

Phasing of coherent femtosecond X-ray diffraction from size-varying nanocrystals

John C. H. Spence,^{1,*} Richard A. Kirian,¹ Xiaoyu Wang,¹ Uwe Weierstall,¹ Kevin E. Schmidt,¹ Thomas White,² Anton Barty,² Henry N. Chapman,^{2,3} Stefano Marchesini,⁴ and James Holton^{4,5}

¹Department of Physics, Arizona State University, Tempe, AZ 85287 USA.

²Center for Free Electron Laser Science DESY University of Hamburg, Notkestrasse 85, 22607, Hamburg, Germany

³University of Hamburg, Luruper Chaussee 149, Hamburg 22761, Germany

⁴Advanced Light Source, Lawrence Berkeley Laboratory, Berkeley, CA. 94720 USA

⁵Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158 USA.

*spence@asu.edu

Abstract: The scattering between Bragg reflections from nanocrystals is used to aid solution of the phase problem. We describe a method for reconstructing the charge density of a typical molecule within a single unit cell, if sufficiently finely-sampled “snap-shot” diffraction data (as provided a free-electron X-ray laser) are available from many nanocrystals of different sizes lying in random orientations. By using information on the particle-size distribution within the patterns, this digital method succeeds, using all the data, without knowledge of the distribution of particle size or requiring atomic-resolution data.

©2011 Optical Society of America

OCIS codes: (100.5070) Phase retrieval, (260.1960) Diffraction theory; (290.5840) Scattering, molecules; (320.7100) Ultrafast measurements.

References and links

1. N. Kasai, and M. Kakudo, “*X-ray diffraction by macromolecules*” (Springer New York 2005).
2. W. A. Barletta, J. Bisognano, J. N. Corlett, P. Emma, Z. Huang, K. J. Kim, R. Lindberg, J. B. Murphy, G. R. Neil, D. C. Nguyen, C. Pellegrini, R. A. Rimmer, F. Sannibale, G. Stupakov, R. P. Walker, and A. A. Zholents, “Free electron lasers: Present status and future challenges,” *Nucl. Instrum. Methods* **A618**, 69–96 (2010).
3. R. A. Kirian, X. Y. Wang, U. Weierstall, K. E. Schmidt, J. C. H. Spence, M. Hunter, P. Fromme, T. White, H. N. Chapman, and J. Holton, “Femtosecond protein nanocrystallography-data analysis methods,” *Opt. Express* **18**(6), 5713–5723 (2010).
4. H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, S. Stern, K. Nass, R. Andritschke, C. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. Holton, S. Marchesini, S. Schorb, D. Rupp, M. Adolph, T. Gorkhaver, M. Svenda, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson, and J. C. H. Spence, “Femtosecond X-ray protein nanocrystallography,” *Nature* Feb 3, (2011) in press.
5. D. Sayre, “Some implications of a theorem due to Shannon,” *Acta Crystallogr.* **5**(6), 843 (1952).
6. M. F. Perutz, “The structure of haemoglobin,” *Proc. Royal Soc.* **A264**, 264–286 (1954).
7. R. H. T. Bates, “Fourier phase problems are uniquely solvable in more than one dimension,” *Optik (Stuttg.)* **61**, 247–262 (1982).
8. J. C. H. Spence, in *Science of Microscopy*, eds., P. Hawkes and J. C. H. Spence, (Springer, 2006. New York).
9. J. C. H. Spence, and R. B. Doak, “Single molecule diffraction,” *Phys. Rev. Lett.* **92**(19), 198102 (2004).
10. C. Riek. “Recent developments in microdiffraction on protein crystals,” *J. Synchrotron Radiat.* **11**, 4–6 (2004). Also R. Von Dreele (2009) Personal communication: Results of powder X-ray diffraction from growth media.
11. M. von Laue, “The external shape of crystals and its influence on interference phenomena in crystalline lattices,” *Ann. Phys.* **26**, 55–68 (1936).

12. I. K. Robinson, and R. Harder, "Coherent X-ray diffraction imaging of strain at the nanoscale," *Nat. Mater.* **8**(4), 291–298 (2009).
 13. M. G. Rossmann, and J. W. Erickson, "Oscillation photography of radiation-sensitive crystals using a synchrotron source," *J. Appl. Cryst.* **16**(6), 629–636 (1983).
 14. A. G. Leslie, "The integration of macromolecular diffraction data," *Acta Crystallogr. D Biol. Crystallogr.* **62**(1), 48–57 (2006).
 15. F. Maia, C. Yang, and S. Marchesini, "Compressive auto-indexing in femtosecond nanocrystallography," *Proc. SPIE* **7800**, 78000F (2010).
 16. G. Xu, G. Zhou, and X. Zhang, "Phase recovery for X-ray crystallography," *Phys. Rev. B* **59**(14), 9044–9047 (1999).
 17. W. Press, B. P. Flannery, A. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*. C.U.P. (2007).
 18. A. Brunetta, M. Sanchez del Rio, B. Golosiob, A. Simionovicic, and A. Somogyi, "A library for X-ray-matter interaction cross sections for X-ray fluorescence applications," *Spectrochim. Acta [A]* **B59**, 1725–1731 (2004).
 19. D. A. Shapiro, H. N. Chapman, D. DePonte, R. B. Doak, P. Fromme, G. Hembree, M. Hunter, S. Marchesini, K. Schmidt, J. Spence, D. Starodub, U. Weierstall, *J. Synchrotron Radiat.* **15**, 593–599 (2008).
 20. S. Marchesini, "A unified evaluation of iterative projection algorithms for phase retrieval," *Rev. Sci. Instr.* **78**(1), 011301 (2007).
 21. V. Elser, I. Rankenburg, and P. Thibault, "Searching with iterated maps," *Proc. Natl. Acad. Sci. U.S.A.* **104**(2), 418–423 (2007).
 22. P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre, "Reconstruction of a yeast cell from X-ray diffraction data," *Acta Crystallogr. A* **62**(4), 248–261 (2006).
-

1. Introduction

We consider the phase problem for scattering from a large number of nanocrystals of varying size, an important problem in the field of protein crystallography. Over the past century, in view of the importance of structure-factor phases in determining atomic structure, each new solution to the crystallographic phase problem has led to major advances in structural biology. Anomalous diffraction, isomorphous replacement, molecular replacement and direct methods are widely used, but all have their limitations (needing chemical modifications to the sample, atomic-resolution data, limits on the number of atoms, or a similar solved protein [1]). It has been said that the discovery of a genuinely new structure in molecular biology requires phase *measurement*, rather than phasing by modeling, based on the protein database, since the return of such new structures to the database may contribute to model bias in the database. With the invention of the hard X-ray laser (e.g. Linac Coherent Light Source (LCLS) [2]), new types of diffraction data are becoming available [3,4] which offer new possibilities for direct phasing, not limited in this way. There is a particular need for a direct phasing method which does not require atomic-resolution data.

It has long been appreciated that if scattering could be obtained at points between Bragg reflections, the additional information would greatly assist solution of this problem [5]. For protein crystals, variations in water content have been used to vary the cell dimensions for this purpose [6]. Since they contain chiral alpha-helices, protein crystals are acentric, however structure-factors may be real for projections with inversion symmetry. The complex molecular transform "sampled" at the Bragg condition is sufficient to reconstruct the molecular charge density, consistent with the Shannon sampling theorem (we assume one molecule per unit cell throughout). Sampling of intensities at the half-integral Bragg condition completely determines the autocorrelation of the molecular density, from which a unique solution of phases can be determined [7]. We show how the finite size of a nanocrystal can provide the inter-Bragg scattering needed to recover the continuous molecular transform between Bragg peaks, so that iterative phasing methods [8] may then be used to solve the phase problem and recover a molecular image. If enough projections could be obtained from the same nanocrystal, one might treat an entire nanocrystal as a single, non-periodic object and phase it using these methods, however radiation damage and the difficulty of orienting sub-micron objects prevents this. We describe here a more efficient approach using large numbers of nanocrystals which does not require goniometry, and which allows for variations in crystal size.

The type of data required might be provided [3] by the new generation of free-electron hard X-ray lasers (e.g. LCLS [2]) in the serial crystallography geometry [9], where a steady stream of hydrated protein nanocrystals, of varying sizes and in random orientations, is sprayed in single file across the femtosecond X-ray beam, and the detector is read out after every pulse, ideally with one nanocrystal per exposure. It is then found that a ten femtosecond X-ray pulse terminates before significant radiation damage occurs to the sample. As a result of the photoelectron cascade which occurs after the incident pulse has terminated (and detection has ceased) the sample is vaporized [4]. By comparison with single-molecule diffraction, perfect nanocrystals increase peak (Bragg) intensity by N^6 (for N molecules on a side), provide a solution to the molecular orientation problem through Miller indexing, and provide a path to single-molecule imaging as N is reduced by filtration, while also allowing study of proteins which fail to produce large crystals [10]. Using short pulses instead of freezing to reduce damage allows structural analysis and in situ “snap-shot chemistry” at room temperature, while the continuous flow of fresh material allows pump-probe study of irreversible processes. Finally, we show here that this geometry offers an efficient direct solution to the phase problem. In the simplest optical terms, we show that a finite grating solves the phase problem.

For a finite crystallite, the sharp Bragg reflections are convoluted with a crystal shape-transform function, and the result is then modulated by the molecular transform [11]. The shape-transform therefore provides the required inter-Bragg scattering needed to solve the phase problem. (This has been used for the study of strain in inorganic nanocrystals recently [12]). Our simulations (Fig. 1(a)) show resulting “partial” reflections [13], in random orientations, each formed by one slice of the Ewald sphere through the three-dimensional shape transform at each lattice point [3]. As for a finite grating, the number of unit cells N between crystal facets along direction \mathbf{g} is equal to $p + 2$, if there are p subsidiary fringes between Bragg reflections along direction \mathbf{g} .

2. Phasing nanocrystals

Consider diffraction from a finite parallelepiped crystal, illuminated by a pulse of plane-polarized monochromatic incident radiation, with wavevector \mathbf{k}_i ($|\mathbf{k}_i| = 1/\lambda$) and small beam divergence. The diffracted photon flux I_n (photons/pulse/ pixel) at $\Delta\mathbf{k} = \mathbf{k}_i - \mathbf{k}_o$ produced by the n -th crystallite, consisting of $N(n) = N_1 \times N_2 \times N_3$ unit cells, is given by [3]

$$I_n(\Delta\mathbf{k}) = J_o r_e^2 P |F(\Delta\mathbf{k})|^2 |S_n(\Delta\mathbf{k})|^2 \Delta\Omega \quad (1)$$

Here $F(\Delta\mathbf{k})$ is the continuous scattering from one unit cell (molecule). J_o is the incident photon flux density (photons/pulse/area), r_e^2 the electron cross section, P a polarization factor, and $\Delta\Omega$ the solid angle subtended by a detector pixel at the sample. For small scattering angles Θ the resolution $d = \lambda/\Theta$. $S_n(\Delta\mathbf{k})$ is the transform of the truncated crystal lattice (an interference function resembling a sinc function laid down at each reciprocal lattice point). It has the form

$$\prod_{i=1}^3 \sin^2(N_i \Psi_i) / \sin^2(\Psi_i)$$

where the Ψ_i define the scattering geometry. We assume an integral number of primitive unit cells. Equation (1) is remarkable in that it separates the result of complex interference effects into a simple product of intensity factors (unlike in-line holography), the second of which (S_n) is purely geometric and provides “samples” of $|F(\Delta\mathbf{k})|^2$ sufficiently fine to permit phasing. Experimentally, the set of observed $\Delta\mathbf{k}$ may be determined using autoindexing software [3,13] to find the Miller indices of all Bragg reflections, using fractional indexing between lattice points, and this determines the orientational relationship between diffraction patterns. The molecular transform $F(\Delta\mathbf{k})$ between Bragg reflections is identical for all nanocrystals, but

the lattice transform $S_n(\Delta\mathbf{k})$ depends on the size and shape of a crystal, and differs between crystals. However $S_n(\Delta\mathbf{k})$ is the same in the neighborhood of every lattice point for a given crystal. These characteristics allow the two intensities $|F(\Delta\mathbf{k})|$ and $|S_n(\Delta\mathbf{k})|$ to be disentangled.

The oversampled molecular transform may be “demodulated” from the diffracted intensity in principle by dividing the interference function in Eq. (1) into the measured intensity on the left. The continuous molecular transform modulus $|F(\Delta\mathbf{k})|$ may then be phased by iterative methods, and so inverted to give a three-dimensional electron density map of the contents of one unit cell.

We first assemble the indexed patterns from all nanocrystals into a three-dimensional diffraction volume, and then average these oriented intensities over all snapshot diffraction patterns (from crystals of different size), to obtain the quantity

$$\langle I_n(\Delta\mathbf{k}) \rangle_n = J_o r_e^2 P \langle |F(\Delta\mathbf{k})|^2 |S_n(\Delta\mathbf{k})|^2 \rangle_n \Delta\Omega \quad (2)$$

Since $F(\Delta\mathbf{k})$ is identical in each pattern,

$$\langle I_n(\Delta\mathbf{k}) \rangle_n = J_o r_e^2 P |F(\Delta\mathbf{k})|^2 \langle |S_n(\Delta\mathbf{k})|^2 \rangle_n \Delta\Omega \quad (3)$$

and the molecular transform may be expressed as

$$|F(\Delta\mathbf{k})|^2 = \frac{\langle I_n(\Delta\mathbf{k}) \rangle_n}{J_o r_e^2 P \langle |S_n(\Delta\mathbf{k})|^2 \rangle_n \Delta\Omega} = \frac{\langle I_n(\Delta\mathbf{k}) \rangle_n}{D(\Delta\mathbf{k})} \quad (4)$$

It remains to determine the denominator $D(\Delta\mathbf{k})$, the mean shape transform, which depends on the particle size distribution, and which we now discuss. We note that unlike Eq. (1), which can only be inverted for individual nanoparticles when $|S_n(\Delta\mathbf{k})|^2$ is non-zero, $\langle |S_n(\Delta\mathbf{k})|^2 \rangle_n$ is non-zero and therefore can be inverted over a wide range around the Bragg condition.

If we knew the particle size distribution $G(N(n))$ where $N(n)$ defines the crystallite dimensions, we could write

$$\langle |S_n(\Delta\mathbf{k})|^2 \rangle_n = \frac{\sum_n |S_n(\Delta\mathbf{k})|^2 G(N(n))}{\sum_n G(N(n))}$$

avoiding the need to determine individual values of $N(n)$ for each nanocrystal from the fringe systems. The denominator D in Eq. (4) (the mean shape-transform) may however be simply obtained without knowledge of the particle size distribution using the fact that a sum over corresponding pixels around different reflections will smooth out the molecular transform (which is different at each lattice point) but preserve the shape transform (which is the same around each lattice point for one nanocrystal). Qualitatively we find D as follows. First define a Wigner-Seitz cell around every reciprocal lattice point, extending half-way to the neighboring lattice point. We then perform a periodic average over all these Brillouin zones, by summing them all into one, then redistributing this summed cell periodically throughout reciprocal space. Since the molecular transform varies between cells (see Fig. 1(b)), while the shape transform does not, the sum accumulates to give D in Eq. (4). In detail, we average all Bragg reflections in Eq. (2) located at reciprocal lattice points \mathbf{g}_{hkl} :

$$\langle I_n(\Delta\mathbf{k} - \mathbf{g}_{hkl}) \rangle_{n,hkl} = J_o r_e^2 P \langle |F(\Delta\mathbf{k} - \mathbf{g}_{hkl})|^2 |S_n(\Delta\mathbf{k} - \mathbf{g}_{hkl})|^2 \rangle_{n,hkl} \Delta\Omega \quad (5)$$

Angle brackets denote an average over n and \mathbf{g}_{hkl} . We sum corresponding pixels (differing by \mathbf{g}) in the three-dimensional diffraction volume within a small range of $\Delta\mathbf{k}$ around every reflection over all reflections, and also over all patterns from particles of different size (bring

the intensity around every \mathbf{g}_{hkl} to the origin, and sum the result over all crystals). The summed cell at the origin is then redistributed about each lattice point.

Since the lattice transform is identical when translating by any \mathbf{g}_{hkl} , we have

$$\langle I_n(\Delta\mathbf{k} - \mathbf{g}_{hkl}) \rangle_{n,hkl} = J_o r_e^2 P \langle |F(\Delta\mathbf{k} - \mathbf{g}_{hkl})|^2 |S_n(\Delta\mathbf{k})|^2 \rangle_{n,hkl} \Delta\Omega \quad (6)$$

The molecular transform intensity is uncorrelated with the finite-lattice transform so that

$$\langle I_n(\Delta\mathbf{k} - \mathbf{g}_{hkl}) \rangle_{n,hkl} = J_o r_e^2 P \langle |F(\Delta\mathbf{k} - \mathbf{g}_{hkl})|^2 \rangle_{n,hkl} \langle |S_n(\Delta\mathbf{k})|^2 \rangle_n \Delta\Omega \quad (7)$$

Rearranging terms, we may now express the needed average finite-lattice transform as

$$\langle |S_n(\Delta\mathbf{k})|^2 \rangle_n = \frac{\langle I_n(\Delta\mathbf{k} - \mathbf{g}_{hkl}) \rangle_{n,hkl}}{J_o r_e^2 P \Delta\Omega}, \quad (8)$$

which we insert into Eq. (4) to obtain an expression for the molecular transform, purely in terms of the measured diffracted intensities

$$|F(\Delta\mathbf{k})|^2 = \frac{\langle I_n(\Delta\mathbf{k}) \rangle_n}{\langle I_n(\Delta\mathbf{k} - \mathbf{g}_{hkl}) \rangle_{n,hkl}} \quad (9)$$

This self-normalising expression does not require experimental scaling factors.

The smallest crystal which can be phased by this method will be one which is too small to be indexed. The largest crystal which can be phased will depend on both the dynamic range of the detector, and the number of detector pixels. The method fails for crystals much larger than a mosaic block, or when there is insufficient scattering between Bragg reflections. A larger number of inter-Bragg pixels than the Nyquist requirement is already needed to extract structure factors from these partial reflections by the Monte Carlo method [3], which requires the three-dimensional volume of the shape transforms to be estimated. The number of such detector pixels on a side is about $4L/d$, for resolution d , with L the largest nanocrystal's largest dimension. (For sub-micron nanocrystals, the effects of crystal size on the angular width of a Bragg reflection dominate the mosaicity, beam divergence and energy spread effects normally considered in conventional protein crystallography). Nyquist sampling requires 4 independent measurements of the molecular transform for each reflection ("half-orders" in three dimensions). However these samples need not be equally spaced [15]. Eq. (9) is also limited by the dynamic range of the detector, which can be effectively increased using summed data. In general, the envelope of the shape transform falls as $1/\Delta k$ [16], so that midway between lattice points, as D approaches the noise limit, fluctuations in the denominator of Eq. (13) will amplify noise on the retrieved molecular transform. This may be controlled by using a Weiner filter [17]. This smooth damping function applied to Eq. (9) can be obtained from interpolation of the power spectrum of the numerator of Eq. (9), and has values near unity when noise is low, and small values otherwise. The need for such a function was avoided in these simulations by collecting sufficient data. To determine the number of diffraction patterns which must be recorded we consider a sum of M diffraction patterns from one subclass belonging to the same orientation (within error). With a small background count N_b per shot per pixel, and N_s signal counts per shot per pixel ($N_s \ll 1$ at high angles), the signal to noise ratio S , assuming Poisson noise is $S = MN_s / \sqrt{M(N_s + N_b)} \approx \sqrt{MN_s} (1 - N_b / 2N_s)$. By choosing the constant S (a commonly accepted value is $S = 5$) and using N_s from I_n in Eq. (1), and given N_b , this expression then gives the number of shots M in this orientation class for statistical significance of the "oversampled" intensity. Saturation of the most intense pixel may require several data sets

with different gains. Signal-to-noise ratio also provides a condition on the inversion of Eq. (4) by optimization methods. Shannon interpolation (for the Fourier relationship between the autocorrelation of the molecular density and $|F(\Delta\mathbf{k})|^2$) gives $K_{2a}(\Delta\mathbf{k})^{**}|F(\Delta\mathbf{k})|^2 = |F(\Delta\mathbf{k})|$, where $K_{2a}(\Delta\mathbf{k})$ is the Fourier transform of the doubled unit cell and $**$ denotes convolution. Combining this with Eq. (4):

$$\langle I_n(\Delta\mathbf{k}) \rangle_n = D(\Delta\mathbf{k})K_{2a}(\Delta\mathbf{k})^{**}|F(\mathbf{k})|^2 + \varepsilon \quad (10)$$

where ε is the noise contribution. This may be expressed in matrix notation as the linear equation

$$I = DK_m F + \varepsilon = AF + \varepsilon \quad (11)$$

where the condition number of \mathbf{A} must now exceed the signal-to-noise ratio S . Equation (11) may be solved using optimization methods, based on the pseudoinverse of \mathbf{A} .

3. Simulations

Figure 1 shows simulations [18] at 8 kV for the protein alpha-conotoxin PnIB (1AKG), crystallized with one molecule in a cubic unit cell ($a = b = c = 5.84$ nm, $\alpha = \beta = \gamma = 90^\circ$). A 512×512 pixel detector is assumed with 100 mm camera length and 120 micron pixels. The mean number of molecules $N = 20$ on a side of the nanocrystal (as observed experimentally [19]) with standard deviation $\sigma = 2$. Figure 1(a) shows the partial reflections expected from one nanocrystal. Figure 1(b) shows the modulus of the 3D molecular transform $|F(\Delta\mathbf{k})|$ on one plane through the origin of q -space. The beam-stop obscures the central maximum at the direct beam. Figure 2(a) shows the merged sum of $M = 10^6$ oriented diffraction patterns on this plane, assuming 10^{13} incident photons per shot with 0.5 micron beam diameter, giving 1.46×10^6 scattered photons per shot. The average finite lattice sum (Eq. (8)) shows a similar form to Fig. 2(a) but without intensity modulation, while Fig. 2(b) shows the molecular transform recovered using Eq. (9) for this noise level and particle size distribution. The results are evaluated using a standard goodness-of-fit R-factor $R(M)$ [5] comparing $F(\Delta\mathbf{k})_{\text{model}}$ with $F(\Delta\mathbf{k})_{\text{recov}}$ from Eq. (9), as shown in Fig. 3(a) (bold curve) for M summed patterns. Our results suggest that 10^6 patterns would be needed for phasing under these conditions. If the incident flux is reduced to 10^{11} photons per pulse, we find that 10^8 patterns are needed.

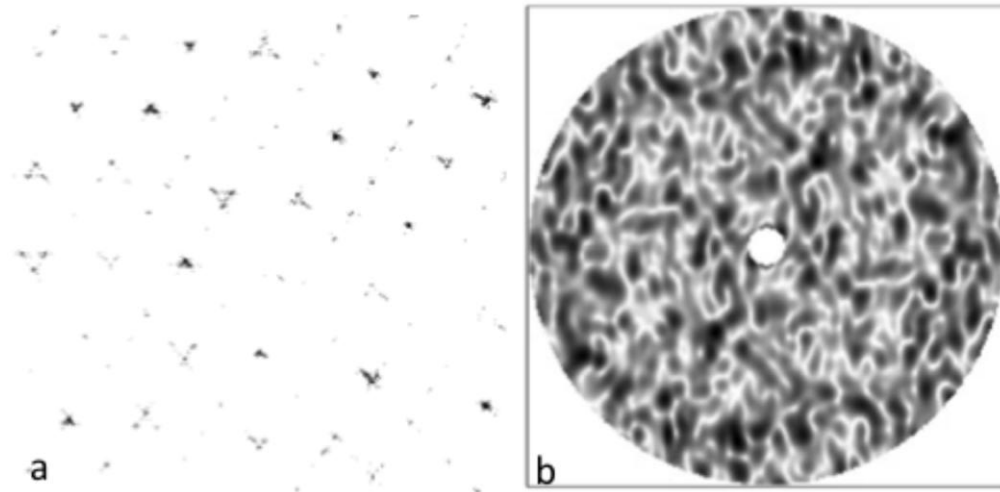


Fig. 1. (a) Simulated snap-shot diffraction pattern for X-rays in transmission through a sub-micron protein nanocrystal in a random orientation for some inner reflections near the origin. Shapes are partial reflections formed by 2D intersection of Ewald sphere with 3D Fourier

Transform of shape of nanocrystal near Bragg conditions. (b) $|F(\Delta\mathbf{k})|$ for 1AKG protein on [001] in \mathbf{q} -space, resolution $d = 0.523$ nm at side of pattern, 6 samples between Bragg spots.

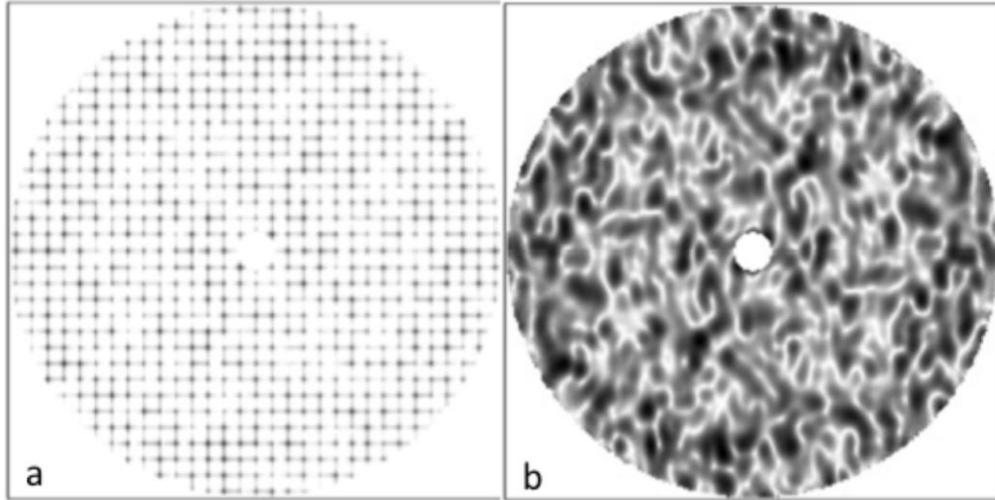


Fig. 2. (a) Sum of many patterns (Eq. (3)); crystal size $N = 20$, $\sigma = 2$, also on [001]. (b) The recovered molecular transform. Compare with 1(b)

The reconstruction of a charge-density map of the molecule then proceeds by phasing this transform, which may be based on any of the established iterative phasing methods [20,21] reviewed elsewhere [20,22], using a doubled cell. Results using shrinkwrap are shown in Fig. 3(b), and will depend on the detailed experimental conditions and size of beamstop. We note that the support of the molecule (the unit cell) is known. A large fraction of the unit cell for real protein crystals consists of water, not included here. Shannon's sampling theorem may be used as a constraint on predictions of the inter-Bragg scattering, once estimates of complex values at the reciprocal lattice points exist.

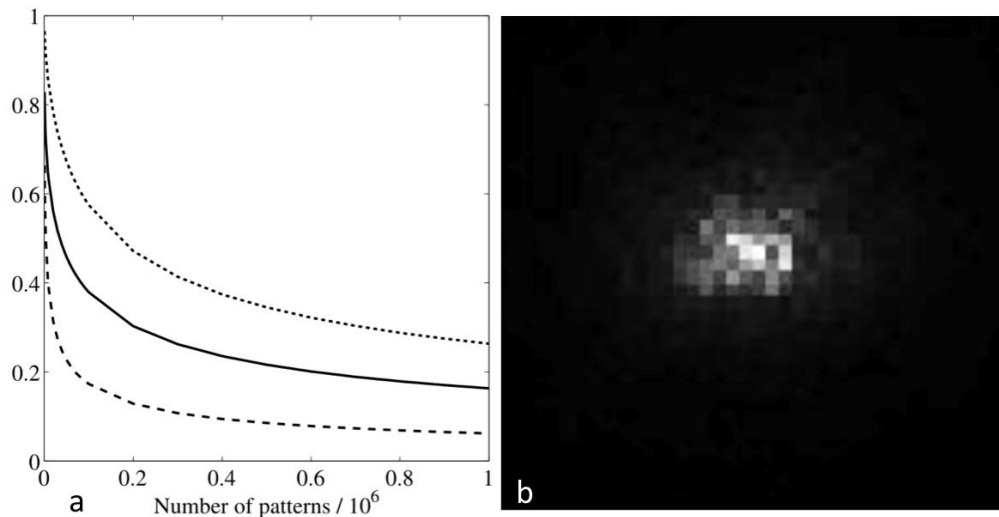


Fig. 3. (a) Bold curve: overall goodness of fit index $R(M)$, for $1 < M < 10^6$ crystals. Upper curve: $R(m)$ evaluated for the 50% of intensities which lie between reciprocal lattice points only. Lower: $R(m)$ evaluated for remaining 50% which lie near lattice points. (b) Projected low-resolution electron density slice in real space, after iterative phasing, with [100] normal.

The distance across the entire displayed area is 20nm, molecule enlarged for clarity.

In summary we have described the principle of a new crystallographic phasing method for data consisting of many diffraction patterns from nanocrystals of various sizes, in random orientations, containing identical molecules. No data of the required type (from untwinned nanocrystals) has thus far been collected at the LCLS at the fine pixel sampling rate required. We have assumed a transverse coherence width greater than the size of the largest nanocrystal, a requirement for this method. Calculations for higher resolution would be possible using greater computational time and memory - our purpose here is to demonstrate the principle of this method in a simple case. It is clear that this method could be combined with modelling information based on the molecular replacement technique. We provide an example of the method applied to simulated protein nanocrystal diffraction, and show that this new direct method for solving the phase problem does not require knowledge of the particle size distribution, independent measurement of experimental parameters, or atomic-resolution data.

Acknowledgements

We thank Profs. V. Elser and J. Hadju for useful discussions. Supported by NSF award MCB 1021557.