

# Structure of Single Particles from Randomly Oriented Ensembles Using an X-Ray Free Electron Laser

SS Kim, S Wibowo and DK Saldin\*

Department of Physics, University of Wisconsin-Milwaukee, USA

Received: December 01, 2017; Published: December 07, 2017

\*Corresponding author: DK Saldin, Department of Physics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA, Email: dksaldin@uwm.edu

## Abstract

The method of time-resolved crystallography offers a method of studying fast changes in structure during a chemical reaction. However, being a crystallographic method it is restricted to crystal structures. The problem with this is that most chemical reactions under physiological conditions happen in solution and not on molecules which are part of a crystal. The main problem with trying to find chemical reactions in solution is that one needs a powerful source of X-rays which will give a measurable signal even from small numbers of randomly oriented molecules. The newly developed X-ray free electron laser allows this when combined with a novel theoretical technique. Of course we understand one of the strengths of working with crystals is that the signal comes not from one but from trillions of identically oriented unit cells. The fact that the unit cells are identically oriented means that the X-rays do not have to scatter off a single molecule (or unit cell) to give a sensible signal. Since all molecules are identically oriented information about an idealized average molecule may be obtained from scattering by a crystal even though the different photons scatter off different molecules. The problem is that most chemical reactions do not take place in crystals.

Two main developments allow us to overcome this limitation. One is the development of the X-ray free electron laser (XFEL) which is capable of producing X-rays many orders of magnitude brighter than any previous X-ray source. If it is possible to determine the structure of a single particle without the need for crystals we will have achieved our goal. Luckily there has been a corresponding increase in understanding of scattering by disordered arrays. One of the things that has become realized recently is that if one concentrates not on the bare intensities of scattering, but on what are known as their angular correlations these said correlations are characteristic, over most of their range, of the structure but on not of the orientation of the molecules, provided the scattering is from a dilute disordered ensemble. Consequently, the way has been opened for the study of molecules via scattering by ensembles that are not identically oriented. What is more, techniques have further been developed for following fast changes in the structure of such molecules in a pump-probe experiment.

**Abbreviations:** XFEL: X-ray Free Electron Laser; LCLS: Linac Coherent X-ray Source; PYP: Photoactive Yellow Protein; LCLS: Linac Coherent Light Source

## Introduction

As evidenced by the Krebs (or citric acid) cycle [1] life functions via a series of chemical reactions. The experimental study of such reactions is therefore of great interest. Methods of studying such reactions date back at least to fem to second time-resolved photoelectron spectroscopy [2]. High spatial resolution was added by the use of x-rays in the technique of time-resolved crystallography [3]. This latter method originates in the era where there existed only weak X-ray sources. At that time X-rays were used to obtain information about a single molecule only by scattering off a large number of identically oriented copies. Luckily there exist natural structures (crystals) that many substances form which allow this. A typical crystal used in such work contains perhaps a trillion ( $10^{12}$ ) identically oriented molecules. A problem is that

most chemical reactions in nature do not take place in crystals. Indeed steric hindrance may prevent many important reactions in solution from ever happening in crystals [4].

Two recent developments may allow this limitation to be overcome. One is the development of a much brighter X-ray source known as a X-ray free electron laser (XFEL). The world's first XFEL, the Linac Coherent X-ray Source (LCLS) is now operational at Stanford in the U.S. Another such machine is now operational near Osaka in Japan. Others are planned, in many countries, e.g. Germany, Switzerland, Korea and China. It is true that these machines are very expensive and require the resources of a rich nation or group of nations. Consequently there is likely, by and large, to be only a single machine per country for the foreseeable future

(although a second XFEL is planned for the US). There has also been a corresponding improvement of theoretical understanding of fundamental diffraction theory as it had been realized the angular correlations amongst the scattered intensities are characteristic of the structure of a molecule but not of its orientation.

Chemical bonds tend to make each molecule identical apart from their orientations in a solution and the method of angular correlations is ideal to study such reactions. It is true the molecules are swimming in a sea of solvent molecules, in this case largely disordered water molecules. The effects of scattering by solvent have not been studied a lot, but the case of cryo-EM where the molecules swim in vitreous ice, the effect the vitreous ice, consisting of largely randomly oriented waier molecules, the solvent is found not to contribute much to the correlated scattering, and so can be ignored, except that coming from the solvent which might be about 100 times more prevalent that the solute may give rise to Poisson scattering that might interfere with the measured correlations. However, Poisson noise reduces the greater the intensity. Consequently, the scattering by a much larger solvent is actually advantageous.

One of the greatest experimental challenges with single particle work at an XFEL is the extremely low "hit rates" of about 0.1%. This means about 99.9% the beam hits empty solvent which contains no particle of interest. The ability we describe in this paper of obtaining structural information from ensembles of particles allows us to use a much higher concentration and overcomes the hit rate problem. Our hit rates are essentially 100%"

## Theory

Before proceeding further we would like to discuss angular pair correlations that play a central part of our theory. Suppose  $A_j$  are the scattered amplitudes from protein  $j$  of a set that is illuminated. The total intensity expected to be measured in a coherent instrument like an XFEL is thus

$$I(\mathbf{q}, \phi) = \left| \sum_j A_j(\mathbf{q}, \phi) \right|^2 \quad (1)$$

$$= \sum_j |A_j(\mathbf{q}, \phi)|^2 + \sum_{j \neq k} k A_j^*(\mathbf{q}, \phi) A_k(\mathbf{q}, \phi) e^{i\mathbf{q} \cdot (\mathbf{R}_k - \mathbf{R}_j)} \quad (2)$$

$$= \sum_j I_j(\mathbf{q}, \phi) + \sum_{j \neq k} k A_j^*(\mathbf{q}, \phi) A_k(\mathbf{q}, \phi) e^{i\mathbf{q} \cdot (\mathbf{R}_k - \mathbf{R}_j)} \quad (3)$$

Here the first summation contained on the incoherent (intensity) summation over different particles (already considered in [5-13] and [14] while the second summation contains the new (coherent) terms which give rise to interparticle interference. Likewise,

$$I(\mathbf{q}, \phi + \Delta \phi) = \sum_j I_j(\mathbf{q}, \phi + \Delta \phi) + \sum_{j \neq k} k A_j^*(\mathbf{q}, \phi + \Delta \phi) A_k(\mathbf{q}, \phi + \Delta \phi) e^{i\mathbf{q}' \cdot (\mathbf{R}_k - \mathbf{R}_j)} \quad (4)$$

Where  $\mathbf{q}'$  is the scattering vector corresponding to the point  $(\mathbf{q}, \phi + \Delta \phi)$  on the detector (in polar coordinates). Here also the first summation contains the incoherent terms and the second summation contains the interparticle interference additions. So for coherent scattering, the pair angular correlation defined by

$$C_2^{(\text{coherent})}(\mathbf{q}, \mathbf{q}, \Delta \phi) = \langle I(\mathbf{q}, \phi) I(\mathbf{q}, \phi + \Delta \phi) \rangle_{\text{DP}} \quad (5)$$

Where the expression  $\langle \dots \rangle_{\text{DP}}$  represents an average over the measured diffraction patterns and the  $I$ 's in the integrand may be taken from (3) and (4). Thus,

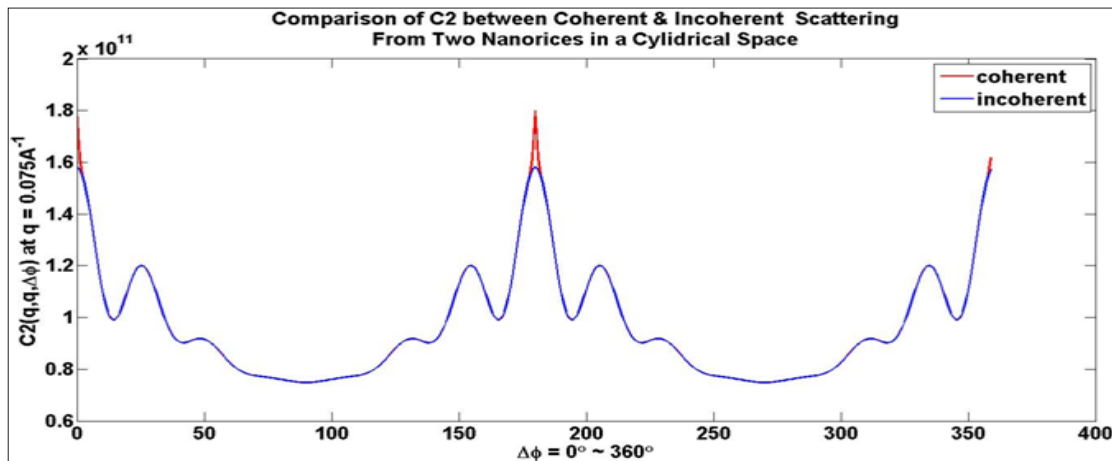
$$\begin{aligned} C_2^{(\text{coherent})}(\mathbf{q}, \mathbf{q}, \Delta \phi) &= C_2^{(\text{incoherent})}(\mathbf{q}, \mathbf{q}, \Delta \phi) + \dots \\ &+ \langle \int I_j(\mathbf{q}, \phi) \sum_{j \neq k} k A_j(\mathbf{q}, \phi + \Delta \phi) A_k^*(\mathbf{q}, \phi + \Delta \phi) e^{i\mathbf{q}' \cdot (\mathbf{R}_j - \mathbf{R}_k)} d\phi \rangle_{\text{DP}} \times \dots \\ &+ \langle \int I_j(\mathbf{q}, \phi + \Delta \phi) \sum_{j \neq k} k A_j^*(\mathbf{q}, \phi) A_k(\mathbf{q}, \phi) d\phi \rangle_{\text{DP}} e^{-i\mathbf{q} \cdot (\mathbf{R}_j - \mathbf{R}_k)} \\ &+ \langle \int \sum_{j \neq k} k A_j^*(\mathbf{q}, \phi) A_k(\mathbf{q}, \phi) A_j(\mathbf{q}, \phi + \Delta \phi) A_k^*(\mathbf{q}, \phi + \Delta \phi) e^{i(\mathbf{q} - \mathbf{q}') \cdot (\mathbf{R}_k - \mathbf{R}_j)} d\phi \rangle_{\text{DP}} \quad (6) \end{aligned}$$

Now in integrating over  $\phi$  neither  $\mathbf{q}$  nor  $\mathbf{q}'$  ever become zero unless one considers  $\mathbf{q}=0$  or  $\mathbf{q}'=0$ , which only exist under the beam stop, and are therefore not measured. Thus in practice the 2nd and 3rd terms on the RHS are negligible since they are associated with random phases. On the other hand, the 4th term has an argument of the exponential that can become zero even on the resolution ring, and therefore does not need to be small at the points where the argument becomes zero. Here  $C_2^{(\text{incoherent})}$  contains only contributions from the first incoherent summations in (3) and (4) in addition to the newer coherent terms. The 4th term contains double sums in (3) and (4). In the end only the 1st and 4th terms survive, and one may write, to a good approximation

$$\begin{aligned} C_2^{(\text{coherent})}(\mathbf{q}, \mathbf{q}, \Delta \phi) &= C_2^{(\text{incoherent})}(\mathbf{q}, \mathbf{q}, \Delta \phi) + \langle \int \sum_j A_j^*(\mathbf{q}, \phi) A_k(\mathbf{q}, \phi) \\ &A_j(\mathbf{q}, \phi + \Delta \phi) A_k^*(\mathbf{q}, \phi + \Delta \phi) x \dots \\ &\dots e^{i(\mathbf{q} - \mathbf{q}') \cdot (\mathbf{R}_k - \mathbf{R}_j)} d\phi \rangle_{\text{DP}} \quad (7) \end{aligned}$$

Where  $C_2^{(\text{incoherent})}$  contains only contribution from the first (incoherent) summations in (3) and (4) and the remaining term contains the specifically interparticle interference effects that arise from the coherence of the XFEL. Here  $\Delta \phi$  represents the vector in the direction of increasing  $\Delta \phi$ . For most values of  $\Delta \phi$ , the randomness of  $\mathbf{R}_j - \mathbf{R}_k$  makes the phases associated with the exponential random and thus the difference between the coherent and incoherent expressions become negligible. An exception occurs when  $\Delta \phi$  approaches zero. Then despite the randomness of  $\mathbf{R}_j - \mathbf{R}_k$  the phases of the exponential terms tend towards zero, which is clearly not random. It will also be noted that at this point  $A_j(\mathbf{q}, \phi + \Delta \phi) \otimes A_j(\mathbf{q}, \phi)$  and  $A_k(\mathbf{q}, \phi + \Delta \phi) \otimes A_k(\mathbf{q}, \phi)$ , and so the integrand of (7)  $\otimes I_j(\mathbf{q}, \phi) I_k(\mathbf{q}, \phi)$  which is always positive.

Consequently,  $C_2^{(\text{coherent})}$  is always greater than  $C_2^{(\text{incoherent})}$ . Thus, the coherent peak will always be larger than the incoherent peak, as seen by the red part of the curve and consequently at the equivalent point at  $\phi = 2\pi$ . For a flat Ewald sphere, Friedel's Law requires that the intensity at  $\Delta \phi = \pi$  be equivalent to that at  $\Delta \phi = 0$ , and hence Figure 1 (which assumes a flat Ewald sphere) has three highly localized red regions consisting of coherence peaks. This is, of course, for two randomly oriented particles. For more randomly oriented particles the height of these peaks will be higher while the widths will be the same (and inversely proportional to the width of the illuminated area via the uncertainty principle) (Figure 1).



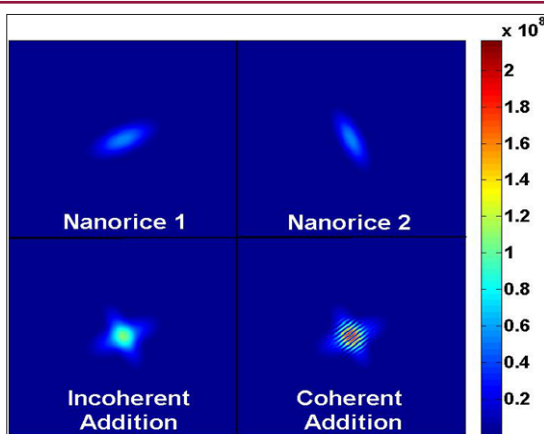
**Figure 1:** Coherent and incoherent  $C_2$  correlations for  $q'=q$  found from the diffraction patterns of randomly spaced particles by averaging all the diffraction patterns.

Note the difference between the correlations is only that the coherent patterns have an extra high but narrow peaks at  $\Delta\phi=0, \pi$  (for a flat Ewald sphere, due to Friedel's Law), and  $2\pi$  (same as 0). Alternatively, if one chooses to plot the correlations for  $|q-q'|$  greater than the width  $2\pi/L$ , where  $L$  is the lateral coherence length, the coherent and incoherent  $C_2$ 's will be equivalent throughout the entire range [15]. For a dilute ensemble of randomly oriented particles therefore, over most of the range of  $\Delta\phi$  our previous assumption that  $C_2$  be represented by the incoherent value remains true. Thus,

$$B_1(q, q) = \int C_2(q, q; \Delta\phi) P_1(\cos(\theta(q))\cos(\theta(q)) + \sin(\theta(q))\sin(q)\cos\Delta\phi) d\Delta\phi \quad (8)$$

is still valid to a good approximation for coherent radiation. For a flat Ewald sphere, one can take  $\theta(q) = \pi/2$  and  $\cos(\theta(q)) = 0$  and  $\sin(\theta(q)) = 1$ , so, to a good approximation,

$$B_1(q, q) = \int C_2(q, q; \Delta\phi) P_1(\cos\Delta\phi) d\Delta\phi \quad (9)$$



**Figure 2:** Diffraction patterns of two individual nanorice particles of random orientations (top panels), the diffraction pattern of two incoherently superimposed ones (bottom left panel) and of coherently superposed ones (bottom right panel).

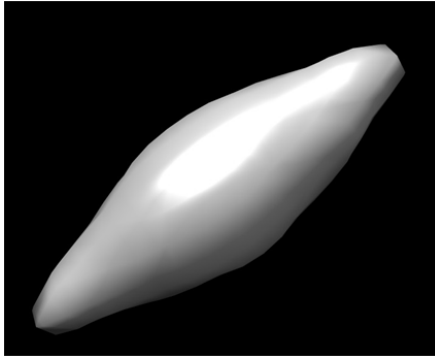
We next tried to simulate the conditions of a nanoparticles representing a small protein like photoactive yellow protein (PYP)

within a coherence volume that one would expect of a focal spot size of 0.1 microns (the design specification of the Linac Coherent Light Source (LCLS) at Stanford, California) and the smallest droplet size claimed of (0.3 microns). This gives an illuminated volume of 0.03 cu. microns. For single particles the resulting diffraction patterns are as shown on the first two panels of Figure 2, When there are two randomly oriented particles the diffraction pattern looks like the one in the 3rd column if one assumes no coherence between the two diffraction patterns. When one introduces coherence the diffraction patterns due to two particles appear like that in panel 4, with a set of narrow interference fringes characteristic of the spacing of the particles. Even in this case the values of  $C_2$  of the incoherent and coherent cases are almost identical over most of the range of  $\Delta\phi$  due to the randomness of the particle positions making the specifically coherent terms sum to a small quantity [15] (Figure 2).

Note that only the coherently added diffraction patterns have interference fringes. The effects of the interference fringes on the  $C_2$ 's almost disappears over most of the range of  $(q, q; \Delta\phi)$  since the fringes are randomly spaced and oriented. This gives rise to essentially the same  $C_2$ 's as in the incoherent case except for isolated regions around at  $\Delta\phi = 0, \pi$  and  $2\pi$  when the phases become zero and are therefore not random. Similar peaks appear in Kam's two-point triple correlations  $C_3(q, q; \Delta\phi)$  [10], i.e. they also have narrow peaks at  $\Delta\phi = 0, \pi$  and  $2\pi$ . On the other hand there are no coherent peaks associated with the correlations between different resolution rings such that  $|q-q'|$  is greater than  $2\pi/L$ , where  $L$  is the transverse coherence length [15].

In this case the integral over the coherent curve in Figure 1 gives almost the same answer as an integral over the incoherent curve in the same figure that has been considered before, in order to find  $B_1(q, q)$ . Similarly an integral over  $C_3(q, q; \Delta\phi)$  may be performed to find a good approximation to  $T_1(q, q)$ . Since both  $B$  and  $T$  are found from the same diffraction pattern (whether it comes from a single particle or several) it is possible to reconstruct an image of a single particle since both  $C_2$  and  $C_3$  are proportional to the number of particles. Since both  $C_2$  and  $C_3$  are derived from the same set of diffraction patterns the ratio of  $C_2$  to  $C_3$  will be the same

as for single particle, thus allowing exactly the same algorithm for any collection of diffraction patterns corresponding to an arbitrary number of particles. This allows one to reconstruct an image of a single particle from the angular correlations of many, even with the existence of interparticle interference. For incoherent scattering this has been demonstrated in 2D even with experimental data [6]. However, we think this is the first time this capability has been demonstrated in 3D (Figure 3).



**Figure 3:** Density of single particle of nanorice of size approximately 50 Å recovers from an illuminated area of approximately 1000 Å squares from 8000 two particles simulated diffraction patterns including noise and interparticle interference.

In cases where the extra coherent peaks give rise to significant changes in the values of  $B$ , and  $T$ , an integral that avoids these points gives almost the same value as the incoherent case. This is a bit like the principal part of an integral in complex analysis, except that the ordinate is not infinite in this case, the result of the integral over these regions remains very small. Consequently, the integral over the restricted range for the coherent case gives almost the same result as the integral over the whole range for the incoherent case [15]. It should be noted these deviations between the coherent and incoherent cases are confined to cases where  $q=q'$  (same resolution ring). Alternatively, one may avoid this problem if one retrieves information only from different resolution rings (off-diagonal  $C_2$ 's and  $C_3$ 's) (or  $q \neq q'$  (since the deviation between the coherent and incoherent cases only occur for the correlations with  $q=q'$  at these localized points). Even if one chooses to obtain information from the terms with  $q=q'$  one may avoid the sharp coherent peaks and obtain the quantity  $B_1(q,q')$  to a good approximation [15].

The  $B$ 's can be shown to be essentially

$$B_1(q,q') = N \sum_m I_m^*(q) I_m(q) \quad (10)$$

Where  $N$  is the number of particles scattered off. This quantity may contain vital information about the diffraction volume we seek.

Thus, even in the coherent case, one can define

$$C_2(q,q; \Delta \phi) = \int I(\phi) I(q, \phi + \Delta \phi) d\phi \quad (11)$$

Where  $I$  is the total intensity of the diffraction pattern, provided one avoids the narrow regions around  $\Delta \phi = 0$  (and their Friedel counterparts) that are characteristic of coherent radiation. At least in 2D it is clear [10] that this is a quantity independent of the orientation of the particle and hence diffraction pattern. (It is in

3D also [7,8]). The quantity depends only on the difference,  $\Delta \phi$ , the azimuthal angle between two points involved in the calculation of the correlations. Effectively one is comparing the scattering of a single particle in various relative directions, a quantity that is independent of the particle orientation. The fact that this is independent of particle orientation makes it suitable for use with an ensemble of identical particles differing only in orientation. The total signal will be proportional to the number of particles, thus we gain the advantage of a crystal (or at least of a microcrystal) with a signal contributed by multiple particles. Since there are some questions about the feasibility of the signal-to-noise ratio with a large number of particles, the initial tests will be on a small number of particles.

This will be possible since an XFEL will be expected to be initially focused on an area of radius 100 nm, perhaps 20 times larger in linear dimensions than a small protein. The Legendre transform of  $C_2$ , see Eq. (9), gives rise to a quantity that can be shown to be characterized by the spherical harmonic expansion of the scattered intensities. That is, it can be shown that if the coefficients  $I_{lm}(q)$  may be extracted from these quantities, one may reconstruct a 3D diffraction pattern of the scattered intensities via the expansion

$$I(q) = \sum_{lm} I_{lm}(q) Y_{lm}(\theta, \phi) \quad (12)$$

Where  $Y_{lm}(\theta, \phi)$  is a spherical harmonic of angular momentum quantum numbers  $(lm)$ .  $\theta$  and  $\phi$  are 3D polar angles. A problem is that, in general, the extraction of the  $I_{lm}$  coefficients from the  $B$ 's is not a trivial task, although there are specific high-symmetry circumstances when this has been shown to be possible [11,12]. One of these is when there is azimuthal symmetry in the particle for example a grain of nanorice. In this case the azimuthal symmetry of the particle results in an azimuthal symmetry of the diffraction volume [15] If the diffraction volume is azimuthally symmetric, one may take  $m=0$  in (10). This immediately suggests

$$|I_{l0}(q)| = \sqrt{B_1(q,q)} \quad (13)$$

This gives the magnitudes of all allowed values of the expansion coefficients. The only remaining task is to find their phases. They may be found from the diagonal triple correlations defined by

$$C_3(q,q; \Delta \phi) = \langle \int I_2(q, \phi) I(q, \phi + \Delta \phi) d\phi \rangle_{DP} \quad (14)$$

Which may be evaluated from the ensemble of diffraction patterns just as easily as the pair correlations (we don't need to be concerned about the coherent peak as it is so narrow that we can ignore it His still allows us to capture the incoherent peak Alternatively one can devise an algorithm that uses only  $C_2(q,q; \Delta \phi)$  with  $|q'-q| > 2\pi/L$  where  $L$  is the width of the illuminated area. It may be shown that over most of the range of  $\Delta \phi$  even for coherent radiation (once again it would be best to avoid the red high-intensity peaks in Figure 1 at  $\Delta \phi = 0, \pi$ , and  $2\pi$  if  $q=q'$  (correlations taken over the same resolution ring) [16].

### Reconstruction of a Single Particle from Multiple Particle Diffraction Patterns

The importance of scattering by multiple particles is because there is still a considerable gap between the focal spot size aimed



for by the XFEL instrument manufacturers and the sizes of typical proteins. Suppose one uses a concentration of 0.6 moles/ m<sup>3</sup> as in SAXS. The design specification of the world's first X-ray free electron laser (XFEL), the Linac Coherent Light Source (LCLS), is for a focal spot size is about 0.1 micron square, and since the minimum size of a liquid droplet claimed is about 0.3 microns [17], the volume illuminated will be about 3x10<sup>-21</sup>m<sup>3</sup>, or else about 1.8x10<sup>-21</sup> moles. But Avogadro's number is about 6x10<sup>23</sup> molecules/mole. Thus the minimum number of molecules illuminated will be about would be about 400.

Thus we have to conclude that for the study of small proteins such as PYP, single molecule studies are still some way in the future, and that, in the meanwhile, we need either nanocrystals or a number of independently randomly oriented molecules, and a theory of type presented here if clustering of the molecules can be avoided. If the molecules cluster, it's probably better to use a "hit-finder" to eliminate multiple particle hits beforehand and then to use any of the theories proposed (including this one) for structure solution

Although this has been claimed before [7,8], we have now actually tested this proposition in real simulations, not stopping at showing that the correlations are the same (apart from a scaling factor), but reconstructing an image from two particles per shot from the correlations. We have shown that a particle may be reconstructed from both B<sub>l</sub>(q,q) and T<sub>l</sub>(q,q)'s from a single particle if m=0. We show here that exactly the same method may be used for multiple particles in independent random orientations in each snapshot diffraction pattern since the scaling factor of both B and T are the same for a given number of particles. This will always be the case: the diffraction patterns from which B and T are evaluated will come from exactly the same numbers of particles since in general they will be from the same diffraction patterns. For purposes of illustration here we considered only two particles in independent random orientations, but this illustrates the main point. We simulated

$$I'_m(q) = \sum_m [D_{lm'm}(\alpha_1, \beta_1, \gamma_1) + D_{lm'm}(\alpha_2, \beta_2, \gamma_2)] I_m(q) \quad (15)$$

Where  $(\alpha_1, \beta_1, \gamma_1)$  and  $(\alpha_2, \beta_2, \gamma_2)$  are independent sets of random Euler angles. We did explicit simulations of the expected diffraction patterns and calculated from them the angular correlations B and T and reconstructed from them, by the same algorithm for the single-particle case, a real space image. The image we found is shown in Figure 3. If indeed the conditions of a dilute ensemble of identical particles can be achieved, there will be no need for a hit-finder program [18] to reject multiple-particle hits provided the particles are truly random in position (dilute ensemble). The fact that the particle-number-dependent scaling factors are the same for the pair and two-point triple correlations means it is not necessary to know at the outset exactly how many particles there are in the ensemble since both the pair correlation and two-point triple correlations are derived from diffraction patterns with exactly the same number of particles.

The possibility of using diffraction patterns possibly from multiple particles adds considerably to the capability of the use for

the XFEL for structure determination of individual particles as it will add greatly to the "hit-rate". Our simulations suggest that here may be even advantages to considering diffraction patterns from ensembles of multiples in terms of convergence. In light of this the only case we see for hit-finder program at least one that determines single particle hits is when the particles tend to stick together and agglomerate. In such a case it might be better to determine the structure from single particles. It should be pointed out that despite the fact that it is capable of dealing with cases of multiple randomly oriented particles, it is also a method of determining the structure of a single particle from diffraction patterns of a single particle in random orientations [18].

The values the magnitudes of the spherical harmonic expansion coefficients are determined from (13). Also, as B<sub>l</sub>(q,q) is determined by the integral of a real quantity C<sub>2</sub> with a real Legendre polynomial, it is real and has a positive square root. The only remaining task is to determine the sign of I<sub>l0</sub>(q). We can determine this sign from the triple correlations by assuming, as it must be for nanocrystals, that azimuthal symmetry of the amplitudes implies azimuthal symmetry of the intensities from the usual Clebsch-Gordon rules for adding angular momenta. In this case, all magnetic quantum numbers are equal to zero, and the triple correlation reduces to a sum over only the angular momentum quantum numbers l, and not also magnetic quantum numbers m. Thus,

$$T_l(q,q) = N \sum_{\lambda, \kappa} I_{l\lambda 0}(q) I_{l\kappa 0}(q) G(l0; \lambda 0; \kappa 0) \quad (16)$$

Where G is a Gaunt coefficient [19] Note that the two-point triple correlations are scaled for multiple particles by exactly the same factor N as the quantities B<sub>l</sub>.

Since an ellipsoid has azimuthal symmetry about a particular axis, we can choose that particular axis as z-axis, thus eliminating any other components of the magnetic quantum number except m=0. |I<sub>l0</sub>(q)| can be obtained directly from B<sub>l</sub>(q,q) via (13). The only unknown here is sign of I<sub>l0</sub>(q). The sign can be determined by fitting all possible signs of I<sub>l0</sub>(q) to the values T<sub>l</sub>(q) of the triple correlations calculated directly from the diffraction patterns of random particle orientations in (16). It should be stressed that the number of equations is equal to the number of distinct T<sub>l</sub>(q,q) values, namely the numbers of q and l values, as is the number of unknowns I<sub>l0</sub>(q), so there is no information deficit [20] (Figure 3).

After obtaining the signs of the I<sub>l0</sub>(q), the single particle diffraction volume can be calculated from

$$I(q) = \sum_l I_{l0}(q) Y_{l0}(q) \quad (17)$$

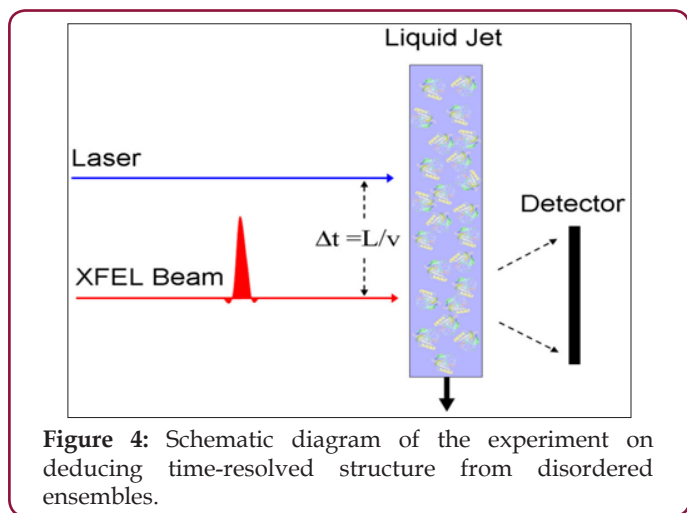
And an iterative phasing algorithm produces the real space image of Figure 3.

It should be pointed out that, for a single particle, the structure of such a simple particle had been determined even experimentally by a different method [21], which is a true single particle method and does not have the ability to reconstruct a single particle from diffraction patterns of multiple ones. In fact, the ability to recover an image of a single particle from the diffraction patterns of many is unique to our method. It should also be pointed out that Figure

3 was obtained from diffraction pattern intensities simulated under the coherence assumption since the difference between the coherent and incoherent intensities are highly localized at  $\Delta \phi = 0, \pi$  and  $2\pi$  in the angular correlations.

**Experimental Determination of the Time-Resolved Structure of Proteins**

Armed with these facts about angular correlations, and their remarkable ability to study structure from disordered ensembles, we return now to the main problem addressed in this paper, namely the recovery of time-resolved information about proteins from XFEL diffraction patterns. Paradoxically, it is easier to extract useful information from more complex particles in a time-resolved experiment of the form of Figure 3, although the advantages suggested above of not needing to focus on a single particle remain. In this arrangement the particles are incident on an XFEL in a liquid jet. A short time before the incidence of the X-ray beam on the liquid jet it is illuminated by light from an optical laser which puts suitable molecules into a light-induced excited state. A schematic diagram of the apparatus is shown in Figure 4. Although this experiment describes excitation by light, other similar experiments may be considered in which one studies the set of molecules immediately after they has been mixed with a substrate, for example. As we have already seen.



**Figure 4:** Schematic diagram of the experiment on deducing time-resolved structure from disordered ensembles.

A quantity that is independent of the particle’s orientation (although dependent of the particle structure) is those related, example to the angular pair correlations,  $C_2$ . This will undoubtedly change  $B_l(q,q')$  by a quantity  $\delta B(q,q')$ , which in turn may be measured by measuring the values of  $B$ . before and after the reaction. Interestingly, it is possible to find a linear relation between  $\delta B_l(q,q')$  and the change in the electron density  $\delta \rho(r)$  of the molecule. Since in general there will be a time delay between the optical laser excitation of the molecule and its interrogation by the X-ray beam, if the structure after the excitation may be deduced from the measured X-ray diffraction patterns, one would get a handle into the structure a specified time after the excitation. Since this time interval may be varied by just choosing to illuminate the particle beam a different distance from the incidence of the X-ray beam, one is able to follow the time-resolved changes in the

structure of the photo excited molecule that is the course of light-induced chemical reaction.

The central experimental quality one uses is  $B_l(q,q')$  which we have seen is the Legendre transform of  $C_2(q,q';\Delta \phi)$  which, being a pair correlation, is characteristic of the particle structure, but not of its orientation [7,8]. We ultimately want to find the change  $\delta p(r)$  in the electron density of each molecule (assumed identical, as one must expect of molecules) corresponding to an  $\delta B_l(q,q')$  which in turn is derived from the experimentally measured  $C_2(q,q';\Delta \phi)$ . It can be shown that the quantity  $B_l(q,q')$  can be rewritten in the alternative form

$$B_l(q,q') = N \sum_m I_m^*(q) I_m(q') \tag{18}$$

Differentiating which, we obtain

$$\delta B_l(q,q') = N \sum_m \delta I_m(q) I_m(q') + I_m(q) \delta I_m(q') \tag{19}$$

In terms of the spherical harmonic expansion coefficients  $I_m(q)$  of the diffraction volume of a single particle. Also note that

$$\delta I_m(q) = \int \delta I(q) Y_m(q) dq. \tag{20}$$

In order to relate this to the change  $\delta \rho$  in the electron density one should convert small changes in the intensity to small changes  $\delta A$  in the amplitude, which one can do by differentiating. Since

$$I(q) = |A(q)|^2 \tag{21}$$

It follows taking the variation again that

$$\delta I(q) = \delta A(q) A^*(q) + A^*(q) \delta A(q) \tag{22}$$

At first sight it is a bit disappointing that both  $\delta A$  and its complex conjugate appear on the RHS. However as we will see, this is of little consequence since the quantity we are ultimately after, namely the change in the electron density of the molecule,  $\delta \rho(r)$ , is real. Since

$$\delta A(q) = \int \delta \rho(r) e^{i q \cdot r} dr \tag{23}$$

And the complex conjugate of this equation is

$$\delta A^*(q) = \int \delta \rho(r) e^{-i q \cdot r} dr \tag{24}$$

It’s possible to relate both to the change  $\delta \rho(r)$  in the real electron density. In summary, it is possible to find a linear real relation

$$\delta B_l(q,q') = N M_{lqq'} \delta \rho(r) \tag{25}$$

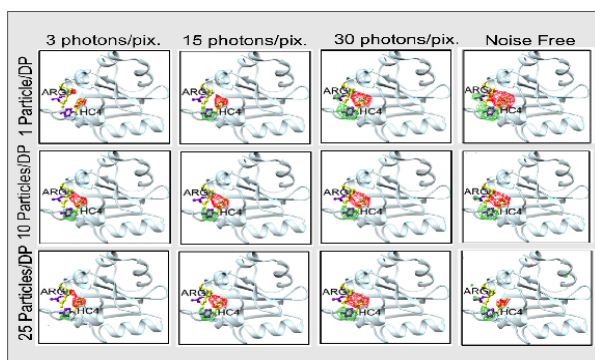
$$\text{Where } M_{lqq'} = \sum_m I_m^*(q) C_l(q',r_j) + C_l^*(q,r_j) I_m(q') \tag{26}$$

$$\text{Where } C_{lm}(q,r_j) = e^{i q \cdot r_j} 4 \pi \sum_k f_k i^l j^l(q|r_k - r_j|) Y_{lm}^*(|r_k - r_j|) \tag{27}$$

Where  $f_k$  is the form factor of the  $k$ -th atom in the molecule Eq. (24) is the relationship we sought [22,23]. Note that since  $B_l$  is a quantity independent of the number of identical particles, all our arguments for the scattering is not off one but an ensemble of particles of random orientations still hold. This is of advantage experimentally as we have pointed out before. This relationship has many similarities to the *difference Fourier* formula used in conventional time-resolved crystallography [3,24]. They both require as input the “dark” structure, i.e. the structure of the

unexcited molecule. They both recover the difference density  $\delta \rho$  in the frame of reference of the “dark” structure. In both cases, the latter factor allows the difference electron density to be displayed ultimately using standard crystallographic software superimposed on a model of the “dark” structure that the changes in the structure on photoexcitation are more obvious. It should be noted that although individual particles are in random orientations an initial reconstruction of the angular correlations allows one to reconstruct the electron density in an orientation of one choosing (since all orientations correspond to the same angular correlations). We choose the frame of reference of the “dark” structure, by a suitable definition of the matrix  $M$ , because it allows the display of the electron density changes using exactly the same software as time-resolved work with crystals.

One thing we should point out is that like the difference Fourier method, this is a non-iterative way of getting at the difference electron density directly from the measured data. It seems to be accurate enough to determine structural changes at the level of the residues of a protein. This is seen in Figure 5 [22] in which is shown the expected difference electron density calculated by our algorithm of the expected difference structure of photoactive yellow protein (PYP) 2 ms after photoexcitation. This figure suggests that the reconstruction survives Poisson noise which is included in the simulations. It will be noted that after 2ms two residues are expected to move simultaneously, the chromophore as well as the nearby ARG residue which moves in response. Some evidence of the movement of both is just seen in the reconstructions, although the larger movement of the chromophore is more evident. As is customary in time-resolved work, the red regions indicate a deficiency of electron density in the final image and a green lobe indicates an increase in the electron density. Thus looking at these electron density maps gives an indication of the structural changes a short time measured time after photoexcitation.



**Figure 5:** Difference electron density expected 2ms after photoexcitation. These calculations offer a realistic simulation as they assume that the diffraction patterns include Poisson noise. We consider three different numbers of particles as well as three different incident photon counts.

In keeping with our featured capability of the correlation method it will be noted that the right hand column features as many as 25 randomly oriented particles. These simulations were performed assuming an incoherent source of X-rays as the intensity

from many particles is assumed to be the sum of intensities from single particles. The distinction with the coherent case is not very important over most of the range of the correlations as the randomness of particle positions give the scattering a kind of incoherence [9]. The only region where the coherence plays a part in when  $q=q'$  and  $\delta \phi$  is approximately zero,  $\pi$  (due to Friedel symmetry) or  $2\pi$  (same as zero). We have suggested how to deal with these differences [15] (Figure 5)

## Conclusion

The significance of this work is that if these simulations are realized in practice, we now have, for the first time, a method of finding the structures of particles such as molecules from diffraction patterns of copies of many of the molecules even if they are not of identical orientations as in a crystal. In the last application we have shown that even time-resolved structure may be determined of molecules that are in random orientations, thus opening to experimentally observing realistic chemical reactions. It should be pointed out that many of the effects we have ignored, such as scattering by solvent, are fairly unimportant for the time-resolved problem since here we look at differences in the correlations between the photo excited and ground state molecules, where such factors tend to subtract out. Indeed in the time-resolved problem we study only features of the diffraction patterns that are different due to the photoexcitation. Of course solvent is perhaps 100 times as prevalent as solute and could give rise to Poisson noise that does not cancel between the photo excited and ground state structures so some care has to be exercised.

An experiment has recently been reported in which the structure of the Mimi virus has been determined experimentally by a variant of the single particle methods described earlier [23]. It has been previously shown by us [10,11] that under various high-symmetry situations, that it is possible to determine the structure of the particle from simulated diffraction patterns. In an experiment to recover time-resolved structural variations, starting from a knowledge of a nearby structure, using many of the same quantities, namely, the angular correlations, have also been shown to be capable of recovering time-resolved changes in a structure from the knowledge of a closely related structure of a single molecule in realistic simulations, including shot noise [24] even from diffraction patterns of multiple particles, as we demonstrate here. This possibility is unprecedented in structural work with XFEL diffraction patterns. This method also works on diffraction patterns of multiple particles in independent random orientations.

## Acknowledgment

We acknowledge support for this work from the National Science Foundation, grant numbers MCB-1158138 and STC-1231306, we thank Kanupriya Pande, Ruslan Kurta, Peter Zwart and Jeffery Donatelli, for helpful discussions, and Kanupriya Pande for comments on the manuscript.

## References

1. HA Krebs, PDJ Weizman (1987) The Krebs or citric acid cycle: half a century and still turning London: Biochemistry Society p. 24.

2. A Stolow, AE Bragg, DM Neumark (2004) Femosecond Time-Resolved Photoelectron Spectroscopy. *Chem Rev* 104(4): 1719-1757.
3. K Moffat, D Szebenyi, D Bilderback (1984) X-ray Laue diffraction from protein crystals. *Science* 222(4645): 1423-1425.
4. TW Kim, JH Lee, J Choi, KH Kim, LJ van Wilderen, et al. (2012) Protein structural dynamics of photoactive yellow protein in solution revealed by pump-probe X-ray solution scattering. *J Am Chem Soc* 134(6): 3145-3153.
5. DK Saldin, VL Shneerson, M Howells, S Marchesini, HN Chapman, et al. (2010) Structure of a single particle from scattering by many particles randomly oriented about an axis: towards structure solution without crystallization. *New J Physics* 12: 035014.
6. DK Saldin, HC Poon, M Bogan, S Marchesini, D Shapiro, et al. (2011) new light on disordered ensembles: ab initio structure determination of one particle from scattering fluctuations of many copies. *Phys Rev Lett* 106(11): 115501.
7. Z Kam (1978) Determination of macromolecule structure in solution by spatial correlations. *Macromolecules* 10: 927.
8. DK Saldin, VL Shneerson, R Fung, A Ourmazd (2009) Structure of isolated biomolecules from ultra-short x-ray pulses: exploiting the symmetry of random orientation. *J Phys Condens Matter* 21(13): 134014.
9. DK Saldin, HC Poon, VL Shneerson, M Howells, HN Chapman, et al. (2010) beyond small angle X-ray scattering: exploiting angular correlations. *Phys Rev B* 81: 174105.
10. HC Poon, DK Saldin (2011) beyond the crystallization paradigm: structure determination from diffraction patterns of ensembles of randomly oriented particles. *Ultramicroscopy* 111(7): 788-806.
11. DK Saldin, HC Poon, P Schwander, M Uddin, M Schmidt (2011) Reconstructing an icosahedral virus from single particle diffraction experiments. *Opt Exp* 19(18): 17318-17335.
12. HC Poon, P Schwander, M Uddin, DK Saldin (2013) Fiber diffraction without fibers. *Phys Rev Lett* 110(26): 265505.
13. R Kirian, KE Schmidt, X Wang, RB Doak, JCH Spence (2011) Signal, noise, and resolution in correlated fluctuations from snapshot small angle x-ray scattering. *Phys Rev* 84(1): 011921.
14. DK Saldin, HC Poon, VL Shneerson, M Howells, HN Chapman, et al (2010) beyond small angle X-ray scattering: exploiting angular correlations. *Phys Rev B* 81: 174105.
15. RP Kurta, M Altarelli, E Weckert, IA Vartnyants (2012) X-ray cross correlation analysis applied to disordered two dimensional systems. *Phys Rev B* 85: 184204.
16. We have earlier considered another case where the diffraction volume is dominated by  $m=0$ . This is the case of a helical structure, which is also characterized by intensities only on a set of "layer planes". In the present case the azimuthally symmetric diffraction volume is not characterized by "layer planes". The fact that the intensities appear throughout reciprocal space provides the distinction.
17. DP Deonte, JT McKeown, U Weirstall, RB Doak, JCH Spence (2011) Towards ETEM serial crystallography: electron diffraction from liquid jets. *Ultramicrosc* 111(7): 824-827.
18. JB Pendry (1974) *it Low Energy Electron Diffraction*. Academic, London, UK.
19. V Elser (2011) Strategies for processing diffraction patterns from randomly oriented particles. *Ultramicrosc* p. 111.
20. ND Loh, MJ Bogan, V Elser, A Barty, S Boutet, et al. (2010) Cryptotomography: reconstructing 3D Fourier intensities from randomly oriented single-shot diffraction patterns. *Phys Rev Lett* 104: 225501.
21. K Pande, P Schwander, M Schmidt, DK Saldin (2014) Deducing fast electron density changes in randomly oriented uncrystallized biomolecules in a pump-probe experiment. *Phil Trans Roy Soc B* 369(1647): 20130332.
22. K Pande, M Schmidt, P Schwander, DK Saldin (2015) Simulations on time-resolved structure determination of uncrystallized biomolecules in the presence of shot noise. *Struct Dyn* 2(2): 024103.
23. J Drenth (2007) *Principles of Protein X-Ray Crytalslography* Berlin: Springer.
24. T Ekeberg, M Svenda, C Abergel, FR NC Maia, V Seltzer JM Claverie, et al. (2015) The dimensional reconstruction of the giant mimivirus particle with an x-ray free electron laser. *Phys Rev Lett* 114: 098102.



#### Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<http://biomedres.us/>