crystallogr. A* **5**, 843. (doi:10.1107/S0365110X52002276)

15. Perutz MF. 1954 The structure of haemoglobin. 3. Direct determination of the molecular transform. *Proc. R. Soc. Lond. A* **225**, 264–286. (doi:10.1098/rspa.1954.0202)

16. Chen JP, Spence JC, Millane RP. 2012 Phase retrieval in femtosecond X-ray nanocrystallography. Proceedings of the 27th Conference on Image and Vision Computing New Zealand 2012, pp. 43–48. New York, NY ACM.

17. Dilanian RA, Streltsov VA, Quiney HM, Nugent KA. 2012 Continuous X-ray diffractive field in protein nanocrystallography. *Acta Crystallogr.* **A69**, 108–118. (doi:10.1107/S0108767312042535)

18. Elser V. 2013 Direct phasing of nanocrystal diffraction. *Acta Crystallogr.* **A69**, 559–569. (doi:10.1107/S0108767313023362)

19. Chen JPJ, Spence JCH, Millane RP. 2014 Direct phasing in femtosecond nanocrystallography. I. Diffraction characteristics. *Acta Crystallogr.* **A70**, 1–11. (doi:10.1107/S2053273313032038)

20. Chen JPJ, Spence JCH, Millane RP. 2014 Direct phasing in femtosecond nanocrystallography. II. Phase retrieval. *Acta Crystallogr.* **A70**, 1–8. (doi:10.1107/S2053273313032725)

21. Williams G, Pfeifer M, Vartanyants I, Robinson I. 2003 Three-dimensional imaging of microstructure in Au nanocrystals. *Phys. Rev. Lett.* **90**, 175501. (doi:10.1103/PhysRevLett.90.175501)

22. Pfeifer MA, Williams GJ, Vartanyants IA, Harder R, Robinson IK. 2006 Three dimensional mapping of a deformation field inside a nanocrystal. *Nature* **442**, 63–66. (doi:10.1038/nature04867)

23. Boutet S, Robinson IK. 2008 Coherent X-ray diffractive imaging of protein crystals. *J. Synchrotron Radiat.* **15**, 576–583. (doi:10.1107/S0909049508029439)

24. Xu G, Zhou GE, Zhang XY. 1999 Phase recovery for X-ray crystallography. *Phys. Rev. B* **59**, 9044–9047. (doi:10.1103/PhysRevB.59.9044)

25. Miao J, Sayre D. 2000 On possible extensions of X-ray crystallography through diffraction pattern oversampling. *Acta Crystallogr. A* **56**, 596–605. (doi:10.1107/S010876730001031X)

26. Marchesini S. 2007 A unified evaluation of iterative projection algorithms for phase retrieval. *R. Sci. Instrum.* **78**, 011301. (doi:10.1063/1.2403783)

27. Hahn T. 2005 *International tables for crystallography, space-group symmetry*. New York, NY: Wiley.

28. Elser V, Millane RP. 2008 Reconstruction of an object from its symmetry-averaged diffraction pattern. *Acta Crystallogr. A* **64**, 273–279. (doi:10.1107/S0907444907066164)

29. Elser V. 2003 Phase retrieval by iterated projections. *J. Opt. Soc. Am. A* **20**, 40–55. (doi:10.1364/JOSAA.20.000040)

30. Gerchberg RW, Saxton WO. 1972 Practical algorithm for determination of phase from image and diffraction plane pictures. *Optik* **35**, 237.

31. Marchesini S, He H, Chapman HN, Hau-Riege SP, Noy A, Howells MR, Weierstall U, Spence JCH. 2003 X-ray image reconstruction from a diffraction pattern alone. *Phys. Rev. B* **68**, 140101. (doi:10.1103/PhysRevB.68.140101)

32. Martin AV *et al*. 2012 Noise-robust coherent diffractive imaging with a single diffraction pattern. *Opt. Expr.* **20**, 16 650–16 661. (doi:10.1364/OE.20.016650)

7

rstb.royalsocietypublishing.org   *Phil. Trans. R. Soc. B* **369**: 20130331