Challenges in solving structures from radiation-damaged tomograms of protein nanocrystals assessed by simulation

Ariana Peck, Qing Yao, Aaron S. Brewster, Petrus H. Zwart, John M. Heumann, Nicholas K. Sauter and Grant J. Jensen

1 Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
2 Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
3 Center for Advanced Mathematics in Energy Research Applications, Lawrence Berkeley National Laboratory, Berkeley CA, 94720, USA
4 Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO 80309, USA

(Dated: September 18, 2020)

Structure determination methods are needed to resolve the atomic details that underlie protein function. X-ray crystallography has provided most of our knowledge of protein structure but is constrained by the need for large, well-ordered crystals and the loss of phase information. The rapidly developing methods of serial femtosecond crystallography, micro-electron diffraction, and single-particle reconstruction circumvent the first of these limitations by enabling data collection from nanocrystals or purified proteins. However, the first two methods also suffer from the phase problem, while many proteins fall below the molecular weight threshold required by single-particle reconstruction. Cryo-electron tomography of protein nanocrystals has the potential to overcome these obstacles of mainstream structure determination methods. Here we present a data processing scheme that combines routines from X-ray crystallography and new algorithms we developed to solve structures from tomograms of nanocrystals. This pipeline handles image processing challenges specific to tomographic sampling of periodic specimens and is validated using simulated crystals. We also assess the tolerance of this workflow to the effects of radiation damage. Our simulations indicate a trade-off between a wider tilt-range to facilitate merging data from multiple tomograms and a smaller tilt increment to improve phase accuracy. Since phase errors but not merging errors can be overcome with additional datasets, these results recommend distributing the dose over a wide angular range rather than using a finer sampling interval to solve the protein structure.

I. Introduction

Protein structure determination critically advances our understanding of biochemical mechanisms and the molecular basis of disease. X-ray crystallography has been the principal method of structure determination for decades, accounting for 90% of atomic models deposited in the Protein Data Bank [1]. However, two limitations constrain this method’s broader applicability. First, some proteins do not readily form sufficiently large and well-ordered crystals for characterization at synchrotron sources [2, 3]. Second, the phase information needed to solve a protein’s structure cannot be experimentally measured. Estimating this lost information requires experimental perturbations that some crystals are not amenable to, very high resolution diffraction data, or the availability of a homologous structure [4]. Such prior information is often limited for proteins that are difficult to crystallize, compounding the challenge of structurally characterizing new protein families.

The need for large crystals has been overcome in part by the development of techniques suitable for sub-micron sized nanocrystals. One of these methods is serial femtosecond crystallography (SFX), which relies on femtosecond-length X-ray pulses that are orders of magnitude brighter than synchrotron radiation [5]. SFX has significantly advanced structural studies of difficult-to-crystallize membrane proteins [6, 7] and crystals sensitive to radiation damage [8, 9] but suffers from low throughput. A second method is micro-electron diffraction (microED). This modality of cryo-electron microscopy (cryo-EM) takes advantage of the strong interaction between electrons and matter to enable data collection from crystals less than 1 µm thick [10]. Similar to SFX, microED has enabled structural studies of amyloid proteins that do not readily form large, well-ordered crystals [11, 12]. However, neither method can measure the crystallographic phases, so this information must be inferred indirectly or supplied by other techniques [4].

Single particle reconstruction (SPR) is another cryo-EM modality that is increasingly effective for high-resolution structure determination. In SPR, projection images of purified proteins are recorded, aligned, and merged to determine the protein’s structure [13]. This method thus overcomes the main limitations of X-ray crystallography, bypassing the need for crystallization and retaining the phase information. Further, recent technological advances have enabled SPR to achieve reconstructions with high-resolution limits comparable to X-ray crystallography [13]. However, without the signal amplification that results from coherent scattering by a crystal, molecular weight is a limiting factor. As
the object’s scattering mass decreases, errors in aligning projection images increase and attenuate the high-resolution signal [14, 15]. To date, the smallest macromolecule solved by SPR to better than 4 Å resolution is a 40 kDa riboswitch, and only seven proteins <100 kDa have been solved to similar resolution [16]. By contrast, the median mass of proteins in the human proteome is 41 kDa [17], rendering SPR unsuitable for many proteins.

Cryo-electron tomography (cryo-ET) is a third modality of cryo-EM, and its application to nanocrystals could overcome the principal obstacles of mainstream structure determination methods [18]. In cryo-ET, a tilt-series of projection images is collected from a sample embedded in vitreous ice and reconstructed into a tomogram, a volume that contains 3D structural information about the specimen. This method has traditionally been used to study cellular ultrastructure, and the reconstruction approach of subtomogram averaging has recently determined near-atomic resolution (<5 Å) structures of select purified proteins and ribosomes in situ [19–21]. Like microED, cryo-ET is suitable for protein nanocrystals, with an expected optimal sample thickness of 50-300 nm [22, 23]. Like SPR, imaging rather than diffraction data are collected, so the phase information is retained during the experiment [24]. Further, collecting data in imaging mode provides a unique opportunity to spatially characterize and computationally correct for disorder [25, 26]. In 2D electron crystallography, this was achieved by a procedure called lattice “ unbending” [26, 27]. Alternatively, disorder could be corrected in real space with subtomogram averaging algorithms [20, 28–32]. Another benefit of applying cryo-ET to nanocrystals is the potential to develop a hybrid method that combines high-resolution diffraction intensities from microED and low-resolution phases from imaging for structure solution. Even phases to intermediate resolution (~7 Å) should suffice to resolve α-helices and provide a robust starting model for phase extension [33, 34]. Development of this hybrid method should be straightforward, as sample preparation and the microscope are shared between the two techniques.

While electron imaging of 3D crystals has revealed lattice structure at nanoscale lengths [25, 35, 36], to date structure determination has not been achieved from tomograms of such samples. In our first attempts to solve structures from experimental tomograms of nanocrystals, we found that low completeness was particularly severe at high tilt-angles and prohibitive to merging datasets. These observations prompted us to analyze the challenges of merging incomplete datasets and explore ways to improve data acquisition from nanocrystals. Additional difficulties, such as accounting for multiple scattering and the contrast transfer function, are anticipated during data processing [26, 37]. However, we focused on factors that predicted the ability to merge multiple datasets, as merging is critical to overcome the missing wedge of information inherent to tomography and loss of completeness due to radiation damage. Here we describe a data processing scheme to solve structures from tomograms of nanocrystals that leverages software from X-ray crystallography and algorithms we developed to handle challenges unique to tomography data. We also assess this workflow’s tolerance to the effects of radiation damage using simulated crystals. Our results recommend a data collection strategy that maximizes the angular spread of the reflections recorded in each tomogram to increase the likelihood of successfully merging datasets.

II. Design of a data processing pipeline

Structure determination from tomograms of nanocrystals could leverage algorithms from crystallography or, alternatively, rely on subtomogram averaging [38] (see Discussion). Here we focus on the former approach. Since retention of phase information is unique to imaging methods, we emphasize the steps required to recover crystallographic phases from the tomogram’s Fourier transform. Our data processing pipeline leverages functions available in the SPARX [39, 40], DIALS [41, 42], and cctbx [43] software packages in addition to providing new algorithms developed specifically for tomographic data of crystal specimens. The steps are presented schematically in Fig. S1A and described in the following sections.

To develop this workflow and explore different ways of processing these data, we simulated tomograms of protein nanocrystals. We selected lysozyme in P1 symmetry (PDB ID: 6D6G) and a peptide inhibitor in P21 symmetry (PDB ID: 4F1) as model crystal systems to test distinct space group symmetries and protein folds. For each crystal system, the protein coordinates from the Protein Data Bank were tiled in Chimera to generate 3D nanocrystals containing 10 unit cells along each dimension [44]. Density maps were simulated from the nanocrystal’s atomic coordinates to 2.5 Å resolution using electron scattering factors with the cctbx software package and rotated to randomly-sampled orientations [43]. These intact volumes were projected into tilt-series spanning either a ±60° tilt-range with 3° increments or a ±40° tilt-range with 2° increments between projection images. No optical aberrations from the microscope, including the contrast transfer function, were simulated. The resulting tilt-series were reconstructed into tomograms using IMOD [45]. The workflow used to generate these simulated datasets is shown in Fig. S1B.

A. Eliminating phase splitting

Real crystals are characterized by imperfections that result in loss of exact periodicity. Imaging introduces further non-idealities, such as interpolation errors from discrete sampling and truncation of crystal edges due to a finite field of view. These deviations from perfect crystallinity in real space result in spatially-smeared intensities and phase splitting at Bragg peak positions in reciprocal space (Fig. S2). Specifically, the phase values
A. 1KL plane h,k,l=(1,1,1)  

B.  

FIG. 1. Elimination of phase splitting at Bragg peaks.  
(A) Left: The 1KL plane is visualized for a simulated crystal, which is imperfectly periodic because the unit cell dimensions do not span an integer number of pixels. The brightness and color of each pixel are determined by its intensity and phase values, respectively. Right: Inset of the (1,1,1) reflection. Pixels with intensity values within three-fold of the maximum intensity are visualized, and the phase values of these high-intensity pixels are noted in degrees. In (B), a tapering function was applied to the crystal. The density was centered by auto-convolution and shifted to the volume’s origin before computing the Fourier transform.

of pixels immediately surrounding peak centers shift by 180° between adjacent octants (Figs. 1A, S1). The split phases result from the presence of a circular discontinuity at the point considered the origin by the discrete Fourier transform, causing the phase to alternate sign between frequency bins in Fourier space. There is no such discontinuity for signals that are exactly periodic in the window of the Fourier transform, and hence no phase splitting for ideal (albeit finite) crystals.

For imperfect crystals, we found that phase splitting could be eliminated by applying a symmetric tapering function in real space, followed by centering the density within the volume. Shifting the center of the resulting volume to the origin removes the circular discontinuity at the origin of the Fourier transform calculation. Here we chose a tapered cosine (Tukey) window with a tapering fraction of 0.5 and computed the translation needed to center the crystal density using the SPARX library’s auto-convolution function [39, 40]. Applying these pre-processing steps to crystals with imperfect periodicity eliminates phase splitting and results in a consistent phase value in the immediate vicinity of each Bragg peak (Figs. 1B, S2).

B. Spot-finding and indexing

We adapted algorithms from the crystallography software package DIALS to index the Bragg peaks in the tomogram’s 3D Fourier transform [41]. Peaks were identified by scanning the Fourier transform along the x-axis (parallel to the axis of rotation), and high-intensity pixels in each slice were identified using the pixel thresholding algorithm described in Ref. [42]. Sets of contiguous bright pixels in adjacent slices were assembled into an initial spots list, which was then filtered according to the specified gain (the ratio of electrons per detector pixel to reported counts), resolution limits, and minimum and maximum number of included pixels [42].

Spot centroids were then mapped to reciprocal space. We used a 1d Fast Fourier transform algorithm to index the spots and estimate the crystal’s orientation and unit cell constants to within a magnification scale factor [42, 46]. The provisional P1 cell constants were refined by enforcing constraints of the known space group symmetry. For experimental data, it is unclear whether the intensity information from imaging will be sufficiently accurate to determine the Laue class. Further, systematically forbidden reflections may be present due to multiple scattering or lie in an unsampled region of reciprocal space, preventing identification of screw axes [47]. However, symmetry could readily be determined by microED instead (see Discussion). Space group determination is critical for locating the crystallographic phase origin, as the phases do not reflect the space group symmetry until a crystallographic phase origin is found [47].

C. Bragg peak fitting

Peak fitting involves assigning pixels to Bragg peaks, followed by integrating the intensities and averaging the phases of the assigned pixels. In X-ray crystallography and microED, profile fitting algorithms are used to integrate diffraction peaks [42, 48–50]. Profile fitting assumes a standard spot shape and models how each reflection is sampled based on its orientation with respect to the rotation axis, crystal mosaicity, and beam divergence [49]. In the case of tomographic data, however, a generic reflection profile cannot be assumed. Each reflection is only partially measured due to the large spacing between tilt increments and lack of continuous rotation (Fig. S3A), and at present we do not have adequate models to account for these effects. Developing a model to compensate for reflection partiality would improve the estimated intensities, but it would be challenging to devise a corresponding correction for the estimated phases.

Given these challenges, we implemented the following heuristic approach to determine reflection profiles from tomographic data. First, the pixel coordinates of each lattice point in the tomogram’s Fourier transform were estimated from the crystal’s indexing matrix. The reflection was discarded if it was predicted to lie in the missing
FIG. 2. **Peak fitting for a representative reflection.** (A) The intensity profiles of slices are visualized along the indicated direction and centered on the (-2,3,-8) reflection of a simulated crystal tomogram. In (B), the intensities are shown in 2D, with high-intensity pixels assigned to the peak marked by an X. In (C), the phases in the vicinity of the peak are visualized using the same color scheme as in Fig. 1. The phase values of pixels retained as part of the peak are noted in degrees. The high-intensity pixel marked by a yellow X in (B) had a phase value of 17° and was discarded as an outlier.

The likelihood of including high-intensity outliers in the background region. We then estimated the values of the masked pixels by trilinear interpolation to approximate the background intensity beneath the peak. We used interpolation rather than assuming a constant background to better account for the anisotropic structure of the non-Bragg intensity, which results from the measured data being oriented along skew slices of finite thickness that are commonly separated by an angular increment of 2 or 3°. As a result, the orientation and magnitude of the background vary among peaks, depending on the peak’s position relative to the planes of sampled signal in Fourier space (Fig. S3). The estimated background values were then subtracted from the original intensities of the corresponding pixels.

The assignment of pixels to Bragg peaks was then refined using the phase values (Fig. 2C). Specifically, the intensity-weighted standard deviation of the assigned pixels’ phases was computed. If this metric exceeded a specified threshold, outlier pixels were iteratively removed until this threshold was reached. Empirically we found that a threshold of 15° enabled the recovery of accurate phase values. Ensuring consistent phase values within the peak is critical, as the phase can change sharply outside the Bragg peak (Fig. 2C). Finally, the reflection’s intensity and phase were respectively estimated from the sum of the intensity values and the intensity-weighted mean of the phase values of the retained pixels.

Both the intensity and phase components of the Bragg peaks were fit using tomograms pre-processed as described in Section IIA, which eliminated phase splitting. Although application of a tapered cosine function in real space reduced the intensities, it also prevented sharp edges in the window of the Fourier transform from being convolved with the shape of the Bragg peaks in the frequency domain. As a result, the reflection profiles were more spherical and a higher number of Bragg peaks were retained by the peak-fitting algorithm described above, which benefited from this rounder shape.

### D. Merging datasets

In cryo-ET, the experimental geometry limits the accessible tilt range to ±70° relative to the sample’s untitled orientation [38, 51]. In principle, this rotation range should suffice to collect a complete dataset from a single crystal for most point groups [52]. In practice, however, 140° is an overestimate of the useful rotation range because intermediate to high-resolution information is lost both at high tilt angles due to increased specimen thickness and in images recorded late in the tilt-series due to accumulated dose [53]. Further, a popular data collection strategy in cryo-ET uses a 3° tilt increment for a tilt-range of ±60° [53], so a fraction of Bragg peaks in the nominal tilt-range will fall between recorded images and not be sampled. The size of this fraction is anticipated to vary between samples and depend on fac-
tors like mosaicity and the resolution range. Achieving
high completeness thus requires merging data from mul-
tiple crystals in different orientations.

Merging phases from different crystals requires posi-
tioning the datasets on a common phase origin. Un-
less this is also a crystallographic origin, the phases of
symmetry-equivalent reflections are not related and the
phase data effectively have P1 symmetry. Here we es-
Established a common phase origin by treating one crystal
as reference and shifting the phase origins of the remain-
ing datasets to this reference origin (Fig. 3). The frac-
tional unit cell row vector, \( \mathbf{u} \), that shifts the phases of
a second crystal to the reference origin was determined
by minimizing the intensity-weighted mean residual be-
tween the reference phases and the shifted phases of the
second crystal:

\[
\arg \min_{\mathbf{u}} \sum_{\mathbf{h}} \bar{I}_{\mathbf{c}} |\langle \phi_{\mathbf{c}_2} (\mathbf{h}) - 2\pi \mathbf{r} \mathbf{h} \rangle - \langle \phi_{\mathbf{c}_1} (\mathbf{h}) \rangle| \tag{1}
\]

where the minimization is performed over all discrete val-
es of \( \mathbf{u} \) that fractionally sample the unit cell in real
space according to a specified interval along each cell
dimension. The sum is over the shared reflections be-
tween datasets \( \mathbf{c}_1 \) and \( \mathbf{c}_2 \). \( \mathbf{h} \) is a column vector of Miller
indices, and \( \bar{I}_{\mathbf{c}} \) is the mean intensity for reflection \( \mathbf{h} \).
Eq. 1 is similar to the phased translation function used in
molecular replacement searches, though here we compute
intensity-weighted phase residuals rather than calculat-
ing complex structure factors [54]. The term in paren-
thesis, \( \langle \phi_{\mathbf{c}_2} (\mathbf{h}) - 2\pi \mathbf{r} \mathbf{h} \rangle \), describes how the phases of \( \mathbf{c}_2 \)
change when its phase origin is shifted by \( \mathbf{u} \). This ex-
pression was then used to position the phases of all re-
flections of \( \mathbf{c}_2 \) on the reference origin. To merge the next
crystal, the reference phase set was updated to include
all reflections recorded in the original \( \mathbf{c}_1 \) and \( \mathbf{c}_2 \) datasets.
The reference phase set thus expanded with each merge
while the reference origin remained roughly fixed, with
only minor adjustments after averaging the phases from
different crystals. Datasets were merged in the order the
maximized the number of shared reflections between the
reference data and the dataset being merged at each step.
Once the final dataset was merged, the phase of each
reflection was estimated as the intensity-weighted mean
phase for all observations of that reflection.

Intensity information is unaffected by the choice of
phase origin due to translational invariance. For consis-
tency, however, we also treated the intensities as having
P1 symmetry during merging. Intensities from a second
dataset, \( I_{\mathbf{c}_2} \), were uniformly scaled to the first dataset,
\( I_{\mathbf{c}_1} \), by minimizing the sum of squared residuals:

\[
\arg \min_{m,b,\sigma} \sum_{\mathbf{h} \in (\mathbf{c}_1 \cap \mathbf{c}_2)} \left( \log(me^{-q^2\sigma^2}I_{\mathbf{c}_2}(\mathbf{h}) + b) - \log(I_{\mathbf{c}_1}(\mathbf{h})) \right)^2 \tag{2}
\]

where \( q \) is the magnitude of the reciprocal space vector
associated with reflection \( \mathbf{h} \), the exponential term is the
Debye-Waller factor [55], and least-squares optimization
is used to determine the scaling parameters \( m, b, \) and
\( \sigma \). We found that the use of logarithmic residuals im-
proved the stability of the optimization algorithm. The
next dataset was then scaled to the averaged intensities
of the original \( \mathbf{c}_1 \) and \( \mathbf{c}_2 \) datasets. Once all datasets were
merged, each reflection’s intensity was computed as the
mean of all observations of that reflection.

E. Locating a crystallographic phase origin

As noted above, it is unlikely that the phase origin of
the merged dataset will coincide with a crystallographic
origin consistent with the crystal’s space group. In 2D
electron crystallography, the crystallographic origin was
found by testing fractional cell positions in the plane of
the repeating unit for fulfillment of phase constraints im-
posed by symmetry [47, 56]. If the plane group was not
known in advance, this origin refinement procedure was
performed for every possible plane group, and the plane
group that yielded the lowest phase residual was selected.
Crystal symmetry determination is easier for 2D crystals,
however, as there are only 17 plane groups in contrast to
230 space groups [47]. Here we assumed that the space
group was already known and extended the method of
origin refinement to 3D crystals as follows.

The unit cell was sampled using a real space grid with
equally-spaced nodes along each dimension, and each node
or fractional cell position was considered a candidate
origin. The phases of the merged data, \( \phi_0 \), were
shifted by the fractional cell vector, \( \mathbf{u} \), to this origin:

\[
\phi_\mathbf{u}(\mathbf{h}) = \phi_0(\mathbf{h}) - 2\pi \mathbf{r} \mathbf{h} \mathbf{u}. \tag{3}
\]

We then evaluated the following three metrics. The first
metric corresponded to the phase residual of symmetry-
equivalent reflections. The mean phase value for a
given reflection \( \mathbf{h} \) can be estimated from its symmetry-
equivalent reflections as follows [57]:

\[
\langle \phi_\mathbf{u}(\mathbf{h}) \rangle = \langle \phi_\mathbf{u}(\mathbf{h}_a) - 2\pi \mathbf{h}_a \mathbf{t}_a \rangle. \tag{4}
\]
where the symmetry-equivalent reflections, \( \mathbf{h} \) and \( \mathbf{h}_s \), are related by the translation vector, \( \mathbf{t}_s \), and the phase origin is \( \mathbf{u} \). We then computed the symmetry phase residual as the sum of the phase differences between this mean value and independent observations from the symmetry-equivalent reflections:

\[
R_{\text{sym, } \phi_u} = \frac{\sum_{\mathbf{h}} \sum_{\mathbf{h}_s} I(\mathbf{h}_s) \left| (\phi_u(\mathbf{h}_s) - 2\pi \mathbf{h}_s \cdot \mathbf{t}_s) - \langle \phi_u(\mathbf{h}) \rangle \right|}{\sum_{\mathbf{h}} \sum_{\mathbf{h}_s} I(\mathbf{h}_s)}
\]

(5)

where the double sum enumerates the set of unique reflections and, for each, the set of symmetry-equivalent reflections. The residual was intensity-weighted to favor stronger reflections.

The second metric considered the difference between centric reflections and their expected phase values. Centric reflections satisfy the condition \( \mathbf{h} \mathbf{R} = -\mathbf{h} \), where \( \mathbf{R} \) is the rotation matrix associated with reflection \( \mathbf{h} \) [57]. When reflection data are positioned on a crystallographic phase origin, the phases of centric reflections are restricted to a limited set of possible values. We computed the intensity-weighted mean residual between the observed phase values at each candidate origin \( \mathbf{u} \) and the expected phase values for these reflections as follows:

\[
R_{\text{cen, } \phi_u} = \frac{\sum_{\mathbf{h}} I_{\mathbf{h}} \min \left\{ \left| \phi_u(\mathbf{h}) - \pi \mathbf{t} \cdot \mathbf{h} \right|, \left| \phi_u(\mathbf{h}) - \pi (\mathbf{h} + 1) \cdot \mathbf{t} \right| \right\}}{\sum_{\mathbf{h}} I_{\mathbf{h}}}
\]

(6)

where \( \mathbf{t} \) is the translation vector associated with reflection \( \mathbf{h} \), and the sum is restricted to the observed centric reflections [58]. This metric was omitted for datasets that did not contain any centric reflections.

For the third metric, an electron density map of the unit cell was computed from the intensities of the merged dataset and the shifted phases, \( \phi_u \), using the ccdbz library [43]. The skew of the density values was evaluated, as the density distribution of macromolecular crystals exhibits a positive skew in contrast to the Gaussian distribution characteristic of random maps. This distinction is used to judge map quality during automated structure solution after experimental phasing [59]. The advantage of this metric is that it uses all available reflections, rather than just the subset of either centric reflections or reflections with high multiplicity. For consistency with the phase residual metrics, the negative of the skew was computed such that lower values indicated more probable origins.

These three metrics were normalized and summed with equal weighting to score each candidate origin given by the fractional cell vector, \( \mathbf{u} \). The fractional cell position with the lowest combined score was selected, and the phases of the merged dataset were shifted to this origin. The intensities and shifted phases were then reduced to the asymmetric unit, and symmetry constraints were imposed. Specifically, the phases of symmetry-equivalent reflections were mapped to their expected values in the asymmetric unit (see Eq. 4) and the intensity-weighted mean phase value was computed. The intensities of symmetry-equivalent reflections were averaged.

As an alternative approach, individual datasets could be positioned on a crystallographic origin prior to merging. In this case, the merging procedure would require a search over a limited number of positions rather than the entire unit cell, as space group symmetry restricts the number of crystallographic origins. However, we found that the process of merging two datasets was slightly more robust than positioning individual datasets on a crystallographic origin. We compared each strategy using simulated datasets of a \( P2_12_12_1 \) crystal with unit cell dimensions \( a,b,c=16.2, 29.1, 47,7 \) Å generated with a range of mean phase errors (0-40°), completeness (10-40% prior to accounting for space group symmetry), and relative B-factors (described in more detail in Section III; see Eq. 7). The searches for the common origin between datasets and crystallographic origin were both performed using a sampling interval of 0.2 Å; the phase origin was considered to be correctly found if the distance between the correct origin and the origin found by the algorithm was less than twice the sampling interval. We observed a 4% higher success rate for finding a common phase origin between two datasets than locating a crystallographic origin for each individual dataset (Fig. S6A). Further, identifying a crystallographic origin was successful in 8% more cases when the procedure was performed on the merged data from two tomograms than on a single tomogram (Fig. S6B). In contrast, the success rate for merging two and three datasets was similar (Fig. S6C). These trends suggest a modest advantage to merging datasets to a common phase origin prior to finding a crystallographic origin, as the latter procedure appeared to be more sensitive to the number of available reflections.

### III. Tests of robustness to radiation damage

We validated the full workflow presented in Section II (Fig. S1A) by evaluating the accuracy of the reflection data recovered from simulated tomograms. The processed data were compared to phases and intensities computed directly from the atomic model, in addition to reflection data recovered from “intact volumes” processed in the same manner as the simulated tomograms. Intact volumes refer to rotated crystal densities generated in the same manner as the simulated tomograms (see Section II), except without projection into tilt-series and reconstruction into tomograms. This comparison between intact volumes and simulated tomograms enabled assessing the inaccuracy and loss of completeness introduced by tomographic sampling beyond the baseline interpolation errors from simulating and rotating the crystal densities onto a discrete grid. For each structure (6D6G/4BFH) and type of volume (intact volume/tomogram), performance was judged based on merging five datasets. This merging procedure was repeated ten times with unique sets of five datasets.

Metrics evaluating the quality of the recovered reflection data are shown in Table 1. The information from the
FIG. 4. Merging undamaged tomograms recovers the reference density. Density maps were computed from reflection data for a representative intact volume (left), one tomogram (center), or the merge of five tomograms (right). The simulated crystal system had either (A) triclinic or (B) orthorhombic symmetry. Each panel visualizes the map for the entire protein (upper) and the subset of residues circled in red (lower).

intact volumes was consistent with the reference phases ($R_{\text{ref, } \phi} < 2.5^\circ$) and intensities (CC $> 0.97$) computed from the initial atomic models. For the simulated tomograms, merging multiple datasets was required to achieve high completeness. When the tilt-range spanned $120^\circ$, or 67% of reciprocal space, the data recovered from each tomogram were only 40% complete in the absence of internal symmetry (Table 1). The unexpectedly low completeness stems from the large angular increment of the tilt-series: many Bragg peaks lie between tilt images and were ei-
TABLE I. Quality of merged data

| No. crystals merged | Intact volumes | Undamaged tomograms, c=0.1 | Mildly damaged tomograms, c=0.2 | Moderately damaged tomograms, c=0.3 | Severely damaged tomograms, c=0.4 |
|---------------------|----------------|--------------------------|-------------------------------|-------------------------------|-------------------------------|
|                     |                 | ±60                      | ±60                          | ±60                          | ±60                          |
|                     | 1               | 5                        | 1                             | 5                             | 1                             |
| Completeeness, c^a  | 0.96            | 0.99                     | 0.39                          | 0.89                          | 0.34                          | 0.85                          | 0.30                          | 0.72                          | 0.30                          | 0.78                          |
| R_{ref,φ}^b (°)     | 2.4             | 2.6                      | 9.2                           | 6.2                           | 7.4                           | 5.5                           | 18.9                         | 15.0                          | 10.4                          | 8.8                           |
| CC(I_{vim}, I_{vor}) | 1.00            | 1.00                     | 0.71                          | 0.74                          | 0.79                          | 0.81                          | 0.59                         | 0.57                          | 0.65                          | 0.65                          |
| CC(map,model)^d     | 0.98            | 0.99                     | 0.58                          | 0.85                          | 0.56                          | 0.88                          | 0.50                         | 0.72                          | 0.50                          | 0.79                          |
|                     | 1               | 5                        | 1                             | 5                             | 1                             | 5                             | 1                            | 5                             | 1                             | 5                             |
| Completeeness       | 0.95            | 1.00                     | 0.73                          | 0.99                          | 0.67                          | 0.99                          | 0.61                         | 0.96                          | 0.62                          | 0.97                          |
| R_{ref,φ}^b (°)     | 1.0             | 1.0                      | 6.1                           | 2.8                           | 5.2                           | 2.2                           | 11.0                         | 4.9                           | 8.4                           | 3.5                           |
| CC(I_{vim}, I_{vor}) | 0.98            | 0.98                     | 0.81                          | 0.87                          | 0.86                          | 0.90                          | 0.72                         | 0.76                          | 0.73                          | 0.76                          |
| CC(map,model)^d     | 0.97            | 0.99                     | 0.79                          | 0.92                          | 0.78                          | 0.95                          | 0.69                         | 0.87                          | 0.72                          | 0.88                          |

^a To a high-resolution limit of 3.3 Å. Entries show the mean value from ten runs of merging the indicated number of datasets consisting of intact volumes, undamaged tomograms, or tomograms simulated with radiation damage to achieve the specified fractional P1 completeeness (c).

^b The intensity-weighted mean residual between phases from the datasets reduced to the asymmetric unit and reference phases computed directly from the atomic model.

^c The correlation coefficient between the logs of reflection intensities computed from the reduced dataset and atomic model.

^d The correlation coefficient between the real space maps computed from the reduced dataset and atomic model.

ther not observed or discarded as poorly sampled. Generating tomograms using a ±40° tilt-range with 2° increments rather than a ±60° tilt-range with 3° increments resulted in a slight decrease in the overall completeness (~5%) but a modest improvement in the accuracy of the reflection data due to the finer sampling (Table 1). Merging tomograms and the presence of internal symmetry also improved the accuracy of the recovered phases, which showed good agreement with reference (R_{ref,φ} of 6.2° and 2.8° for the P1 and P2_12_1 crystal systems, respectively, when the tilt-range spanned ±60°).

In real space, loss of information due to the missing wedge leads to elongation of density in the direction of the beam [60]. These anisotropic effects were apparent in density maps computed from single tomograms of the P1 crystal, with continuous density along parallel bands and gaps along the protein backbone between these streaks (Fig. 4A). Incorrect connectivity was less pronounced in density maps computed from single tomograms of the orthorhombic crystal, as the presence of internal symmetry mitigated the directionality of the missing wedge effect (Fig. 4B). In both cases, merging multiple crystals improved the correlation with the reference map (Table 1).

Radiation damage is a limiting factor in cryo-ET and results in loss of information, particularly at high-resolution, as data collection progresses [61]. To assess how this impacts recovery of the reference structure, we next applied the workflow to tomograms of these crystal systems with simulated damage. Damage events were assumed to occur at random sites in the crystal volume, and each “hit” was modeled as a blurring of the local density [62]. The blurring was performed by replacing a cubic subvolume of length 5 Å around each selected site with its Gaussian-filtered copy, using a standard deviation of 1 Å along each axis for the 3D Gaussian kernel. Following a dose-symmetric tilt scheme [53], an equal number of hits was applied to the randomly-oriented crystal before computing each projection image to mimic a linear increase in absorbed dose across the tilt-series. Tomograms were reconstructed from these damaged tilt-series, and those that could not be indexed to the correct space group were discarded. For both crystal systems, we calibrated the total number of hits, or effective dose, that on average yielded tomograms with a completeness of 30, 20, or 10% before accounting for space group symmetry, compared to the ~40% P1 completeness observed for undamaged tomograms (Table 1). To focus on the tolerance of data processing to low completeness, ten unique sets of five tomograms were merged if the completeness of each tomogram and the set’s average was respectively within 2% and 1% of the target completeness. Metrics evaluating the average quality of the individual tomograms and merged datasets are reported in Table 1.

For both crystal systems, the reference structure was consistently recovered by merging tomograms that were each 30% complete prior to accounting for space group symmetry. Merging improved the accuracy of the phases and resolved connectivity errors observed in real space maps computed from a single damaged tomogram (Table 1, Fig. 5). While an initial completeness of 20% could also be tolerated by the orthorhombic system, results for the triclinic crystal were variable, with the phase accuracy frequently decreasing during the course of merging despite the increase in multiplicity (Fig. S7). For both crystal systems, an initial completeness of 10% per tomogram was prohibitive to recovering the reference structure. Despite the modest increase in cross-correlation with the reference map, merging reduced phase accuracy under this starting condition, indicating that a common phase origin was not correctly found (Fig. S7). The density of the merged tomograms showed the wrong connectivity, including smearing of density in the solvent region between neighboring chains and gaps along the backbone (Fig. 5, right). In several cases, the merging procedure visibly aligned the missing wedges of the tomograms being merged (Fig. S8), suggesting that the shape of the missing wedge rather than the reflection data dominated the signal. Similar trends in the relationship between...
initial completeness and failure to merge were observed when damaged tomograms were reconstructed from tilt-series spanning either a $\pm 60^\circ$ tilt-range with $3^\circ$ increments or a $\pm 40^\circ$ tilt-range with $2^\circ$ increments, despite the lower initial phase errors of the latter (Table 1).

Although this analysis suggests that $>10\%$ completeness per dataset is required for recovery of the correct structure, loss of overall completeness is not the only hallmark of radiation damage. The progressive accumulation of damage alters the spatial distribution of the reflection data by reducing the number of reflections recorded late in the tilt-series, or at high tilt angles for a dose-symmetric scheme. In addition, the simulated damage introduced phase errors, though the loss of phase accuracy varied for the same amount of simulated dose (Table 1). To disentangle the impact of incompleteness, the angular spread of reflections across the tilt-range, and phase error on the ability to merge tomograms, we simulated radiation damage in reciprocal space according to the following model. For each simulated dataset, the crystal lattice was subjected to a random rotation in 3D. Structure factors were then computed to 3.3 Å, and the positions of Bragg peaks were predicted as a function of tilt angle; reflections in the missing wedge region were excluded. Reflection intensities were modeled as decaying with a B-factor that increased linearly with dose [62]:

$$ I_d(h) = I_0(h)e^{-\frac{1}{8\pi 2 n B_f q^2}} $$

where $I_0$ is the intensity of the undamaged reflection, $q$ is the magnitude of the reciprocal space vector associated with reflection $h$, $B_f$ is a relative B-factor, and $n$ is the image number in the tilt-series, with $n=1, 2, 3, \ldots$ corresponding to tilt angles $0, -3, +3, -6, \ldots$ or $0, -2, +2, -4, \ldots$ for a dose-symmetric scheme spanning $\pm 60$ or $\pm 40$, respec-

FIG. 5. Merging tomograms with damage-induced loss of completeness. Density maps were computed from reflection data recovered from one representative undamaged tomogram or damaged tomogram with the specified $P1$ completeness (upper), and after merging five tomograms of the indicated type (lower). A subset of residues is visualized for the (A) triclinic and (B) orthorhombic crystal systems.
FIG. 6. The dependence of merging success on phase accuracy, completeness, and the angular spread of reflection data. A total of 25,920 simulated datasets were generated across three crystal systems, spanning a range of P1 completeness, phase errors, and relative B-factors. Pairs of datasets with similar starting characteristics and in random orientations were merged. (A) The fractional merge error was computed as the magnitude of the vector error between the true and estimated fractional shifts required to position datasets on a common phase origin. This metric is shown as a function of the mean phase error (left), P1 completeness (center), and how unevenly spread the reflections were across the tilt-range (right) for the datasets being merged. The latter was estimated as the Jensen-Shannon (JS) distance to a uniform angular distribution spanning ±60° (see Eq. 8). Though 2880 pairs of datasets with a phase error of 0° were merged, the results are visually superimposed in the leftmost panel. (B) Merge errors are shown as a function of these three parameters or (C) the two indicated parameters of the datasets being merged. For (B) and (C), red and black respectively indicate datasets for which the incorrect and correct origin shift was determined by the merging algorithm. The error threshold was set to be twice the sampling interval used by the algorithm, such that the correct phase origin was found only when the fractional merge error was <0.03.

Structure solution from tomograms of nanocrystals depends on finding the origin shift that correctly aligns the phases of multiple datasets. We assessed the accuracy of this merging procedure based on the merge error, the magnitude of the vector difference between the fractional origin shift estimated by our merging algorithm and the true shift required to align two datasets each subjected to a random phase shift. We considered the correct origin to be found when the merge error was less than twice the sampling interval used by the merging algorithm, corresponding to a fractional merge error of <0.03 for each crystal system. We then examined the dependence of this merge error separately on the phase accuracy, overall completeness, and the spatial distribution of the reflections of the datasets being merged. For the latter, we measured how unevenly the reflection data were distributed across the tilt-range using the Jensen-Shannon (JS) distance. This statistical metric measures the difference between a probability distribution of interest, p, and a reference probability distribution, q:

$$JS = \sqrt{\frac{1}{2} (D(p|m) + D(q|m))}$$

(8)

where p is the normalized distribution of tilt angles for the observed reflections, q is the normalized uniform dis-
tribution spanning ±60°, and \( m = \frac{1}{2}(p + q) \). \( D \) is the Kullback-Leibler divergence given by:

\[
D(x||y) = \sum_{i, x(i) \neq 0} x(i) \log \frac{x(i)}{y(i)}
\]

where the width of the angular bin \( i \) was set to 1°.

Using our chosen reference distribution, the Jensen-Shannon distance measures how unevenly reflections are distributed across a full tilt-range of ±60°, and the score is independent of the dataset's completeness. Datasets that span a narrower tilt-range than ±60° are penalized even when the reflection data are uniformly distributed since the normalized counts differ from the expected frequencies (Fig. 6A). Radiation damage also increases the Jensen-Shannon distance due to loss of Bragg peaks at higher tilt angles as data collection progresses (Fig. S9).

The results from the three simulated crystal systems were pooled to examine trends that held across different space group symmetries and a range of unit cell volumes. We found that the initial phase error and the reflections’ angular spread across the tilt-series, as measured by the Jensen-Shannon distance, better predicted the likelihood of merging success compared to overall completeness (Fig. 6A). Provided that the reflections from individual tomograms were approximately evenly spread across the tilt-range (\( JS < 0.18 \)), the correct origin shift could be determined even in the presence of an average phase error of up to 40°. By contrast, even relatively high completeness (>50%, without accounting for space group symmetry) did not guarantee finding the correct phase origin in the presence of moderate phase errors (Fig. 6B-C). These trends argue for distributing the dose across as wide a tilt-range as possible to maximize the angular spread of reflections available from each tomogram. Although the real-space simulations indicated that a data collection scheme spanning ±40° with 2° increments would reduce phase errors relative to ±60° with 3° increments due to the narrower tilt increment (Table 1), these results predict that the modest increase in phase accuracy will be outweighed by the increased difficulty in correctly merging datasets due to the narrower angular distribution of the reflection data across the tilt-range.

IV. Discussion

Cryo-ET of protein nanocrystals could deliver a method of high-resolution structure determination that both retains experimental phases and circumvents the need for large crystals in conventional X-ray crystallography and high molecular weight proteins in SPR. Here we describe a data processing pipeline to solve structures from tomograms of nanocrystals. This workflow both leverages software from X-ray crystallography and provides new algorithms to handle challenges unique to tomographic data of crystalline specimens. These challenges are related to correctly extracting the reflection phases, including eliminating phase splitting due to imperfect periodicity, excluding partial reflections that result in phase errors of 180°, and extending the phase origin search procedure from 2D plane groups to 3D space groups. We validated this data processing scheme using simulated crystals and found that the recovered reflection information was accurate, yielding maps with a correlation coefficient to reference of ~0.9 after merging five tomograms without any structure refinement.

We also assessed the robustness of this pipeline to radiation damage, which limits the effectiveness of structure determination methods. Merging tomograms of crystals in different orientations is required to compensate loss of completeness, which results from both radiation damage and the missing wedge. We found that this merging procedure was especially sensitive to the angular spread of the reflection data across the tilt-range and phase errors but less so to the completeness of the individual tomograms. These simulations indicate a trade-off between a wider tilt-range to facilitate merging datasets and a finer tilt increment to reduce phase errors. Since including more datasets can overcome phase errors but not incorrect origin shifts, these results recommend data collection strategies that maintain a wide tilt-range rather than decreasing the sampling interval. We also predict that a wider tilt-range would be favored over finer sampling for high-resolution subtomogram averaging to similarly avoid aligning the missing wedge experienced by individual particles during alignment.

While this workflow addresses several data analysis issues specific to tomograms of nanocrystals, additional challenges remain. One critical issue is the optimal specimen thickness: thicker crystals increase signal-to-noise, but the accuracy of the signal suffers due to a larger defocus gradient and increased multiple scattering [26, 37]. We anticipate that the defocus gradient can be largely accounted for by a 3D correction of the contrast transfer function [63, 64], and micorED has shown empirically that multiple scattering introduces only marginal (~5%) errors to the measured intensities for crystals 0.5-1 μm thick [65, 66]. However, the impact of these effects on phase accuracy has not been characterized. Another common tomography challenge is the poorer image quality at high tilt-angles. Given the sensitivity of the merging procedure to the reflections’ distribution in reciprocal space, simulations that additionally account for loss of image quality at high tilt-angles may be useful for further refining data collection strategies from nanocrystals.

Despite these technical challenges, recent successes in cryo-ET of noncrystalline specimens support applying this technique to solve structures from nanocrystals to high resolution. In particular, subtomogram averages of purified immature HIV-1 Gag particles to 3.1 Å and in situ ribosomes to 3.7 Å demonstrate that high-resolution signal is available in tilt-series data [20, 21]. As with SPR, only particles with high molecular weight can be aligned with sufficient accuracy for this high-resolution information to be retained during subtomogram avg-
ing. Alignment requires both determining the relative orientations of particles and positioning them on a common phase origin.

In comparison to SPR, the relative orientations of constituent proteins in a nanocrystal can be readily determined by indexing the coherent signal from the entire crystal. Even tiny nanocrystals will have a higher molecular weight than an individual HIV-1 Gag hexamer or ribosome, thereby extending the reach of cryo-ET to smaller proteins. In comparison to X-ray crystallography, both microED and cryo-ET of nanocrystals require a wider angular range of reflection data for indexing due to the negligible curvature of the Ewald sphere using electron radiation. Though successful indexing of individual electron diffraction stills has been reported [67], in general a tilt-range of ±20° is considered necessary to index microED data [68]. Cryo-ET data pose additional challenges, as imaging but not diffraction data are negatively impacted by several factors including the defocus gradient, translational drift, and loss of coherency due to inelastic scattering. These effects have been predicted to cumulatively reduce the signal-to-noise ratio of projection images by over 90% compared to diffraction stills collected from the same crystal, with increasing loss at higher resolution and for thicker crystals [69]. The diminished signal-to-noise will result in fewer observed reflections when collecting data in imaging versus diffraction mode for a fixed dose or exacerbate radiation damage if the dose is increased to compensate. Either scenario could make indexing more difficult for cryo-ET compared to microED data. Here we adapted algorithms from the crystallography software package DIALS to index the Fourier transforms of simulated nanocrystal tomograms [41, 42]. We found that indexing was robust even for simulated tomograms with low completeness.

While the use of nanocrystals facilitates determining relative particle orientations, the low molecular weight of the constituent proteins makes finding a common phase origin – the other step of particle alignment – more challenging. High-resolution subtomogram averaging has exclusively targeted high molecular weight proteins, and with two exceptions, isolated particles in solvent [19–21, 70]. These characteristics facilitate finding a common phase origin from each particle’s center of mass. For nanocrystals composed of small and densely-packed proteins, a more extensive search over the unit cell volume is required to position datasets on a consistent phase origin. Finding a common phase origin to merge datasets is critical for improving the accuracy of the recovered reflection data and overcoming low completeness.

In the future, we anticipate that cryo-ET of nanocrystals could enable structure determination from disordered crystals, which are typically discarded during diffraction experiments. In crystals, disorder attenuates the high-resolution signal and has been observed at length scales relevant for nanocrystals [25, 35]. In contrast to diffraction methods, imaging permits spatially characterizing and computationally correcting such disorder. One approach is to denoise the tomogram in Fourier space, followed by computationally “unbending” lattice distortions using cross-correlation analysis between the denoised and original tomograms. Lattice unbending overcame the resolution-limiting effects of disorder in 2D electron crystallography, in which specimens were frequently bent or wrinkled [26, 27]. Alternatively, subtomogram averaging could be performed in real space [38]. This technique has achieved subnanometer reconstructions of pleomorphic viruses with imperfect helical symmetry and could prove similarly useful for solving structures from tomograms of disordered nanocrystals [71]. In addition to extending the high-resolution limit of structure determination, the ability to spatially characterize disorder could provide insights into both the organization of proteins that form ordered arrays in vivo and defects in these biological crystals [72–74].

In addition, we anticipate that the development of a hybrid method involving microED and cryo-ET of nanocrystals could become routine. This approach would combine diffraction intensities with low-resolution phases from experimental images to solve an initial structure, followed by phase extension to the high-resolution limit of the microED data [4, 75]. Combining these techniques will be facilitated by the fact that sample preparation is shared between the techniques, and data can be collected on the same instrument [10]. Implementation of both disorder corrections and a hybrid approach with microED will be critical to realize the full potential of cryo-ET of nanocrystals for high-resolution structure determination.

**Code availability**

The code developed to process tomograms of nanocrystals is available at https://github.com/apeck12/cryoetX.

**Acknowledgments**

We thank Lauren Ann Metskas and Florian Schur for valuable discussions, in addition to David Stokes and Steven Ludtke for advice on the phase splitting phenomenon. A.P. is The Mark Foundation for Cancer Research Fellow of the Damon Runyon Cancer Research Foundation (DRG 2361-19). This work was supported by NIH grants R35 GM122588 (to G.J.J), AI150464 (to G.J.J), and GM117126 (to N.K.S).

[1] H. R. Powell, X-ray data processing, Bioscience Reports 10.1042/BSR20170227 (2017).

[2] J. L. Smith, R. F. Fischetti, and M. Yamamoto, Microcrystallography comes of age, Curr. Opin. Struct. Biol.
[3] J. M. Holton and K. A. Frankel, The minimum crystal size needed for a complete diffraction data set, Acta Crystallogr. D Biol. Crystallogr. 66, 393 (2010).

[4] G. Taylor, The phase problem, Acta Crystallogr. D Biol. Crystallogr. 59, 1881 (2003).

[5] I. Schlichting, Serial femtosecond crystallography: the first five years, IUCrJ 2, 246 (2015).

[6] W. Liu, D. Wacker, C. Gati, G. W. Han, D. James, W. I. Weis, A. Barty, M. A. Hough, D. Li, M. Messerschmidt, S. Boutet, G. J. Jensen, R. Chatterjee, S. Gul, F. Fuller, I. Schlichting, Serial femtosecond crystallography: the first five years, IUCrJ 2, 246 (2015).

[7] L. Zhu, U. Weierstall, V. Cherezov, and W. Liu, Serial Femtosecond Crystallography of Membrane Proteins, Adv. Exp. Med. Biol. 922, 151 (2016).

[8] I. D. Young, M. Ibrahim, R. Chatterjee, S. Gul, F. Fuller, S. Koroidov, A. S. Brewster, R. Tran, R. Alonso-Mori, T. Kroll, T. Michels-Clark, H. Laksmono, R. G. Sierra, C. A. Stan, R. Hussein, M. Zhang, L. Douthit, M. Kubin, C. de Lichtenberg, P. Long Vo, H. Nilsson, M. H. Cheah, D. Shevada, C. Saracini, M. A. Bean, I. Seuffert, D. Sokaras, T. C. Weng, E. Pastor, C. Weninger, T. Fransson, L. Lassalle, P. Brauer, P. Aller, P. T. Dockier, B. Andi, A. M. Orville, J. M. Glownia, S. Nelson, M. Sikorski, D. Zhu, M. S. Hunter, T. J. Lane, A. Aquila, J. E. Koglin, J. Robinson, M. Liang, S. Boutet, A. Y. Lyubimov, M. Uervirojnangkoorn, N. W. Moriarty, D. Liesbover, P. V. Afonine, D. G. Waterman, G. Evans, P. Wernet, H. Dobbeek, W. I. Weis, A. T. Brunger, P. H. Zwart, P. D. Adams, A. Zouni, J. Messinger, U. Bergmann, N. K. Sauter, J. Kern, V. K. Yachandra, and J. Yano, Structure of photosystem II and substrate size needed for a complete diffraction data set, Acta Crystallogr. D Biol. Crystallogr. 66, 246 (2015).

[9] A. Ebrahim, T. Moreno-Chicano, M. V. Appleby, A. K. Chaplin, J. H. Beale, D. A. Sherrell, H. M. E. Duyvesteyn, S. Owada, K. Tono, H. Sugimoto, R. W. Strange, J. A. R. Worrall, D. Axford, R. L. Owen, and M. A. Hough, Dose-resolved serial synchrotron and XFEL structures of radiation-sensitive metalloproteins, IUCrJ 6, 543 (2019).

[10] J. A. Rodriguez, D. S. Eisenberg, and T. Gonen, Taking the measure of MicroED, Curr. Opin. Struct. Biol. 46, 79 (2017).

[11] J. A. Rodriguez, M. I. Ivanova, M. R. Sawaya, D. Cascio, F. E. Reyes, D. Shi, S. Sangwan, E. L. Guenther, L. M. Johnson, M. Zhang, L. Jiang, M. A. Arbing, B. L. Nannenga, J. Hattne, J. Whitelegge, A. S. Brewster, M. Messerschmidt, S. Boutet, N. K. Sauter, T. Gonen, and D. S. Eisenberg, Structure of the toxic core of α-synuclein from invisible crystals, Nature 525, 486 (2015).

[12] M. J. de la Cruz, J. Hattne, D. Shi, P. Seidler, J. Rodriguez, F. E. Reyes, M. R. Sawaya, D. Cascio, S. C. Weiss, S. K. Kim, C. S. Hinck, A. P. Hinck, G. Calero, D. Eisenberg, and T. Gonen, Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED, Nat. Methods 14, 399 (2017).

[13] Y. Cheng, Single-Particle Cryo-EM at Crystallographic Resolution, Cell 161, 450 (2015).

[14] R. Henderson, The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules, Q. Rev. Biophys. 28, 171 (1995).

[15] G. J. Jensen, Alignment error envelopes for single particle analysis, J. Struct. Biol. 133, 143 (2001).

[16] M. Wu and G. C. Lander, How low can we go? structure determination of small biological complexes using single-particle cryo-em, Current Opinion in Structural Biology 64, 9 (2020).

[17] L. Brocchieri and S. Karlin, Protein length in eukaryotic and prokaryotic proteomes, Nucleic Acids Res. 33, 3390 (2005).

[18] C. M. Oikonomou and G. J. Jensen, Cellular Electron Cryotomography: Toward Structural Biology In Situ, Annu. Rev. Biochem. 86, 873 (2017).

[19] F. K. Schur, Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging, Curr. Opin. Struct. Biol. 58, 1 (2019).

[20] B. A. Himes and P. Zhang, emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging, Nat. Methods 15, 955 (2018).

[21] D. Tegunov, L. Xue, C. Dienemann, P. Cramer, and J. Mahamid, Multi-particle cryo-em refinement with m visualizes ribosome-antibiotic complex at 3.7 Å inside cells, bioRxiv 10.1101/2020.06.05.136341 (2020).

[22] M. W. Martynowycz, C. Glynn, J. Miao, M. Jason de la Cruz, J. Hattne, D. Shi, D. Cascio, J. Rodriguez, and T. Gonen, Microed structures from micrometer-thick protein crystals, bioRxiv 10.1101/152504 (2017).

[23] V. Lučić, A. Rigort, and W. Baumeister, Cryo-electron tomography: the challenge of doing structural biology in situ, J. Cell Biol. 202, 407 (2013).

[24] P. N. Unwin and R. Henderson, Molecular structure determination by electron microscopy of unstained crystalline specimens, J. Mol. Biol. 94, 425 (1975).

[25] I. Nederlof, Y. W. Li, M. van Heel, and J. P. Abraham, Imaging protein three-dimensional nanocrystals with cryo-EM, Acta Crystallogr. D Biol. Crystallogr. 69, 852 (2013).

[26] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy, J. Mol. Biol. 213, 899 (1990).

[27] A. D. Schenk, D. Castano-Diez, B. Gipson, M. Arheit, X. Zeng, and H. Stahlberg, 3D reconstruction from 2D crystal image and diffraction data, Meth. Enzymol. 482, 101 (2010).

[28] D. Nicastro, C. Schwartz, J. Pierson, R. Gaudette, M. E. Porter, and J. R. McIntosh, The molecular architecture of axonomes revealed by cryoelectron tomography., Science 313, 944 (2006).

[29] J. M. Heumann, A. Hoenger, and D. N. Mastronarde, Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging, Nat. Methods 15, 955 (2018).

[30] D. Wacker, C. Gati, G. W. Han, D. James, W. I. Weis, A. Barty, M. A. Hough, D. Li, M. Messerschmidt, S. Boutet, G. J. Jensen, R. Chatterjee, S. Gul, F. Fuller, I. Schlichting, Serial femtosecond crystallography: the first five years, IUCrJ 2, 246 (2015).

[31] I. D. Young, M. Ibrahim, R. Chatterjee, S. Gul, F. Fuller, I. Schlichting, Serial femtosecond crystallography: the first five years, IUCrJ 2, 246 (2015).
ing subtomogram averaging in RELION, Nat Protoc 11, 2054 (2016).

[32] M. Chen, J. M. Bell, X. Shi, S. Y. Sun, Z. Wang, and S. J. Ludtke, A complete data processing workflow for cryo-ET and subtomogram averaging, Nat. Methods 16, 1161 (2019).

[33] R. N. Jackson, A. J. McCoy, T. C. Terwilliger, R. J. Read, and B. Wiedenheft, X-ray structure determination using low-resolution electron microscopy maps for molecular replacement, Nat Protoc 10, 1275 (2015).

[34] D. I. Stuart and N. G. Abrescia, From lows to highs: using low-resolution models to phase X-ray data, Acta Crystallogr. D Biol. Crystallogr. 69, 2257 (2013).

[35] M. Gallagher-Jones, C. Ophus, K. C. Bustillo, D. R. Boyer, O. Panova, C. Glynn, C. T. Zee, J. Ciston, K. C. Mancia, A. M. Minor, and J. A. Rodriguez, Nanoscale mosaicity revealed in peptide microcrystals by scanning electron nanodiffraction, Commun Biol 2, 26 (2019).

[36] E. van Genderen, Y. W. Li, I. Nederlof, and J. P. Abrahams, Lattice filter for processing image data of three-dimensional protein nanocrystals, Acta Crystallogr. D Struct Biol 72, 34 (2016).

[37] G. Subramanian, S. Basu, H. Liu, J.-M. Zuo, and J. C. Spence, Solving protein nanocrystals by cryo-em diffraction: Multiple scattering artifacts, Ultramicroscopy 148, 87 (2015).

[38] W. Wan and J. A. Briggs, Cryo-Electron Tomography and Subtomogram Averaging, Meth. Enzymol. 579, 329 (2016).

[39] G. Tang, L. Peng, P. R. Baldwin, D. S. Mann, W. Jiang, I. Rees, and S. J. Ludtke, EMAN2: an extensible image processing suite for electron microscopy, J. Struct. Biol. 157, 38 (2007).

[40] M. Hohn, G. Tang, G. Goodyear, P. Baldwin, Z. Huang, P. A. Penczek, C. Yang, R. M. Glaeser, P. D. Adams, and S. J. Ludtke, Sparx, a new environment for cryo-em image processing, Journal of Structural Biology 157, 47 (2007), software tools for macromolecular microscopy.

[41] D. G. Waterman, G. Winter, J. M. Parkhurst, L. Fuentes-Montero, J. Hattne, A. S. Brewster, N. K. Sauter, and G. Evans, The DIALS framework for integration software, CCP4 News. Protein Crystallogr. 49, 16 (2013).

[42] G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gída, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter, and G. Evans, DIALS: implementation and evaluation of a new integration package, Acta Crystallogr D Struct Biol 74, 85 (2018).

[43] R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty, and P. D. Adams, The Computational Crystallography Toolbox: crystallographic algorithms in a reusable software framework, Journal of Applied Crystallography 35, 126 (2002).

[44] E. F. Petterson, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J Comput Chem 25, 1605 (2004).

[45] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD, J. Struct. Biol. 116, 71 (1996).

[46] N. K. Sauter, R. W. Grosse-Kunstleve, and P. D. Adams, Robust indexing for automatic data collection, J Appl Crystallogr 37, 399 (2004).

[47] S. Hovmöller, Crisp: crystallographic image processing on a personal computer, Ultramicroscopy 41, 121 (1992).

[48] J. W. Pfugrath, The finer things in X-ray diffraction data collection, Acta Crystallogr. D Biol. Crystallogr. 55, 1718 (1999).

[49] W. Kabsch, Integration, scaling, space-group assignment and post-refinement, Acta Crystallogr. D Biol. Crystallogr. 66, 133 (2010).

[50] J. Hattne, F. E. Reyes, B. L. Nannenga, D. Shi, M. J. de la Cruz, A. G. Leslie, and T. Gonen, MicroED data collection and processing, Acta Crystallogr A Found Adv 71, 353 (2015).

[51] D. N. Mastronarde, Dual-axis tomography: An approach with alignment methods that preserve resolution, Journal of Structural Biology 120, 343 (1997).

[52] Z. Dauter, Data-collection strategies, Acta Crystallogr. D Biol. Crystallogr. 55, 1703 (1999).

[53] W. J. H. Hagen, W. Wan, and J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging, J. Struct. Biol. 197, 191 (2017).

[54] R. J. Read and A. J. Schierbeek, A phased translation function, Journal of Applied Crystallography 21, 490 (1988).

[55] R. H. Blessing, D. Y. Guo, and D. A. Langs, Statistical expectation value of the Debye-Waller factor and E(hkl) values for macromolecular crystals, Acta Crystallogr. D Biol. Crystallogr. 52, 257 (1996).

[56] L. A. Amos, R. Henderson, and P. N. Unwin, Three-dimensional structure determination by electron microscopy of two-dimensional crystals, Prog. Biophys. Mol. Biol. 39, 183 (1982).

[57] S. Hovmöller, Rotation matrices and translation vectors in crystallography, IUCr: Commission on Crystallographic Teaching 39, 183 (1981).

[58] B. Rupp and K. Kantardjieff, Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology (Garland Science, 2010).

[59] T. C. Terwilliger, P. D. Adams, R. J. Read, A. J. McCoy, N. W. Moriarty, R. W. Grosse-Kunstleve, P. V. Afonine, P. H. Zwart, and L. W. Hung, Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard, Acta Crystallogr. D Biol. Crystallogr. 65, 582 (2009).

[60] M. Radermacher, Weighted back-projection methods, in Electron Tomography: Methods for Three-Dimensional Visualization of Structures in the Cell, edited by J. Frank (Springer New York, New York, NY, 2006) pp. 245–273.

[61] L. A. Baker and J. L. Rubinstein, Radiation damage in electron cryomicroscopy, Meth. Enzymol. 481, 371 (2010).

[62] H. Atakisi, L. Conger, D. W. Moreau, and R. E. Thorne, Resolution and dose dependence of radiation damage in biomolecular systems, IUCrJ 6, 1040 (2019).

[63] G. J. Jensen and R. D. Kornberg, Defocus-gradient corrected back-projection, Ultramicroscopy 84, 57 (2000).

[64] B. Toroño-va, F. K. Schur, W. Wan, and J. A. Briggs, Efficient 3d-ctf correction for cryo-electron tomography using novactf improves subtomogram averaging resolution to 3.4Å, Journal of Structural Biology 199, 187 (2017).

[65] D. Shi, B. L. Nannenga, M. G. Iadanza, and T. Gonen, Three-dimensional electron crystallography of protein microcrystals, 2, e01345 (2013).
B. L. Nannenga, D. Shi, A. G. W. Leslie, and T. Gonen, High-resolution structure determination by continuous-rotation data collection in MicroED, Nat. Methods 11, 927 (2014).

Y. Gevorkov, A. Barty, W. Brehm, T. A. White, A. Tolstikova, M. O. Wiedorn, A. Meents, R. R. Grigat, H. N. Chapman, and O. Yefanov, pinkIndexer - a universal indexer for pink-beam X-ray and electron diffraction snapshots, Acta Crystallogr A Found Adv 76, 121 (2020).

B. L. Nannenga and T. Gonen, The cryo-EM method microcrystal electron diffraction (MicroED), Nat. Methods 16, 369 (2019).

M. T. B. Clabbers and J. P. Abrahams, Electron diffraction and three-dimensional crystallography for structural biology, Crystallography Reviews 24, 176 (2018).

T. A. M. Bharat, D. Kureisaite-Ciziene, G. G. Hardy, E. W. Yu, J. M. Devant, W. J. H. Hagen, Y. V. Brun, J. A. G. Briggs, and J. L?we, Structure of the hexagonal surface layer on Caulobacter crescentus cells, Nat Microbiol 2, 17059 (2017).

W. Wan, L. Kolesnikova, M. Clarke, A. Koehler, T. Noda, S. Becker, and J. A. G. Briggs, Structure and assembly of the Ebola virus nucleocapsid, Nature 551, 394 (2017).

R. H. Lange, Crystalline islet B-granules in the grass snake (Natrix natrix (L.)): tilting experiments in the electron microscope, J. Ultrastruct. Res. 46, 301 (1974).

S. G. Wolf, D. Frenkel, T. Arad, S. E. Finkel, R. Kolter, and A. Minsky, DNA protection by stress-induced biocrystallization, Nature 400, 83 (1999).

L. A. Dadinova, Y. M. Chesnokov, R. A. Kamyshtinsky, I. A. Orlov, M. V. Petoukhov, A. A. Mozhaev, E. Y. Soshinskaya, V. N. Lazarev, V. A. Manuvera, A. S. Orekhov, A. L. Vasiliev, and E. V. Shtykova, Protective Dps-DNA co-crystallization in stressed cells: an in vitro structural study by small-angle X-ray scattering and cryo-electron tomography, FEBS Lett. 593, 1360 (2019).

K. Cowtan, Recent developments in classical density modification, Acta Crystallogr. D Biol. Crystallogr. 66, 470 (2010).