Research

Deducing fast electron density changes in randomly orientated uncrystallized biomolecules in a pump–probe experiment

K. Pande, P. Schwander, M. Schmidt and D. K. Saldin

Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

We propose a method for deducing time-resolved structural changes in uncrystallized biomolecules in solution. The method relies on measuring the angular correlations of the intensities, when averaged over a large number of diffraction patterns from randomly oriented biomolecules in solution in a liquid solvent. The experiment is somewhat like a pump–probe version of an experiment on small angle X-ray scattering, except that the data expected by the algorithm are not just the radial variation of the averaged intensities. The differences of these correlation functions as measured from a photoexcited and dark structure enable the direct calculation of the difference electron density with a knowledge of only the dark structure. We exploit a linear relation we derive between the difference in these correlation functions and the difference electron density, applicable for small structural changes.

1. Introduction

The method of time-resolved crystallography [1] has given us insights into rapid variations of the structure of biomolecules via pump–probe experiments involving photoexcitation by a laser followed very quickly by measurement of X-ray diffraction patterns. Changes in the structure are then found by established methods [2,3]. We present here computer simulations that suggest that such structural changes may be determined even from an ensemble of biomolecules in solution.

The X-ray free-electron laser (XFEL) is a new instrument that promises to revolutionize our study of the atomic architecture of matter [4,5]. The brightness of the X-rays produced by this instrument is some 10 billion times greater than any existing X-ray source (including present-day synchrotrons). This allows the possibility of measuring signals from scattered X-rays of large single molecules, such as proteins. The traditional limitation of X-ray flux for fragile biomolecules can be circumvented completely due to the fact that this very bright radiation is delivered in ultra-short pulses [6–8]. Although the molecules under study will undoubtedly suffer catastrophic radiation damage, the shortness of the pulse enables a signal (albeit weak) to be measured from the particle before its disintegration [4]. In common with other theories developed for extracting structural information from such molecules, we will first assume [9] that a single molecule is hit by the X-ray beam at a time, with random orientations each time. The fact that the particles are hit by the incident X-ray pulses in random orientations enables the full three-dimensional structure of the scattering object to be found from the ensemble of diffraction patterns without deliberate rotations of the object with a goniometer, as is usually done in protein crystallography. However, despite the huge increase in the brightness of an XFEL beam over all previous X-ray sources (a factor of about 10^{10}), the reduction in size from even a 0.1 mm diameter crystal commonly used in X-ray crystallography, to a single molecule (typically of diameter 10^{-8} m) is

even greater (typically a factor of approx. 10^{-12}). Consequently, very few X-ray photons are scattered by a single molecule. Nevertheless, the other feature of XFEL radiation, namely that it is delivered in pulses of a few femtoseconds duration allow a recording of a large number of diffraction patterns per second, thus allowing the measurement of even millions of diffraction patterns from random particle orientations over the course of a single experiment. The search has therefore been on for the development of algorithms that can overcome the low signal-to-noise ratio of any single diffraction pattern by judiciously combining information from a large number of such diffraction patterns. Such an algorithm would allow structure determination of reproducible biomolecules by essentially an unlimited X-ray flux (because the scattering is expected to take place before the molecule's disintegration). This also would be, in one sense, a complete solution of the radiation damage problem. We suggest in this paper that it should also be possible to combine the ultra-brightness [10] of the radiation with the ultra-shortness of its duration [11] to enable the gathering of information never before possible, for example the changes in the structure of an uncrystallized biomolecule in solution as a result of some stimulus, such as photoexcitation, as a function of time since the photoexcitation. Because this time can be very short, the possibility then exists of experimentally following the course of rapid chemical reactions of such uncrystallized biomolecules, as is already done with crystallized ones [11] by the technique of time-resolved diffraction [1,12,13]. This may allow for the first time the study of biochemical reactions of molecules in aqueous solution in which they occur in nature.

An idea proposed for sample delivery of hydrated molecules to an XFEL beam is to inject a continuous stream of a solution containing the molecules into the sample chamber [14–16]. The incident X-rays then scatter off the protein solution. Initially, we assume a single particle per diffraction pattern, as do all other theories for the extraction of structural information from XFEL patterns, but in §6, we point out how this restriction may be relaxed.

2. Structural information from disordered ensembles of molecules

Due to the fact that the particles are presented to the XFEL beam in random orientations, the individual *intensities* will not be obviously related to intensities on other diffraction patterns. However, it has been pointed out [9,17] that the average angular correlations over a large number of such diffraction patterns can yield a quantity that is directly related to the three-dimensional diffraction volume of a single particle. If an oversampled version of the three-dimensional diffraction volume can be extracted from this ensemble of diffraction patterns, modern iterative phasing algorithms [18,19] would be expected to be able to reconstruct the three-dimensional electron density of an individual particle from the multiple diffraction patterns expected to be measured in such an experiment.

We begin by defining an average over a set of diffraction patterns of angular pair correlation function between the intensities of a couple of resolution rings q and q' of diffraction patterns by

$$J(q, q', \Delta\phi) = \langle I(q, \phi) I(q', \phi + \Delta\phi)\rangle_{DP}, \qquad (2.1)$$
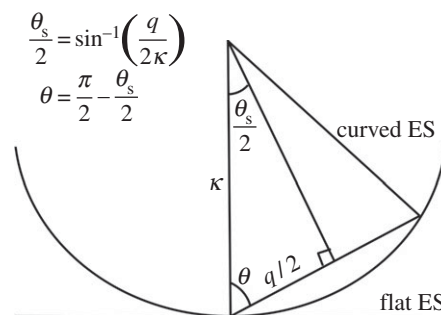

Figure 1. Schematic of diffraction geometry demonstrating the relationship between polar angle $\theta(q)$, scattering angle $\theta_s(q)$, magnitude of scattering vector q and wavenumber $\kappa = 2\pi/\lambda$. ES, Ewald sphere.

where $\langle \cdots \rangle_{DP}$ represents an average over all the diffraction patterns, DP. Note that the orientational averaging of the particles implied by the reasonable assumption that all molecular orientations are equally likely suggests that the LHS of the equation (2.1) will be independent of the value of ϕ chosen on the RHS. In this case, it is possible to show [9] that

$$J(q, q', \Delta\phi) = \sum_l B_l(q, q') P_l(\sin[\theta(q)] \sin[\theta(q')]$$
$$+ \cos[\theta(q)] \cos[\theta(q')] \cos[\Delta\phi]), \qquad (2.2)$$

where

$$\theta(q) = \frac{\pi}{2} - \sin^{-1}\left(\frac{q}{2\kappa}\right), \qquad (2.3)$$

and κ is the wavenumber of the incident radiation. The angle $\theta(q)$ takes account of the curvature of the Ewald sphere, and is almost equal to $\pi/2$ radians for a small resolution ring or a flat Ewald sphere (figure 1). However, this formalism also allows one to take account of the curvature of the Ewald sphere very naturally. In this expression, also

$$B_l(q, q') = \sum_m I_{lm}(q) I_{lm}^*(q'), \qquad (2.4)$$

where $I_{lm}(q)$ are a set of spherical harmonic expansion coefficients of the three-dimensional diffraction volume of a single particle, and P_l is a Legendre function. Finding the coefficients $I_{lm}(q)$ enables the reconstruction of an *oversampled* three-dimensional distribution of scattered *intensities*. Recent developments of iterative phasing algorithms [18,19] suggest that this is all that is needed to reconstruct the electron density of the particle giving rise to the scattered intensity distribution.

Although the quantities $B_l(q, q')$ may be found quite straightforwardly by inverting equation (2.2), finding the $I_{lm}(q)$ coefficients in general from these quantities is far from easy [20], although where the particle has a known symmetry, for example, in the case of an icosahedral [21] or helical virus [22] this may be possible by exploitation of the known symmetry.

Another circumstance in which structural solution is possible is when what is sought is a small change in the structure from a known one. There are no symmetry restrictions to such a difference structure, only that it be close to the known dark structure (the same assumption as made in the difference Fourier method of time-resolved crystallography). Such a solution is of great significance in that it suggests the possibility of following the fast time variation of the structure of a photoexcited biomolecule of uncrystallized biomolecules in solution under conditions which somewhat mimic their functional states in living organisms. Any steric

hindrance caused by neighbouring molecules in a tightly packed crystalline state is circumvented by not having to crystallize the molecules at all, which may also have experimental advantages. Such an experiment truly exploits not only the extreme brightness, but also the fast time structure of XFEL pulses.

We demonstrate this in the case of photoactive yellow protein (PYP), where earlier time-resolved studies have established that, 2 ms after photoexcitation, the primary change in the structure may be regarded as a *cis−trans* isomerization of its chromophore about its C2−C3 axis, and a change of the χ_3 torsional angle of the side chain of the ARG 52 residue, to make room for the structural changes owing to the chromophore isomerization.

3. Proposed time-resolved experiment on dissolved molecules in random orientations

Imagine an experiment where a continuous stream of a solution of photoexcitable molecules is injected into an XFEL sample chamber in the usual manner of the so-called diffract and destroy experiments [4], but where a short distance L before the intersection of the X-rays from the XFEL, the molecules are photoexcited by a powerful laser. If v is the speed of the solvent stream (typically 10 m s^{-1}), then the molecules will be illuminated by the laser a time $\Delta t = L/v$ before it is interrogated by the X-ray beam. Because the time Δt is controllable by varying the distance L, this pump−probe experiment allows exquisite control of the time delay after photoexcitation. It is envisaged that a large number of diffraction patterns may be measured from different regions of the continuous solvent for a given time delay. Diffraction effects owing to solvent scattering may be removed by a Babinet principle argument also used in the theory of small angle X-ray scattering (SAXS) [23]. It is expected that structural information of the photoexcited molecules in the solution resides in the ensemble of such difference patterns [24] (figure 2).

4. Extraction of time-resolved structural changes from randomly oriented molecules

In our earlier paper [24], we demonstrated that the structural content of the $B_l(q, q')$ quantities is extractable from a set of XFEL diffraction patterns of multiple randomly oriented biomolecules. For the case of an assumed structural change of a PYP molecule about 2 ms after photoexcitation which consists of a *cis−trans* isomerization of a chromophore and a simultaneous torsional angle rotation of the side chain of a nearby arginine residue, we showed that the presumed photoexcited structure can be found by optimizing the $B_l(q, q')$ quantity of the photoexcited structure, starting at the value for the known dark structure and assuming that its change on photoexcitation may be characterized entirely by the variation of two torsional angles. Because the number of parameters varied in this case was just two, our simulations suggested the possibility of finding the correct photoexcited structure by performing an exhaustive search through the two-dimensional parameter space to minimize a reliability (or R-) factor between a simulation of the expected 2 ms structure and simulations of model structures for valid values of the structural parameters. Of course for a larger possible number
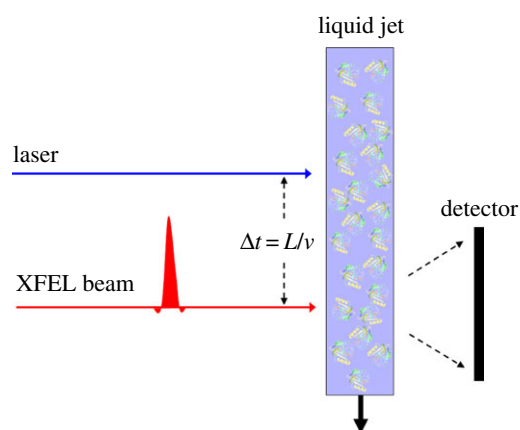

Figure 2. Schematic diagram of the proposed pump−probe experiment. A continuous stream of a solution of identical molecules in random orientations is introduced to the path of the X-ray beam after photoexcitation by a laser a controllable time before interrogation by the X-rays. The time delay Δt between photoexcitation and X-ray incidence is controlled by varying the distance L between the positions of photoexcitation and X-ray incidence.

of varied structural parameters, a more sophisticated search algorithm, such as simulated annealing [25], would need to be used. However, if absolutely nothing was known in advance of the photoexcited structural change, then such a simulated annealing method may not be practical, as the number of structural parameters to be varied may exceed the practical capability of such an algorithm [26].

In such a circumstance, an algorithm akin to the difference Fourier method of traditional time-resolved crystallography, which reveals the difference electron density of the entire molecule, from the experimental data alone, and a knowledge of only the dark structure would be desirable. We show below how such a difference electron density may be recovered from experimentally measurable quantities even from uncrystallized biomolecules in solution, with a knowledge only of the dark structure, just as in traditional time-resolved crystallography.

The idea is that, for small structural changes, one may take the variation of both sides of (2.4) to give

$$\delta B_l(q, q') = \sum_m \delta I_{lm}(q) I_{lm}^*(q') + I_{lm}(q) \delta I_{lm}^*(q'). \quad (4.1)$$

On the RHS, we interpret $I_{lm}(q)$ as the magnitude of the spherical harmonic expansion coefficient of resolution shell q of the (usually known) dark structure and $\delta I_{lm}(q)$ the corresponding quantity associated with the difference in the scattered intensity of the photoexcited and dark structures of a molecule. That is, one may write

$$\delta I_{lm}(q) = \int \delta I(\mathbf{q}) Y_{lm}^*(\theta, \phi) \mathrm{d}\Omega_q, \quad (4.2)$$

where $\mathbf{q} \equiv (q, \theta, \phi)$, $\mathrm{d}\Omega_q$ represents integration over the angles θ and ϕ, and

$$\delta I(\mathbf{q}) = \delta A^*(\mathbf{q}) A(\mathbf{q}) + A^*(\mathbf{q}) \delta A(\mathbf{q}), \quad (4.3)$$

where $A(\mathbf{q})$ may be regarded as the scattered amplitude from the dark structure and $\delta A(\mathbf{q})$ that from the difference structure. Consequently, we can write

$$A(\mathbf{q}) = \sum_k f_k(q) \exp(i\mathbf{q} \cdot \mathbf{r}_k), \quad (4.4)$$

where f_k is the form factor of an atom at position $\mathbf{r}_k$, in the dark (unphotoexcited) molecule, and

$$\delta A(\mathbf{q}) = \sum_p \delta\rho_p \exp(i\mathbf{q}\cdot\mathbf{r}_p), \tag{4.5}$$

and $\delta\rho_p$ is the difference electron density at voxel p at position $\mathbf{r}_p$. Substituting (4.4) and (4.5) into (4.3), we can write

$$\delta I(\mathbf{q}) = \sum_{k,p} \delta\rho_p[f_k(q)\exp[i\mathbf{q}\cdot(\mathbf{r}_k - \mathbf{r}_p)] + \text{c.c.}]. \tag{4.6}$$

from which we deduce that

$$\delta I_{lm}(q) = \sum_p C_{lm}(q;\mathbf{r}_p)\delta\rho_p, \tag{4.7}$$

where

$$C_{lm}(q;\mathbf{r}_p) = 8\pi\sum_k f_k(q)i^l j_l(q|\mathbf{r}_k - \mathbf{r}_p|)Y_{lm}^*(\widehat{\mathbf{r}_k - \mathbf{r}_p}) + c.c. \tag{4.8}$$

The quantity j in (4.8) is a spherical Bessel function, and Y is a spherical harmonic.

Substituting (4.7) and (4.8) into (4.1), we deduce that

$$\delta B_l(q, q') = \sum_p M_{lqq';p}(\delta\rho)_p, \tag{4.9}$$

where

$$M_{lqq';p} = \sum_m [I_{lm}^*(q)C_{lm}(q';\mathbf{r}_p) + I_{lm}(q')C_{lm}^*(q;\mathbf{r}_p)], \tag{4.10}$$

where all quantities on the RHS are known (and real) if the dark structure is known (as it usually is in time-resolved experiments) from which we deduce that

$$(\delta\rho)_p = \sum_{lqq'} (M^{-1})_{p;lqq'}\delta B_l(q, q'). \tag{4.11}$$

Note that because of Friedel's law, namely that $I(\mathbf{q}) = I(-\mathbf{q})$, the only non-zero values of I_{lm} and hence of the elements of M are those of even values of l. It should also be noted that

$$\delta B_l(q, q') = B'_l(q, q') - B_l(q, q'), \tag{4.12}$$

where B' is a quantity measurable from pump–probe experiments of the photoexcited structure, and B that of the dark structure. Thus, the RHS of equation (4.11) consists of a product of a column vector consisting of quantities $\delta B_l(q, q')$ measurable by experiment, and a matrix M^{-1} which depends only on the known dark structure. This yields the sought quantity $\delta\rho_p$, which is the difference electron density at voxel p in the frame of reference of the dark structure, and is therefore easily superimposable on models of the molecule, as we see below. Unlike proposed methods of *ab initio* structure determination of viruses from angular correlations of XFEL intensities [21,22], the method we propose here does not depend on a knowledge of the symmetry of the sought difference structure, only that the difference electron density be small compared with the electron density of the dark structure (the usual assumption of time-resolved crystallography).

5. Model calculations

PYP is a popular (15 kDa) protein for time-resolved structural studies. To make a connection with previous work in time-resolved crystallography [27–30], we illustrate our method on the same molecule. As an initial test of these ideas, we simulated diffraction patterns from individual PYP molecules in random orientations. Of course, owing to a lack of long-range order, such diffraction patterns would be expected to be completely diffuse, with no Bragg spots. First, a set of 10 000 diffraction patterns from PYP molecules in a so-called dark state before photoexcitation were simulated, and then similar sets of 10 000 diffraction patterns each from various assumed excited states involving substantial structural changes to one or two residues. The dark-state structure was taken to be that deposited in the protein data bank (PDB) under the entry 2PHY. Various test structures were determined by making torsional angle rotations of various side chains. The calculations were performed on about 200 computer cores on the AVI computer cluster maintained by the University of Wisconsin-Milwaukee. Thus, for each set of 10 000 diffraction patterns, each core computed just 50 diffraction patterns from molecules oriented over a uniform set of orientations in SO(3), in order to simulate the expected range of orientations in which such molecules are likely to be presented to the X-rays in an XFEL 'diffract-and-destroy' experiment.

The angular correlations $J(q, q'; \Delta\phi)$ were then calculated from each simulated diffraction pattern by the formula

$$J(q, q'; \Delta\phi) = \sum_\phi I(q, \phi)I(q', \phi + \Delta\phi), \tag{5.1}$$

where $I(q, \phi)$ is the expected intensity on resolution ring q and azimuthal angle ϕ of each diffraction pattern. An average was then taken of the quantities J over all the simulated diffraction patterns following equation (2.1). From these quantities, the quantities $B'_l(q, q')$ and $B_l(q, q')$ for the photoexcited and 'dark' structures were calculated by inverting equation (2.2). The difference, $\delta B_l(q, q')$ (4.1) between these two quantities, easily extractable from two sets of measured XFEL diffraction patterns, one from a photoexcited molecule and one from a molecule in a dark state, was then the basic simulated experimental data for our algorithm for recovering the difference electron density $(\delta\rho)_p$ at voxel p in the frame of reference of the dark structure (see e.g. equation (4.11)) from sets of randomly oriented photoexcited and dark molecules of PYP. For the known dark structure, we evaluated the elements of the M matrix that also appears on the RHS, and took its inverse. The result of multiplying the two quantities on the RHS is the recovered difference electron density in the frame of reference of the assumed dark structure.

The difference electron density we so recovered for a hypothetical photoexcited structure owing to the assumed differences between the 2 ms and dark structures (consisting of a *cis−trans* isomerization of the chromophore which is a result of a 180° torsional rotation of the head of the chromophore about its C2–C3 bond, as well as a 77° change of the χ_3 torsional angle of the side chain of the ARG 52 residue of PYP [27–30]), as shown in figure 3.

Because, in this case, the two functional groups that change their positions, namely the chromophore and the ARG 52 residue are in close vicinity, the difference densities from the two are somewhat overlapping. In order to see what would be expected of changes in widely separated functional groups, we also tested a different hypothetical structure involving simultaneous torsional angle rotations of the chromophore, and the PHE 121 residue, as illustrated in figure 4. In this case, the difference electron density owing to the chromophore and the PHE 121 residue are widely enough separated for these difference electron densities to be clearly distinguishable.
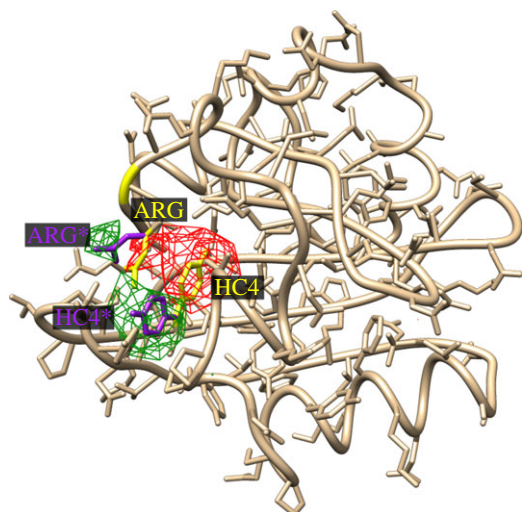
Figure 3. Difference map between the models of the photoexcited and dark structures of PYP assumed in the calculations. The difference between the two structures is that the chromophore (HC4 300) undergoes a *cis − trans* iso-merization which causes it to swing outward towards the edge of the molecule, whereas the ARG 52 residue also swings outwards to make room for the new position of the chromophore in the photoexcited structure. Note that the red lobes of negative electron density are associated with pos-itions of these residues (or functional groups) in the dark structure (yellow bonds) and the green regions of positive difference electron density with their positions (purple bonds) in the excited structure (although the red lobe on the original position of ARG 52 is somewhat cancelled by the fact that it lies between the green lobes of ARG 52 and the chromophore.

Figure 4. Recovered difference electron density from simultaneous torsional angle rotations in a hypothetical structure with structural changes in the widely separated chromophore and PHE 121 residues. Colour conventions same as figure 3.

6. Scattering by multiple particles per diffraction pattern

Next, we wish to comment on a feature of our algorithm which seems to be unique among theories proposed to date for the structure solution of biomolecules from XFEL experiments, namely that it seems to work equally well even if an ensemble of biomolecules in random orientations and random positions contributes to each diffraction pattern, as was realized by Kam [17]. This is only to be expected of a method that seeks aver-aged correlations over a large number of different particle orientations because, regardless of the relative particle orien-tations on a particular snapshot, the averaged values of these quantities over all particle orientations would be expected to be the same.

The importance of this for proposed experiments on indi-vidual protein molecules with a liquid jet [15] is that it would enable the use of a higher concentration of protein molecules than would be possible in an arrangement that illuminates just a single molecule. This increases the relative magnitude of the signal from the molecules compared with solvent and enables the extraction of structural information about a single particle from multiple identical particles in random orientations. Consequently, this would allow the possibility of structure determination of many biomolecules directly from conditions close to their operating conditions in nature, namely in dissolved aqueous environments, without the need for crystallization.

As mentioned above, in order to maximize the protein scat-tering from such a solution stream (and to minimize scattering by the aqueous solution), it would be best to use a solution a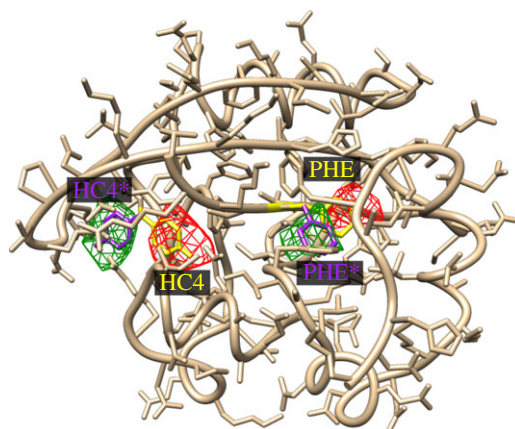s concentrated as possible. It is known that PYP can be concentrated to 150 mg ml^{-1} or higher (M. Schmidt 2011, unpublished data), that is about 10 mmol l^{-1} or higher. In the following, we propose using a concentration of just one-fifth of this amount, namely about 2 mmol l^{-1}. The current design specification of the LCLS is to produce an X-ray beam of per-haps 0.1 μm in diameter at the sample. There is some evidence that it is even possible to focus this beam further, to perhaps 0.01 μm or so (see SPring-8 website at www.spring8. or.jp/en/news_publications/research_highlights/no_50/). If the liquid jet containing the molecules under study can also be reduced to about 0.1 μm in width, then the volume of liquid illuminated by the XFEL beam will be about $\pi 5^2.100$ nm^3. Then, the volume illuminated by the XFEL beam may be arranged to contain just a handful of molecules in random orientations and random positions along the beam. Each XFEL pulse is expected to contain about 10^{12} pho-tons, and the electron cross section may be estimated from the Thomson radius $r_0 = 2.8 \times (10^{-15})$ m to be about $(8\pi/3) \times (2.8 \times 10^{-15})^2$ m$^2 = 0.656 \times 10^{-10}$ nm^2. Now, PYP contains some 1100 atoms or about 7000 electrons. Then, the cross section of scattering by a single molecule will be about 4600×10^{-10} nm^2. If a single XFEL pulse is focused down to a spot diameter of 10 nm, then there will 10^{12} incident photons per 100 nm^2, so the number of photons scattered by a single PYP molecule will be $4600 \times 10^{-10} \times 10^{12} \times 10^{-2} = 4600$ pho-tons, sufficient for a reliable measurement of the angular pair correlations, when averaged over a large number of diffraction patterns, as proposed in this paper. Even in the outermost resolution ring (corresponding to a resolution of approx. 5 Å), the estimated number of photons per Shannon pixel (one may estimate some 120 such pixels from PYP at a resolution of about 5 Å) is about 0.5.

Owing to the expected coherence of an XFEL beam over a volume containing several particles, the scattering from mul-tiple particles involves the addition of the *amplitudes* of the scattered X-rays, rather than only their intensities. However, as we see below, coherent interparticle interference effects are negligible if the average correlations from multiple dif-fraction patterns are calculated, each containing a *different* random distribution of identical particles.

First, we note that the average angular correlations from an ensemble of N particles per diffraction pattern may be written

$$
\begin{aligned}
J(q, q', \Delta\phi) &= \left\langle \left| \sum_{j=1}^{N} A(q, \phi, \Omega_j) e^{iqR_j \cos(\phi - \psi_j)} \right|^2 \left| \sum_{j'=1}^{N} A(q', \phi + \Delta\phi, \Omega_{j'}) e^{iq'R_{j'} \cos(\phi + \Delta\phi - \psi_{j'})} \right|^2 \right\rangle_{DP} \\
&= \left\langle \left[\sum_{j=1}^{N} |A(q, \phi, \Omega_j)|^2 + \sum_{j \neq j'} A^*(q, \phi, \Omega_j) A(q, \phi, \Omega_{j'}) e^{iqR_j \cos(\phi - \psi_j)} e^{iqR_{j'} \cos(\phi - \psi_{j'})} \right] \right. \\
&\quad * \left[\sum_{j'=1}^{N} |A(q', \phi + \Delta\phi, \Omega_{j'})|^2 + \cdots \right. \\
&\quad \left. \left. \sum_{j' \neq j} A^*(q', \phi + \Delta\phi, \Omega_{j'}) A(q', \phi + \Delta\phi, \Omega_j) e^{iq'R_{j'} \cos(\phi + \Delta\phi - \psi_{j'})} e^{iqR_j \cos(\phi + \Delta\phi - \psi_j)} \right] \right\rangle_{DP} \\
&= \left\langle \sum_{j,j'} I(q, \phi, \Omega_j) I(q', \phi + \Delta\phi, \Omega_{j'}) \right\rangle_{DP} \\
&= \left\langle \sum_{j} I(q, \phi, \Omega_j) I(q', \phi + \Delta\phi, \Omega_j) \right\rangle_{DP} + \left\langle \sum_{j \neq j'} I(q, \phi, \Omega_j) I(q', \phi + \Delta\phi, \Omega_{j'}) \right\rangle_{DP},
\end{aligned}
$$

(6.1)

where N is the number of particles (of random orientations) contributing to a particular diffraction pattern, Ω_j is the three-dimensional orientation of particle j (as usually specified by three Euler angles), and R_j and ψ_j are the two-dimensional polar coordinates of the positions of the jth particle in the plane perpendicular to the incident X-rays (this assumes that the scattering vector $\mathbf{q}$ is largely perpendicular to the X-ray beam). We took R_j to be a random number between 0 Å and 200 Å, and ψ_j a random number between 0 and 2π. (Of course, for uniformly distributed particles in real space, the probability of the higher R_j values may be expected to be larger, and it may be necessary to take account of the finite thickness of the liquid jet, and the curvature of the Ewald sphere. We plan to test such more realistic models in future work, though we do not expect our main conclusions to be substantially different.)

A few words are in order about these equations. First, although the simulations assumed perfect coherence between X-rays scattered by different molecules, their random interparticle separations result in very small total contributions from the interference terms between different particles owing to the random phase terms arising from the random particle positions (the second and fourth terms after the second equality). The remaining sums over scattered *intensities* may be further subdivided into the case where the two correlated intensities arise from scattering by the *same* particle, and those where the two intensities arise from the scattering by *dissimilarly oriented* particles. Although the uncorrelated scattering is dominant, it tends to a constant, independent of angle (dashed lines on the left-hand panels of figure 5), a feature that will probably allow their isolation and removal from measured angular correlations (the proof of this will be the subject of a future paper), whereas the correlated scattering by the same particles add up to the same function of $\Delta\phi$ (the solid curves in the left- and right-hand columns of the same figure) as expected for a single particle per snapshot, as illustrated in the right-hand column of figure 5. This should enable the same methods described in the first part of this paper to extract time-resolved structural changes of a single molecule even from experiments on solutions of randomly oriented molecules.

The right-hand column of figure 5 shows the variation of the quantities $J(q, q', \Delta\phi)$ from a set of five particles randomly oriented for any particular snapshot, but averaged over 10 000 diffraction patterns, with each particle in a random orientation in each. These correlations are plotted as a solid line. On these plots, we superimpose the same quantities calculated from diffraction patterns owing to single particles (hatched green curves). For ease of computation, these calculations were performed for an even smaller protein than PYP, namely chignolin (PDB entry 1 UAO). We stress that in both the case of a single particle per snapshot and five particles per snapshot, we averaged the correlations over 10 000 diffraction patterns. Thus, the correlations over the snapshots with five particles per snapshot involved averages over the orientations of some 50 000 particles, whereas those involving a single particle per snapshot involve an average over 10 000 randomly oriented particles. Almost perfect superposition is seen between these two sets of quantities, indicating the diagonal (scattering by same particle) correlations are essentially the same as those from scattering by a single particle. In a future paper, we will describe how to extract the diagonal terms from measured correlations consisting of a sum of diagonal and off-diagonal terms (the latter being angularly independent).

7. Discussion and conclusion

We propose as a method of analysis of the large number of XFEL diffraction patterns each measured from a small number of identical particles in random orientations the calculation of the average angular correlations of their intensities, because this quantity allows the extraction of a quadratic function $B_l(q, q')$ of the spherical harmonic decomposition $I_{lm}(q)$ of intensities from a single particle at each resolution shell q. If the quantities $I_{lm}(q)$ may be extracted from this quadratic function, then this would enable the reconstruction of an *oversampled* three-dimensional diffraction volume of a single particle, and hence a deduction of the electron density of the particle. In the case of *ab initio* structure determination of a single particle, so far this has been demonstrated only for
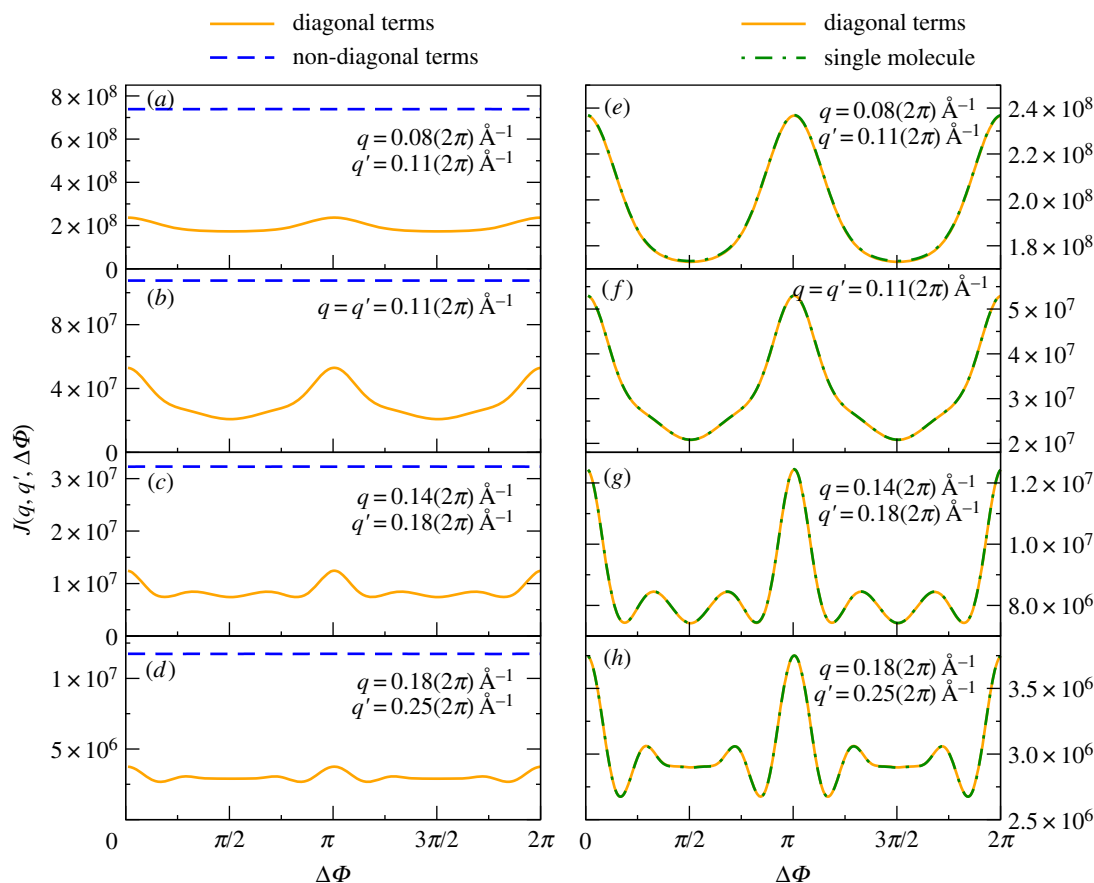
Figure 5. (a−d) Comparison of the 'diagonal' (solid line) and 'off-diagonal' terms (dashed line) in the intensity correlations (first and second terms after the last equality sign in (6.1)). These lines suggest although the 'off-diagonal' terms tend to dominate the measurable angular correlation functions, they tend to be of a relatively simple form (in this case, a constant as a function of angle $\Delta\rho$, a fact that will probably allow their isolation, and hence removal. The remaining 'diagonal' terms contain the structural information of a single particle, as illustrated in (e−h). (e−h) Comparison of $J(q, q', \Delta\phi)$ curves obtained from a single particle per snapshot, and from five randomly oriented particles per snapshot (appropriately divided by 5). These calculations were performed for the small protein chignolin (PDB entry 1 UAO), but illustrate the general point that the sum of the diagonal terms in J (i.e. the intensity correlations from scattering by the same particles) are essentially independent of the number of such particles per snapshot (to within a *factor* of N, the number of particles per snapshot, *even if the particles are randomly oriented*). The radii q and q' of the resolution rings are chosen randomly from our range (up to a maximum value $q_{max} = 2\pi/5$ Å^{-1}, or 5 Å resolution) to demonstrate that the result is essentially independent of the precise values of q and q'. (Online version in colour.)

particles with a known high degree of symmetry, as for example the cases of icosahedral [21] or helical [22] viruses. Another circumstance in which structure solution is possible is when what is sought are small deviations from a known 'dark' structure of a protein in a time-resolved photoexcitation experiment. In this case, it is not necessary to assume *a priori* any symmetry of the difference structure. In an earlier paper, we demonstrated the structural information latent in the quantities $B_l(q, q')$, which may be extracted by a systematic variation of the structural parameters, usually torsional angles, which by and large, specify the differences between the two sets of structures. The problem with this method is that if the residue or small number of residues undergoing the structural change on photoexcitation are unknown, the number of parameters which need to be optimized can be too large for even the best optimization methods available.

Our present simulations suggest that it is possible to find these structural differences at least to the accuracy of the difference Fourier methods of time-resolved crystallography without a prior knowledge of the part of the structure undergoing the structural change on photoexcitation. We do this by pointing out that for the small structural changes typical of fast photoexcitation experiments, the changes $\delta B_l(q, q')$ in the quantities extractable from the photoexcitation experiments described, may be related *linearly* to the difference electron density sought. The linear relationship is specified by a matrix M, whose elements depend only on quantities calculable from a knowledge of the dark structure. What is more, the difference densities sought will be calculated in the frame of reference of the known dark structure, a fact that enables the difference electron density to be displayed on top of the known dark structure using widely used crystallographic display software such as COOT or CHIMERA (although COOT, designed for crystallography, tends to repeat the specified difference electron density periodically as if the structure were a crystal. Of course, in a crystal, the spatial relationship of atoms in different unit cells may be relevant, though it is not relevant in single particle work).

This means that the matrix M (and its inverse) can be calculated once and for all if a dark structure is known, and from this a difference electron density corresponding to each photoexcited structure investigated can be calculated by a simple matrix multiplication with the quantities $\delta B_l(q, q')$ extractable from each experiment. By varying the time delays of each experiment, it should be possible to follow the fast time variation of the structures of photoexcited molecules in solution, rather than of crystallized proteins as in established methods of time-resolved crystallography. A full test on multiple particles of PYP per diffraction pattern will be reported on in a subsequent paper.

Scattering by the solvent would be expected from whatever parts vary rapidly. Subtracting scattered signals from solutions not containing the dissolved molecules would be expected to reduce the scattering from the solvent–air interface common to the solutions with and without the particles. However, there is another interface which exists only for the solutions containing the molecules, namely the solution–molecule interface, which will not be removed by this process. However, we may argue as do exponents of the technique of SAXS [23], that the effect of this interface is not so important at least at relatively low resolutions (below atomic resolution). At such resolutions, the solvent may be regarded as a uniform electron density ρ_s, which we can take to be about $0.33\,e^-\,\text{Å}^{-3}$. The difference from a solvent containing molecules and one without dissolved molecules is the existence of a volume occupied by the molecules that excludes the solvent. However, according to Babinet's principle, the scattering from such a solvent is equivalent to scattering by the volumes occupied by the particles of electron density $-\rho_s$. Adding this quantity to the particle density (say $\rho_p(\mathbf{r})$) results in an effective electron density of the particle of $\rho_p(\mathbf{r}) - \rho_s$ as is argued in the related technique of SAXS [23]. Thus, at least, at the relatively low resolution of our experiment, the effect of solvent scattering is merely to effectively reduce the electron density of the particle by ρ_s, and so the modelling of the scattering is almost the same as that of particles *in vacuo*, except that the effective electron densities of the particles are reduced by ρ_s.

If our results are confirmed in practice, then they suggest it may not be necessary to form crystals for following the fast time variations of the structure of photoexcited molecules. This is desirable for two reasons: (i) many proteins, particularly membrane proteins, are notoriously difficult to crystallize. Indeed, in protein structural work, crystallization is often the rate-limiting step. The freedom from the need to crystallize proteins may allow time-resolved structural changes of a larger class of proteins; (ii) proteins are generally not crystallized when they perform their functions in living entities in nature, but are rather suspended in aqueous solutions, just as in the present experiments. Consequently, there are no extra constraints owing to steric hindrance, which might exist in crystals, in the proposed experiments.

Of course, much still needs to be done to make this a routine method of structure determination. The effects of expected experimental noise need to be added [31] to the multiple simulated XFEL diffraction patterns to see if the same quantities can be recovered in the presence of such noise. And finally, the method still needs to be tested on experiments on known structures with XFEL radiation.

However, the initial indications of the possibility of recovery of information on fast structural changes of uncrystallized biomolecules in solution are exciting portends for the future.

References

1. Moffat K. 1989 Time-resolved macromolecular crystallography. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 309–332. (doi:10.1146/annurev.bb.18.060189.001521)

2. Schmidt M, Rajagopal S, Ren Z, Moffat K. 2003 Application of singular value decomposition to the analysis of time-resolved macromolecular X-ray data. *Biophys. J.* **84**, 2112–2129. (doi:10.1016/S0006-3495(03)75018-8)

3. Schmidt M. 2008 *Ultrashort laser pulses in medicine and biology*. Berlin, Germany: Springer.

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

5. Neutze R, Moffat K. 2012 Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. *Curr. Opin. Struct. Biol.* **22**, 651–659. (doi:10.1016/j.sbi.2012.08.006)

6. Cho A. 2010 What shall we do with an X-ray laser? *Science* **330**, 470–471. (doi:10.1126/science.330.6010.1470)

7. Chapman HN et al. 2011 Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–81. (doi:10.1038/nature09750)

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

9. Saldin DK, Shneerson VL, Fung R, Ourmazd A. 2009 Structure of isolated biomolecules obtained from ultrashort X-ray pulses: exploiting the symmetry of random orientations. *J. Phys. Condensed Matter* **21**, 134014. (doi:10.1088/0953-8984/21/13/134014)

10. Mills DM, Heliwell JR, Kwick Å, Ohta T, Robinson IA, Authier A. 2005 Report of the working group on synchrotron radiation nomenclature brightness, spectral brightness or brilliance? *J. Synchrotron Radiat.* **12**, 385. We adopt here the definition in the paper. (doi:10.1107/S090904950500796X)

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

12. Moffat K. 2001 Time-resolved biochemical crystallography: a mechanistic perspective. *Chem. Rev.* **101**, 1569–1581. (doi:10.1021/cr990039q)

13. Schmidt M, Ihee H, Pahl R, Srajer V. 2005 Protein-ligand interaction probed by time-resolved crystallography. *Methods Mol. Biol.* **305**, 115–154.

14. Weierstall U, Spence JC, Doak RB. 2012 Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83**, 035108. (doi:10.1063/1.3693040)

15. Ganan-Castro AM, De Ponte DP, Herrada MA, Spence JC, Weierstall U, Doak RB. 2010 Liquid capillary micro/nanojets in free-jet expansion. *Small* **6**, 822–824. (doi:10.1002/smll.200901786)

16. Shapiro DA et al. 2008 Powder diffraction from a continuous microjet of submicrometer protein crystals. *J. Synchrotron Radiat.* **15**, 593–599. (doi:10.1107/S0909049508024151)

17. Kam Z. 1978 Determination of macromolecular structure in solution by spatial correlation of scattering fluctuations. *Macromolecules* **10**, 927–934. (doi:10.1021/ma60059a009)

18. Fienup JR. 1982 Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758–2769. (doi:10.1364/AO.21.002758)

19. Oszlányi G, Süto A. 2004 *Ab initio* structure solution by charge flipping. *Acta Crystallogr.* **A60**, 134–141. (doi:10.1107/S0108767303027569)

20. Elser V. 2011 Strategies for processing diffraction data from randomly oriented particles. *Ultramicroscopy* **111**, 788–792. (doi:10.1016/j.ultramic.2010.10.014)

21. Saldin DK, Poon H-C, Schwander P, Uddin M, Schmidt M. 2011 Reconstructing an icosahedral virus from single-particle diffraction experiments. *Opt. Express.* **19**, 17 318–17 335. (doi:10.1364/OE.19.017318)

22. Poon H-C, Schwander P, Uddin M, Saldin DK. 2013 Fiber diffraction without fibers. *Phys. Rev. Lett.* **110**, 265505. (doi:10.1103/PhysRevLett.110.265505)

23. Svergun D, Barberato C, Koch MHC. 1995 CRYSOL: a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates.

J. Appl. Crystallogr. **28**, 768–773. (doi:10.1107/S0021889895007047)

24. Poon H-C, Schmidt M, Saldin DK. 2013 Extraction of fast changes in the structure of a disordered ensemble of photoexcited biomolecules. *Adv. Condensed Matter Phys.* **2013**, 750371. (doi:10.1155/2013/750371)

25. Kirkpatrick S, Gelatt CD, Vecchi MP. 1983 Optimization by simulated annealing. *Science* **220**, 671–680. (doi:10.1126/science.220.4598.671)

26. Rous PJ. 1993 A global approach to the search problem in surface crystallography by low-energy electron diffraction. *Surf. Sci.* **296**, 358–373. (doi:10.1016/0039-6028(93)90030-N)

27. Ihee H *et al*. 2005 From the cover: visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc. Natl Acad. Sci. USA* **102**, 7145–7150. (doi:10.1073/pnas.0409035102)

28. Schmidt M, Pahl R, Srajer V, Anderson S, Ren Z, Ihee H, Rajagopal S, Moffat K. 2004 Protein kinetics: structures of intermediates and reaction mechanism from time-resolved X-ray data. *Proc. Natl Acad. Sci. USA* **101**, 4799–4804. (doi:10.1073/pnas.0305983101)

29. Schotte F *et al*. 2012 Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proc. Natl Acad. Sci. USA* **109**, 19 256–19 261. (doi:10.1073/pnas.1210938109)

30. Jung YO, Lee JH, Kim J, Schmidt M, Moffat K, Srajer V, Ihee H. 2013 Volume-conserving *trans–cis* isomerization pathways in photoactive yellow protein visualized by picosecond X-ray crystallography. *Nat. Chem.* **5**, 212–220. (doi:10.1038/nchem.1565)

31. Kirian RA, Schmidt KE, Wang X, Doak RB, Spence JCH. 2011 Signal, noise, and resolution in correlated fluctuations from snapshot small-angle X-ray scattering. *Phys. Rev. E* **84**, 011921. (doi:10.1103/PhysRevE.84.011921)