

Continuous X-ray diffractive field in protein nanocrystallography

Ruben A. Dilanian,^{a*} Victor A. Streltsov,^b Harry M. Quiney^a and Keith A. Nugent^a

Received 11 May 2012

Accepted 11 October 2012

^aARC Centre of Excellence for Coherent X-ray Science, School of Physics, The University of Melbourne, Melbourne, Victoria 3010, Australia, and ^bARC Centre of Excellence for Coherent X-ray Science, CSIRO Materials Science and Engineering, and Preventative Health Flagship, Parkville, Victoria 3052, Australia. Correspondence e-mail: roubend@unimelb.edu.au

The recent development of X-ray free-electron laser sources has created new opportunities for the structural analysis of protein nanocrystals. The extremely small sizes of the crystals, as well as imperfections of the crystal structure, result in an interference phenomenon in the diffraction pattern. With decreasing crystallite size the structural imperfections play a role in the formation of the diffraction pattern that is comparable in importance to the size effects and should be taken into account during the data analysis and structure reconstruction processes. There now exists a need to develop new methods of protein structure determination that do not depend on the availability of good-quality crystals and that can treat proteins under conditions close to the active form. This paper demonstrates an approach that is specifically tailored to nanocrystalline samples and offers a unique crystallographic solution.

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

Growing crystals of membrane proteins to the size suitable for high-resolution X-ray structure analysis has always been a major challenge in structural biology (Garcia-Ruiz, 2003; Malkin & Thorne, 2004; Baker, 2010). Membrane proteins often tend to form nanoscale crystals (Caffrey, 2003; Fromme & Spence, 2011) which diffract weakly even at synchrotron X-ray sources. The need for long exposure times and, as a result, the delivery of high X-ray doses to record high-resolution data lead to extensive damage of the sample. The recent development of extremely bright X-ray free-electron laser (XFEL) sources has created an opportunity for the structure analysis of such protein nanocrystals (Chapman *et al.*, 2011; Boutet *et al.*, 2012; Johansson *et al.*, 2012; Koopmann *et al.*, 2012). The 'diffract and destroy' approach of XFEL experiments, in which diffraction data are collected from a stream of protein nanocrystals of varying sizes and shapes in random orientations, requires, however, the development of new methods for the structural analysis of proteins.

The size and quality of the protein crystals used in structural analysis are of particular importance. The implications of structural imperfections for solving the structure of proteins by conventional techniques have been identified and analysed (Faure *et al.*, 1994; Mizuguchi *et al.*, 1994; Eyal *et al.*, 2005; Welberry, 2004; Welberry *et al.*, 2011). It has been noted that conventional protein crystallography relies almost entirely on the analysis of Bragg diffraction data. Scattering between the Bragg reflections, however, contains information in addition to that obtainable from the Bragg peaks and should be included in the structural analysis of proteins.

The situation is much more complicated for nanoscale crystals. The extremely small size of the crystals, as well as imperfections of the crystal structure, result in an interference phenomenon in the diffraction pattern and influence the shape of the Bragg peaks, as well as the scattering between them (Vainshtein, 1966; Welberry, 2004; Rafaja *et al.*, 2000, 2004). The mechanisms that suppress the growth of crystals are poorly understood. Some studies suggest that the incorporation of errors leads to a 'poisoning' of the surface (Feher & Kam, 1985; Grant & Saville, 1994). The presence of the lipid or detergent components of the crystallization solution can limit the ideal long-range packing in the formation of crystals. During crystal growth, structural defects and chemical impurities are incorporated until they accumulate to such an extent on the surface of the crystal that further building of a well aligned crystalline lattice becomes energetically unfavourable. If we consider the crystal as a cube of size $L = na$, where a is the unit-cell parameter and n is the number of the unit cells along the a direction, then the ratio of the number of unit cells on the crystal surface, N_s , to the total number of unit cells, $N_v = n^3$, can be estimated to be $N_s/N_v = 1 - [1 - 2(a/L)]^3$. Fig. 1 shows that the fraction of unit cells at the surface of the crystal increases rapidly as the crystal size decreases. As examples, 400 nm-size crystals with 100, 200 and 300 Å unit cells contain, respectively, approximately 14, 27 and 39% surface unit cells. It is clear that the effects due to disorder of a large number of surface unit cells cannot be neglected in the structural analysis of the nanoscale crystals. Moreover, structural imperfections play a role in the formation of the diffraction patterns that is comparable in importance with the size effects if the crystallite size is reduced. Consequently, the

extraction of the integrated intensities of Bragg reflections from the diffraction pattern is a major problem in protein nanocrystallography.

A new method for the structural analysis of protein nanocrystals has recently been demonstrated (Kirian *et al.*, 2010, 2011; Spence *et al.*, 2011). The total intensity of X-rays scattered from the nanocrystals can be described as a simple product of two factors: the form factor of the crystal (CFF) and the molecular form factor (MFF) of the protein. Assuming that the MFF is identical for all crystals but the CFF depends on the size and shape of a crystal, one can derive the electron density of the protein molecule by properly averaging the CFF over all crystals and using iterative phasing methods. The intensities of reflections are integrated within the Wigner–Seitz cell around each Bragg peak, which provides sufficient information for the direct phasing of the MFF.

The method further assumes that crystal size effects dominate in the formation of the diffraction pattern of the protein nanocrystals. It is assumed that the density of structural defects in protein nanocrystals is small and that such effects can, therefore, be neglected or corrected by the static Debye–Waller factor (Kirian *et al.*, 2010). Our analysis suggests, however, that structural imperfections may play an important role in the formation of the diffraction pattern of nanocrystals. Moreover, a reason why crystals may not grow beyond micron size is that they are not ideal, suggesting strongly that the structural imperfections should be included in the structural analysis.

Here we present an approach in which the diffraction pattern of the protein nanocrystals is regarded as a continuous function of the scattering vector, rather than as a discrete set of Bragg reflections. It has been suggested that the diffraction data that are available using XFEL sources offer the possibility of direct phasing of the continuous scattering from single proteins or protein nanocrystals of varying sizes and orientations (Fung *et al.*, 2009; Kirian *et al.*, 2010). With this in mind, one can build a continuous three-dimensional diffraction pattern using a sufficient quantity of two-dimensional diffraction snapshots, collected from a stream of randomly

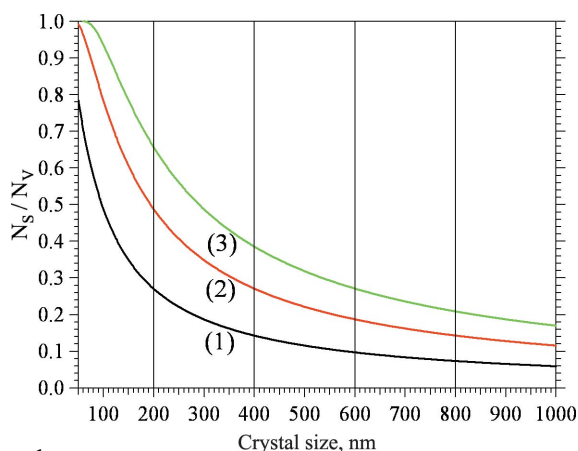


Figure 1
 N_s/N_v ratio as a function of crystal size for different unit-cell parameters: (1) $a = 10.0$ nm, (2) $a = 20.0$ nm and (3) $a = 30.0$ nm.

Table 1
 Notations and symbols.

λ	Wavelength
\mathbf{q}	Scattering vector
$q = \mathbf{q} $	$q = 2\pi S$, where $S = 2 \sin(\theta)/\lambda$
$\mathbf{r}', \mathbf{r}, \mathbf{R}, \mathbf{u}, \Delta \mathbf{r}$	Vectors which define the actual and average positions of atoms, the origin of the molecular cluster, the ideal position of the atom in the molecular cluster, and the displacement of atoms from the average positions, respectively
$\rho_C(\mathbf{r}), \rho_M(\mathbf{r})$	Electron-density distribution in the crystal and in the unit cell, respectively
$O(\mathbf{r})$	Lattice translation function
$S(\mathbf{r})$	Shape function, defines the shape of the crystal
$\Phi(\mathbf{q})$	Molecular form factor (MFF); scattering from the cluster of protein molecules, located in the one unit cell
$T(\mathbf{q})$	Translation factor. Defines structural information, specified by the ideal nuclear positions of atoms (molecules)
$\Lambda(\mathbf{q}), \tilde{\Lambda}(\mathbf{q})$	Interference functions produced by the ideal and disordered lattice of molecular clusters, respectively
$g(\mathbf{q})$	The weighting factor, which describes the degree of coherence of two molecular clusters
$A(\mathbf{q})$	Two-dimensional matrix, elements of which describe the degree of coherence of two waves, scattered from two molecular clusters
$\langle W(\mathbf{q}) \rangle$	An average crystal form factor (CFF)
Δ_m^2	The mean-square displacement (MSD) of the m th neighbouring molecular cluster
$I(\mathbf{q})$ or $I_m(\mathbf{q})$	Intensity of X-rays scattered from the nanocrystal
$I_T(\mathbf{q}), I_{\text{exp}}(\mathbf{q})$	Calculated and measured (simulated) integrated intensities of X-rays scattered from a set of nanocrystals of different shapes, orientations, structural disorders, respectively

oriented nanocrystals. Coherent diffractive imaging (CDI) methods, which have generally been developed for the analysis of non-periodic finite objects (Fienup, 1982; Miao *et al.*, 1998; Spence, 2004; Chapman *et al.*, 2006; Quiney, 2010) are then applied to reconstruct the electron density of the protein molecule. We will demonstrate that the variation of crystal shapes and structural imperfections can be incorporated into the reconstruction process as an additional source of partial coherence of the X-rays scattered from the sample. It has recently been shown that the explicit incorporation of models of partial coherence into the solution of the phase problem significantly improves the quality of reconstruction in diffractive imaging applications (Whitehead *et al.*, 2009; Chen *et al.*, 2009; Abbey *et al.*, 2011; Quiney & Nugent, 2011).

Table 1 summarizes notations and symbols frequently used in this paper.

2. Initial conditions and validity of the method

Diffraction data collected using an XFEL source consist of two-dimensional diffraction patterns (snapshots) of randomly oriented particles (nanocrystals). Each snapshot represents the diffraction pattern from a single particle of finite size and a unique orientation, but each particle possesses its own size, orientation and degree of structural disorder. The structure of a protein molecule is determined from the three-dimensional diffracted intensity distribution by applying the iterative phase retrieval technique, which should be modified to incorporate *a priori* information pertaining to the experimental conditions. Several critical questions, which influence the validity of our method, are discussed below.

2.1. Data acquisition process and the particle orientation problem

In recent years, considerable attention has been given to the problem of the analysis of new types of data obtained using XFEL sources (Shneerson *et al.*, 2008; Loh & Elser, 2009; Fung *et al.*, 2009; Schwander *et al.*, 2010; Kirian *et al.*, 2010; Spence *et al.*, 2011; Tokuhisa *et al.*, 2012). It has been shown that the orientations of individual particles or even single molecules can be determined from the XFEL diffraction data (Fung *et al.*, 2009; Kirian *et al.*, 2010). In the case of nanocrystals, this process is more straightforward, since it is based on indexing very strong Bragg reflections. It has been shown that the existing auto-indexing programs are capable of determining unit-cell parameters and orientations with high accuracy (Kirian *et al.*, 2010).

The scope of this article is to analyse the influence of structural imperfections on the three-dimensional diffraction pattern obtained from a set of randomly oriented, structurally imperfect nanocrystals of different shapes. In our analysis we follow earlier work (Kirian *et al.*, 2010) which describes techniques for merging two-dimensional snapshots into a three-dimensional intensity distribution volume.

2.2. Total number of photons and the signal-to-noise ratio

The successful application of any iterative algorithm depends on the availability of an accurate model for the data acquisition process. A common deviation from the ideal conditions arises because of the impact of noise or low photon counts on the diffraction data. Recent work has investigated the impact of these factors on the quality of reconstruction and explored the point at which the iterative scheme fails to converge to a correct solution (Williams *et al.*, 2007). It has been shown that the critical number of scattered photons for

the successful reconstruction of a two-dimensional projection is $N_c \simeq 10^5$ and the signal-to-noise ratio is $\text{SNR}_c \simeq 290$, where SNR is defined by

$$\text{SNR} = \frac{\sum_{k=1}^N I_k}{\sum_{k=1}^N (I_k)^{1/2} + P_k + B_k} \quad (1)$$

In equation (1), I_k is the number of photons collected in pixel k , P_k is the signal due to photons scattered into the detector by alien sources, and B_k is the bias level arising from the detector and electronics. We will base our simulations on these critical parameters.

2.3. Conditions critical for the CDI approach

Several factors have a strong influence on the setup of a CDI experiment.

(a) The diffraction pattern should satisfy the oversampling criterion

$$\frac{\text{total number of sample points in one dimension}}{\text{portion occupied by object}} \geq 2 \quad (2)$$

for each dimension (Bates, 1982; Quiney, 2010).

(b) The far-field condition, $d^2/(\lambda Z) \ll 1$, where d is the size of the object (nanocrystal), λ is the wavelength of the X-ray source and Z is the object–detector distance.

(c) The discrete Fourier transform relation for the sampling intervals, Δ_S and Δ_D , in the object and detector planes, respectively,

$$\Delta_S = \frac{\lambda Z}{N \Delta_D}, \quad (3)$$

where N is the number of pixels along one side of the detector. Meanwhile, the crystallographic resolution is defined as $d_{\min} = \lambda/2 \sin \theta_{\max}^B$, where θ_{\max}^B is the maximum Bragg angle that can be measured (see Fig. 2a), or

$$d_{\min} = \frac{\lambda}{2 \sin[\frac{1}{2} \text{atan}(N \Delta_D / 2Z)]} \quad (4)$$

As one can see from equations (3) and (4), $d_{\min} \simeq 2\Delta_S$.

(d) Since our approach is based on the analysis of the continuous three-dimensional diffraction pattern, correct sampling of intensities between the Bragg reflections is also important. The number of points between the Bragg reflections can be defined as (see Fig. 2a)

$$m = N \frac{\tan(2\theta_1)}{\tan(2\theta_{\max})}, \quad (5)$$

where $\theta_1 = \text{asin}(\lambda/2a)$. Consider the crystal as a cube of size $L = na$, where a is the unit-cell parameter and n is the number of unit cells along the a direction. Consequently, the number of fringes between the $(h00)$ Bragg reflections is equal to $p = n - 2$. Our analysis shows that the critical number of measured points, m_c , is given by $m_c = 2p + 1$, corresponding to two measured points for each subsidiary fringe.

Taking all of the above requirements into consideration, the parameters chosen for our simulations are summarized in Table 2.

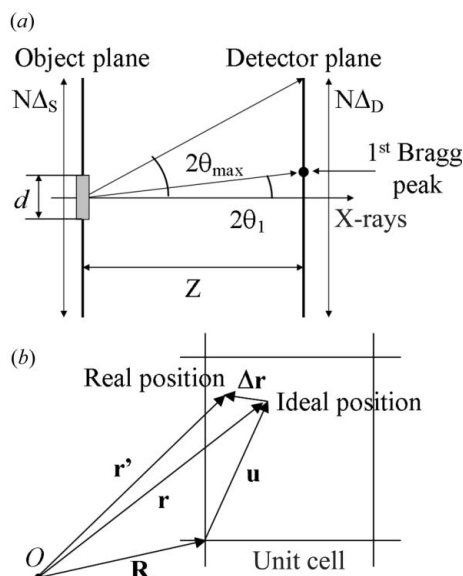


Figure 2
(a) Schematic of the CDI experiment. (b) Vector representation of the position of the atom in the crystal lattice.

Table 2

Experimental parameters used in simulations.

Dimensions of the detector	1800 × 1800
Pixel sizes at the detector plane, Δ _D	110 μm
Wavelength, λ	1 Å
Total number of scattered photons for each two-dimensional projection of the three-dimensional diffraction pattern	10 ⁵
Crystal size	~200 nm
Crystallographic resolution	~4.1 Å

3. Scattering from the partially disordered protein nanocrystal

In this section we consider the idealized case of diffraction from a single nanocrystal of finite size. Although the influence of structural imperfection on the diffraction pattern of protein crystals has been analysed previously (Mizuguchi *et al.*, 1994; Rafaja *et al.*, 2000, 2004; Welberry, 2004; Welberry *et al.*, 2011), it is useful to summarize the main features in order to provide necessary information for the analysis that follows.

We consider a cluster of proteins located in the unit cell as the constituent elements of the crystal structure. We call such a group of proteins a ‘molecular cluster’. In other words, each node of the three-dimensional periodic lattice corresponds to the ideal position of the molecular cluster. We further assume that the molecular cluster consists of identical protein molecules, replicated within the unit cell according to the symmetry of the crystal.

We first consider the intensity of X-rays scattered from an ideal crystal. The amplitude of X-rays scattered by a nanocrystal is represented in the form (see Fig. 2*b*)

$$\Psi(\mathbf{q}) = \sum_l \sum_k f_k(q) \exp(i\mathbf{q} \cdot \mathbf{r}_{kl}) \exp(i\mathbf{q} \cdot \mathbf{R}_l), \quad (6)$$

where \mathbf{R}_l is the position of the *l*th unit cell in the crystal or, according to our definitions, the position of the *l*th molecular cluster in the crystal. The position of the *k*th atom in the *l*th molecular cluster is denoted \mathbf{r}_{kl} , $f_k(q)$ is the atomic scattering vector of the *k*th atom, \mathbf{q} is the scattering vector, $q = |\mathbf{q}|$ and summations are performed over all atoms, labelled *k*, and all unit cells, labelled *l*. Scattering from the molecular cluster is determined by the so-called MFF defined by

$$\Phi_l(\mathbf{q}) = \sum_k f_k(q) \exp(i\mathbf{q} \cdot \mathbf{r}_{kl}). \quad (7)$$

Suppose all molecular clusters in all unit cells are identical and, therefore, $\Phi_l(\mathbf{q}) = \Phi(\mathbf{q})$ for all clusters in the crystal. Using * to signify the complex conjugate, the intensity of X-rays scattered from an ideal crystal is obtained through the relation $I(\mathbf{q}) \propto \Psi(\mathbf{q})\Psi^*(\mathbf{q})$, so that

$$I(\mathbf{q}) \propto |\Phi(\mathbf{q})|^2 \sum_{j,l} \exp(i\mathbf{q} \cdot \mathbf{R}_j) \exp(-i\mathbf{q} \cdot \mathbf{R}_l), \quad (8)$$

or

$$I(\mathbf{q}) \propto |\Phi(\mathbf{q})|^2 \Lambda(\mathbf{q}). \quad (9)$$

In equation (9), $\Lambda(\mathbf{q})$ represents the interference function produced by the three-dimensional periodic lattice of mole-

cular clusters. For large crystals, the scattered intensity is negligible everywhere except near the positions of the Bragg reflections.

To incorporate structural imperfections into the analysis of the diffraction data we define the positions of atoms as $\mathbf{r}'_{kl} = \mathbf{r}_k + \Delta\mathbf{r}_{kl}$, where \mathbf{r}_k is the average position of the *k*th atom in the molecular cluster, which is identical for all clusters, and $\Delta\mathbf{r}_{kl}$ denotes the displacement of the *k*th atom from the average position in the *l*th molecular cluster (Fig. 2*b*).

Consequently, the intensity of X-rays scattered from the partially disordered nanocrystal can be rewritten in the following form [see Vainshtein (1966) or Rafaja *et al.* (2000) for more details]:

$$I(\mathbf{q}) \propto |\Phi(\mathbf{q})|^2 \tilde{\Lambda}(\mathbf{q}), \quad (10a)$$

where

$$\tilde{\Lambda}(\mathbf{q}) = \sum_{j,l} \exp(i\mathbf{q} \cdot \mathbf{R}_j) g_{jl}(\mathbf{q}) \exp(-i\mathbf{q} \cdot \mathbf{R}_l). \quad (10b)$$

In equation (10*b*), $g_{jl}(\mathbf{q})$ is the weighting factor which describes the degree of the coherence of two molecular clusters located at positions \mathbf{R}_j and \mathbf{R}_l . It is equal to unity for an ideal crystal, and equal to zero for the totally incoherent formation of molecular clusters. It is defined by

$$g_{jl}(\mathbf{q}) = \langle \exp[i\mathbf{q} \cdot (\Delta\mathbf{r}_{kj} - \Delta\mathbf{r}_{li})] \rangle, \quad (11)$$

where angle brackets denote an average over all molecular clusters of the crystal. As one can see from equation (10*b*), the behaviour of the interference function, $\tilde{\Lambda}(\mathbf{q})$, is governed by the form of $g_{jl}(\mathbf{q})$. The $g_{jl}(\mathbf{q})$ factors describe the structural imperfections, such as the loss of the coherence of molecular clusters due to their misorientations (Rafaja *et al.*, 2004),

$$g_{jl}(\mathbf{q}) = \min \left[\frac{1 - (12q^2 D_m^2)^{-1}}{qQ(\mathbf{q})D_m}, 1 \right], \quad (12a)$$

or the displacement of molecular clusters from their ideal positions (Rafaja *et al.*, 2000),

$$g_{jl}(\mathbf{q}) = \exp(-2\pi^2 q^2 \Delta_m^2). \quad (12b)$$

Here, $m = |j - l|$, D_m is the distance between two molecular clusters located at positions \mathbf{R}_j and \mathbf{R}_l , and $Q(\mathbf{q})$ denotes the ‘texture’ or ‘preferred orientation’ function which characterizes the angular distribution of the molecular clusters. The mean-square displacement (MSD) of the *m*th neighbouring molecular cluster is denoted Δ_m^2 , and can be defined as $\Delta_m^2 = m\Delta_1^2$ (Vainshtein, 1966) where Δ_1^2 is the MSD of the nearest-neighbour molecular clusters.

According to equation (10*b*), the presence of structural imperfections in protein nanocrystals leads to broadening of diffraction peaks in addition to the size effects. Fig. 3 shows the interference function calculated for the ideal and the partially disordered protein nanocrystal. In our analysis, we assumed that $\Delta_1 = (\Delta_{1a}, \Delta_{1b}, \Delta_{1c})$, $\Delta_{1a} = \alpha_1 d_a$, $\Delta_{1b} = \alpha_1 d_b$ and $\Delta_{1c} = \alpha_1 d_c$, where d_a , d_b and d_c are the interplanar distances along the crystallographic directions *a*, *b* and *c*, respectively, and α_1 , α_2 and α_3 are variable parameters. As one

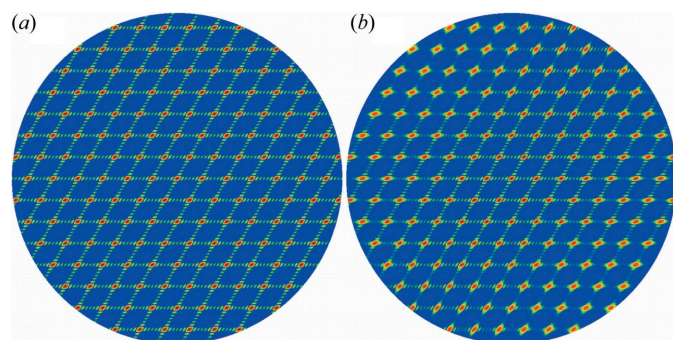


Figure 3
The $(hk0)$ plane of the interference function calculated for (a) the ideal nanocrystal and (b) for the partially disordered nanocrystal. The $g_{jl}(\mathbf{q})$ is calculated using equation (12b), assuming that $\Delta_1 = \alpha d_{100}$, where d_{100} is the interplanar distance along the (100) direction of the crystal lattice and $\alpha = 0.005$.

can see from Fig. 3(b), the broadening effect is more pronounced at high- q regions of the diffraction pattern. Detailed analysis of the interference function, $\tilde{\Lambda}(\mathbf{q})$, however, shows that the appearance of diffraction fringes between the Bragg reflections is extremely sensitive to the $g_{jl}(\mathbf{q})$ function, even at the low- q regions (Fig. 4).

In our analysis, we have assumed that the molecular clusters consist of identical protein molecules. Protein crystals are, however, often grown as complexes of two or more proteins which introduces so-called conformational disorder. Conformational disorder does not affect the shape of Bragg peaks, but reduces the intensity of reflections by the additional Debye–Waller factor and also leads to continuous diffuse scattering in inter-Bragg regions (Vainshtein, 1966; Welberry, 2004).

The interference phenomenon which can be observed in the diffraction patterns of nanocrystals depends on contributions to $g_{jl}(\mathbf{q})$ (Vainshtein, 1966; Welberry, 2004). As follows from the analysis, the structural imperfections of nanocrystals lead to broadening of the Bragg reflections, to decay of their peak intensities and to modification of the scattering between the Bragg reflections. In conventional protein crystallography, critical assumptions are usually made: the density of structural defects is small and displacements of atoms are mutually uncorrelated. In this case, the static Debye–Waller factor can be used to account for the influence of structural imperfections on the diffraction pattern. The intensity of X-rays scattered from the partially disordered nanocrystal is transformed into the form $I(\mathbf{q}) \propto |\Phi(\mathbf{q})|^2 \Lambda(\mathbf{q}) \exp(-Bq^2)$, where B is the Wilson factor (Giacovazzo, 2011). Generally, however, $g_{jl}(\mathbf{q})$ is a non-symmetrical, continuous function that exists in the entire range of \mathbf{q} and not just around the Bragg reflections. The averaging of intensities only around the positions of the Bragg reflections ignores most of the effects of the structural imperfections, leading to the loss of important structural information. From this point of view, extracting the integrated intensities of the Bragg reflections from the diffraction pattern is a formidable task.

To overcome this problem one should consider the diffraction pattern from the partially disordered protein

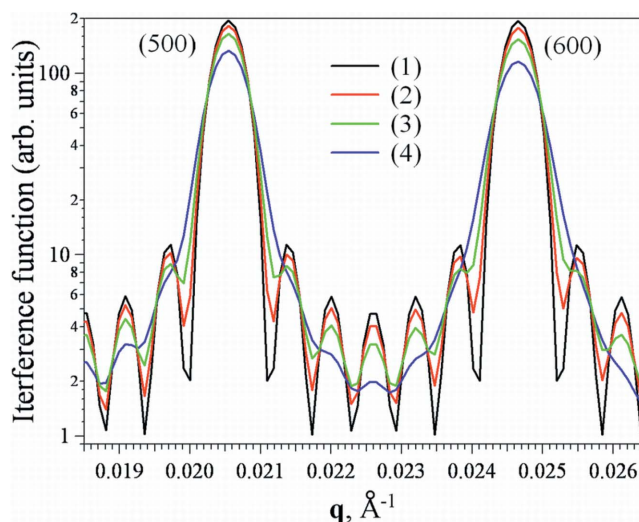


Figure 4
The variation of the interference function, $\tilde{\Lambda}(\mathbf{q})$, between (500) and (600) Bragg reflections for the hexagonal unit cell, $a = b = 281$, $c = 165$ Å. The $g_{jl}(\mathbf{q})$ is calculated using equation (12b), assuming that $\Delta_1 = \alpha d_{100}$, where d_{100} is the interplanar distance along the (100) direction of the crystal lattice. (1) $\alpha = 0.0$, (2) $\alpha = 0.002$, (3) $\alpha = 0.005$, (4) $\alpha = 0.008$.

nanocrystals not as a discrete set of intensities of the Bragg reflections, but as a continuous function of \mathbf{q} . In this case, the approach that is now known as the CDI technique can be applied for structure analysis of the protein nanocrystals. The ability of the CDI approach to reconstruct the object comprised of a small number of periodic elements has recently been demonstrated (Chen *et al.*, 2009). The CDI reconstruction technique is based on Fienup's extensions of the algorithm first proposed by Gerchberg & Saxton (1972) and Fienup (1982). The Gerchberg–Saxton–Fienup (GSF) algorithm propagates a numerical representation of a scalar wavefield between 'object' and 'detector' planes using instances of the Fourier transformation. The wavefield in these fixed planes is constrained by the application of *a priori* information, and the iteration between planes is continued until self-consistency is achieved (Fienup, 1982). The *continuous diffractive field* produced by the nanocrystals contains all the information necessary for structure analysis and, unlike conventional crystallography, provides a unique solution in a single set of data (Bates, 1982). According to the analysis presented in this section, $g_{jl}(\mathbf{q})$ may be regarded as describing the degree of coherence of the X-rays scattered from two molecular clusters, located at positions \mathbf{R}_j and \mathbf{R}_l . The structural imperfections in a protein nanocrystal can, therefore, be considered as a source of partially coherent X-rays, scattered from the ensemble of molecular clusters arranged in a three-dimensional periodic lattice. Recent developments in CDI show that the coherent properties (spatial and temporal) of X-ray sources can be incorporated directly into the CDI reconstruction process (Whitehead *et al.*, 2009; Chen *et al.*, 2009; Abbey *et al.*, 2011). Moreover, the radiation damage caused by the interaction of X-rays of the XFEL source with the protein molecules can also be incorporated into the structural analysis by introducing models of partial coherence (Quiney & Nugent, 2011); we have utilized the same general

approach in our analysis. Furthermore, $\Phi(\mathbf{q})$ and, therefore, the electron density distributed in the unit cell of the protein nanocrystal can be reconstructed from measured intensities $I(\mathbf{q})$ by introducing models of $g_{jl}(\mathbf{q})$ into the algorithm.

4. Total intensity of X-rays scattered from protein nanocrystals of different shapes

In this section, we assume that sufficient two-dimensional diffraction patterns of randomly oriented particles have been collected to build a three-dimensional intensity distribution volume. Each snapshot represents the diffraction pattern from a single particle of finite size and a unique orientation, but each particle has its own size, orientation and degree of disorder. The orientation of each nanocrystal is determined from the two-dimensional diffraction pattern by the indexing approach presented in Kirian *et al.* (2010). The diffraction patterns for each orientation are accumulated to achieve the critical number of scattering photons, $N_c \simeq 10^5$, from which the three-dimensional diffraction pattern is ultimately constructed. We also assume that all of the criteria described in §2 are satisfied.

We consider diffraction from the set of M distinct nanocrystals of different shapes $S = \{S_1(\mathbf{r}), \dots, S_M(\mathbf{r})\}$, where $S_m(\mathbf{r})$ defines the shape of the m th nanocrystal. We assume that each nanocrystal fits a box, Ω , with dimensions $L_a \times L_b \times L_c$, formed by the ideal unit cell of the nanocrystal translated periodically along the lattice vectors \mathbf{a} , \mathbf{b} and \mathbf{c} , shown in Fig. 5. The position of the j th atom of the m th crystal in the Ω lattice is defined as $\mathbf{r}'_{jm} = \mathbf{r}_j + \Delta\mathbf{r}_{jm}$, or $\mathbf{r}'_{jm} = \mathbf{R}_l + \mathbf{u}_{kl} + \Delta\mathbf{r}_{jm}$, where \mathbf{r}_j is the average position of the j th atom in the crystal, \mathbf{R}_l is the origin of the l th unit cell (molecular cluster) in Ω , \mathbf{u}_{kl} is the ideal position of the k th atom in the l th molecular cluster and $\Delta\mathbf{r}_{jm}$ denotes the displacement of the j th atom from the average position in the m th crystal (Fig. 2*b*). Each unit cell of Ω is assigned an occupation factor, $p^m(\mathbf{r}_j)$, which is equal to unity if $\mathbf{r}_j \in S_m(\mathbf{r})$ and zero otherwise.

The intensity of X-rays scattered from the m th crystal can be written in the following general form (see also §3):

$$I_m(\mathbf{q}, t) \propto I_0(t) \sum_{j,k} f_j^m(q, t) f_k^m(q, t) p^m(\mathbf{r}_j) p^m(\mathbf{r}_k) T_j(\mathbf{q}) T_k^*(\mathbf{q}) g_{jk}^m(\mathbf{q}). \quad (13)$$

Here, the summation is performed over all atoms located in Ω , $I_0(t)$ is the time-dependent intensity of the incident pulse, $f_j^m(q, t)$ is the elastic scattering factor of the j th atom in the m th crystal, $T_j(\mathbf{q}) = \exp(i\mathbf{q} \cdot \mathbf{r}_j)$ and $g_{jk}^m(\mathbf{q}) = \exp[i\mathbf{q} \cdot (\Delta\mathbf{r}_{jm} - \Delta\mathbf{r}_{km})]$.

The integrated intensity, $I_T(\mathbf{q})$, of X-rays scattered from a set of nanocrystals of different shapes can then be defined as

$$I_T(\mathbf{q}) \propto \sum_{j,k} T_j(\mathbf{q}) A_{jk}(\mathbf{q}) T_k^*(\mathbf{q}), \quad (14)$$

where

$$A_{jk}(\mathbf{q}) = \langle I_0(t) f_j^m(q, t) f_k^m(q, t) p^m(\mathbf{r}_j) p^m(\mathbf{r}_k) g_{jk}^m(\mathbf{q}) \rangle. \quad (15)$$

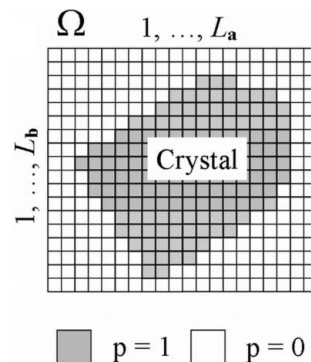


Figure 5
Schematic of the crystal structure representation; the (ab) plane of the Ω box is shown.

Angle brackets denote an average over time, all crystal shapes and all structural distortions. As one can see from equations (14) and (15), all of the structural information, specified by the ideal nuclear positions of atoms, is contained in $T_j(\mathbf{q})$, while all structural imperfections are combined within $A_{jk}(\mathbf{q})$. One may determine one of these quantities from the measurements of $I_T(\mathbf{q})$ if the other is known. The goal of our presentation is the structural analysis of proteins to determine $T_j(\mathbf{q})$. Since $A_{jk}(\mathbf{q})$ is a continuous function of \mathbf{q} , one can also characterize the structural imperfections from measured intensities, $I_T(\mathbf{q})$, in crystals with known ideal (average) structure. We consider potential applications in the following examples.

4.1. Scattering from ideal crystals of different shapes

We assume that the density of structural defects is small and, therefore, the influence of such defects on the diffraction pattern can be neglected or taken into account by the static Debye–Waller factor, $\exp(-Bq^2)$. The total intensity, $I_T(\mathbf{q})$, can be rewritten in the form

$$I_T(\mathbf{q}) \propto |\Phi(\mathbf{q})|^2 \sum_{j,k} T_j(\mathbf{q}) A_{jk}(q) T_k^*(\mathbf{q}) \quad (16a)$$

and

$$A_{jk}(\mathbf{q}) = \langle I_0(t) p^m(\mathbf{r}_j) p^m(\mathbf{r}_k) \rangle, \quad (16b)$$

where $T_j(\mathbf{q}) = \exp(i\mathbf{q} \cdot \mathbf{R}_j)$. Equation (16*a*) can also be represented in the form

$$I_T(\mathbf{q}) = |\Phi'(\mathbf{q})|^2 \langle S(\mathbf{q}) \rangle, \quad (17)$$

where $\Phi'(\mathbf{q}) = \Phi(\mathbf{q}) \exp(-Bq^2)$ and $\langle S(\mathbf{q}) \rangle$ is the average CFF. Since the CFF is the same in the neighbourhood of every Bragg reflection for a given crystal, equation (17) can be transformed into the same form discussed in Spence *et al.* (2011) and Kirian *et al.* (2010).

4.2. Radiation damage or substitution disorder effects

The integrated intensity, $I_T(\mathbf{q})$, can be rewritten in the form

$$I_T(\mathbf{q}) \propto \sum_{j,k} p_j p_k T_j(\mathbf{q}) A_{jk}(q) T_k^*(\mathbf{q}) \quad (18a)$$

and

$$A_{jk}(\mathbf{q}) = \langle I_o(t)f_j(q, t)f_k(q, t) \rangle, \quad (18b)$$

where $T_j(\mathbf{q}) = \exp(i\mathbf{q} \cdot \mathbf{r}_j)$. It is readily shown that equation (18a) can be transformed into the same form discussed in Quiney & Nugent (2011).

4.3. Structural disorder

This case is the target of our analysis. We consider only sources of structural disorder, such as mutual displacement or misorientations of molecular clusters. The current analysis does not include conformational disorders, though these are readily incorporated into the analysis. The integrated intensity, $I_T(\mathbf{q})$, is represented in the form

$$I_T(\mathbf{q}) \propto \sum_{j,k} T_j(\mathbf{q})A_{jk}(\mathbf{q})T_k^*(\mathbf{q}) \quad (19a)$$

and

$$A_{jk}(\mathbf{q}) = \langle p^m(\mathbf{R}_j)p^m(\mathbf{R}_k) \exp[i\mathbf{q} \cdot (\Delta\mathbf{r}_{jm} - \Delta\mathbf{r}_{km})] \rangle, \quad (19b)$$

where $T_j(\mathbf{q}) = \Phi(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{R}_j)$. The elements of the matrix $A(\mathbf{q})$ describe the degree of coherence of two waves, scattered from two molecular clusters separated by the distance $|\mathbf{R}_j - \mathbf{R}_k|$. Consequently, the structure of the protein crystal can be derived from $I_T(\mathbf{q})$ by introducing models of coherency, $A_{jk}(\mathbf{q})$.

According to the coherent mode formulation of coherence theory, proposed by Wolf (1982), the relationship between the coherent properties of molecular clusters and a far-field scattered intensity can be represented in terms of a modal expansion of the form

$$I_T(\mathbf{q}) = \sum_{k=1}^M \eta_k \Psi_k(\mathbf{q})\Psi_k^*(\mathbf{q}), \quad (20)$$

where η_k are real, non-negative numbers representing the occupancy of the mode $\Psi_k(\mathbf{q})$ and M is the number of such modes. The modes are themselves mutually incoherent. It has recently been demonstrated (Whitehead *et al.*, 2009; Chen *et al.*, 2009; Quiney & Nugent, 2011; Abbey *et al.*, 2011) that the incorporation of models of partial coherence into CDI analysis significantly improves the quality of the reconstruction; we have utilized the same approach in our analysis here. We consider only one type of structural imperfection, namely a type of distortion of the crystal lattice caused by correlated displacements of the molecular clusters from their ideal positions [equation (12b)]. We used this type of distortion as an illustrative example and as a tool to explore the generic properties of the approach.

5. Simulations

The protein molecule photosystem I (Jordan *et al.*, 2001) was used as a target for the structure reconstruction. We simulated both ideal and disordered nanocrystals (~ 200 nm linear size) of photosystem I containing $7 \times 7 \times 12$ unit cells. Each unit cell consists of six protein molecules (the molecular cluster) according to the space group $P6_3$.

The electron-density distribution of the crystal, $\rho_C(\mathbf{r})$, can be represented by the convolution

$$\rho_C(\mathbf{r}) = \rho_M(\mathbf{r}) \otimes O(\mathbf{r}), \quad (21)$$

where $\rho_M(\mathbf{r})$ is the electron density of the molecular cluster and $O(\mathbf{r})$ is the lattice function defined by Vainshtein (1966),

$$O(\mathbf{r}) = L\delta(\mathbf{r} - \mathbf{R}_0) + \sum_{l=1}^L p_l \delta(\mathbf{r} - \mathbf{R}_l). \quad (22)$$

The summation is performed over all L lattice translations (unit cells) of the crystal. In equation (22), \mathbf{R}_l denotes the l th translational position of the three-dimensional lattice, p_l is the occupancy factor and δ is the Dirac delta function; the occupancy factor equals unity when the position is fully occupied by the molecular cluster and zero if the position is vacant. Starting with the central position, $\mathbf{R}_0 = (0, 0, 0)$, all subsequent translational positions, \mathbf{R}_l , of the molecular cluster are $\mathbf{R}_l = \mathbf{R}_{l-1} + \mathbf{H} + \Delta\mathbf{H}$, where $\mathbf{H} = (\mathbf{a}, \mathbf{b}, \mathbf{c})$ is the elementary translation vector and $\Delta\mathbf{H}$ denotes the displacement of the molecular cluster from the ideal position; for the ideal crystal $\Delta\mathbf{H} = (0, 0, 0)$. To simulate the partially disordered nanocrystal, a set of 100 random vectors, $\{\Delta\mathbf{H}\}$, is generated for each position \mathbf{R}_l using a Gaussian distribution with a mean of $\mathbf{R}_l = \mathbf{R}_{l-1} + \mathbf{H}$, from which one displacement vector $\Delta\mathbf{H}$ is randomly selected. The number of unit cells, L_m , for each crystal was randomly generated, with a Gaussian distribution corresponding to a mean length of 200 nm and with a standard deviation of 10%. The three-dimensional intensity distribution volume was calculated as follows; the MFF function, $\Phi(\mathbf{q})$, was calculated using

$$\Phi(\mathbf{q}) = \sum_{j=1}^N f_j(q) \sum_{k=1}^K \exp\{i2\pi[\mathbf{q} \cdot (\hat{U}_k \mathbf{u}_j)]\}, \quad (23)$$

where N is the number of atoms in the protein molecule, K is the number of symmetry operations, $f_j(q)$ is the atomic scattering factor of the j th atom, \mathbf{u}_j is the position of the j th atom and \hat{U}_k is the k th symmetry operation matrix. The interference function for each crystal, $\Lambda_m(\mathbf{q})$, was calculated for each nanocrystal using

$$\Lambda_m(\mathbf{q}) = \hat{F}[O_m(\mathbf{r}) \otimes O_m(\mathbf{r})], \quad (24)$$

where \otimes denotes the convolution operator. The intensity scattered from the crystal was calculated using

$$\begin{aligned} I_m(\mathbf{q}) &= \left| \hat{F}[\rho_C(\mathbf{r})] \right|^2 = \left| \hat{F}[\rho_M(\mathbf{r})] \hat{F}[O_m(\mathbf{r})] \right|^2 \\ &= |\Phi(\mathbf{q})|^2 \Lambda_m(\mathbf{q}). \end{aligned} \quad (25)$$

Finally, the total intensity was calculated as

$$I_{\text{exp}}(\mathbf{q}) = \sum_{n=1}^M I_n(\mathbf{q}), \quad (26)$$

where M is the total number of three-dimensional diffraction patterns simulated. The two-dimensional cross section of the three-dimensional intensity distribution, $I_{\text{exp}}(\mathbf{q})$, for the ideal and imperfect photosystem I nanocrystals is shown in Fig. 6.

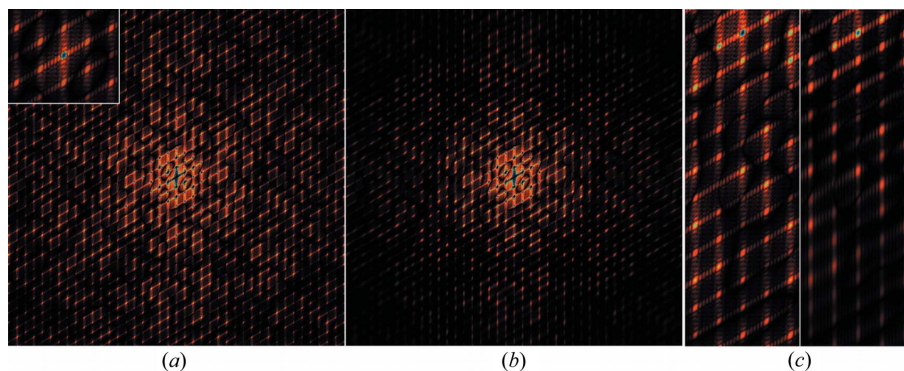


Figure 6
 Simulated diffraction patterns of the photosystem I nanocrystal [(*ab*) plane]: (a) diffraction from the ideal crystal; (b) diffraction from the set of disordered nanocrystals. (c) A magnified view of a segment from the diffraction patterns of the ideal (left) and the disordered (right) nanocrystals.

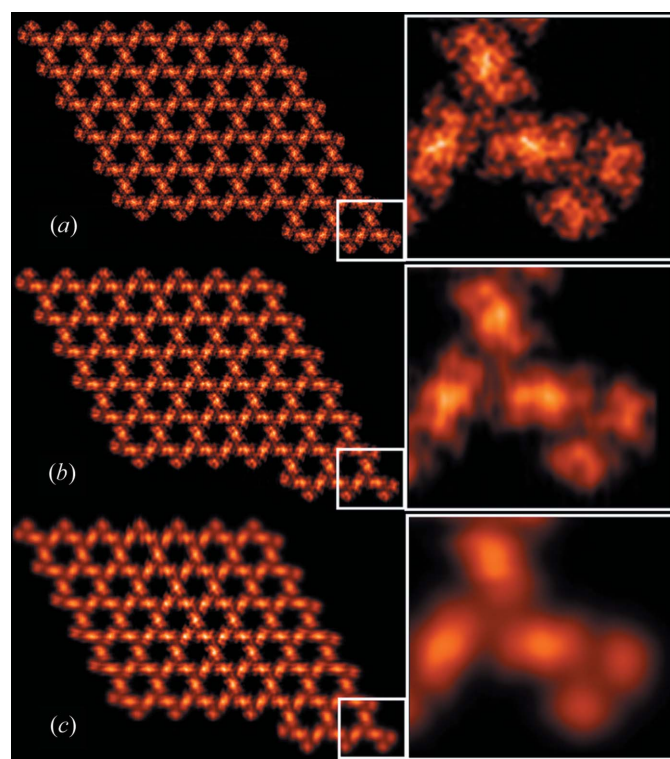


Figure 7
 The (001) projection of the electron density for the ideal (a) and the disordered (b), (c) photosystem I protein crystals, reconstructed at a resolution of 12 Å without incorporation of the partial coherency model into the reconstruction process.

6. Reconstruction

Elements of the $A(\mathbf{q})$ matrix, presented in equation (19b), can be rewritten in the equivalent form

$$A_{jk}(\mathbf{q}) = \frac{1}{M} \sum_{m=1}^M \int_{-\infty}^{\infty} P_{jk}^m(\mathbf{R}) G(\Delta_{jk}) \exp(i\mathbf{q}\Delta_{jk}^m) d\Delta_{jk}^m, \quad (27)$$

where $\Delta_{jk}^m = \Delta\mathbf{r}_{jm} - \Delta\mathbf{r}_{km}$, $G(\Delta_{jk})$ is the distribution function which describes the lattice distortion and $P_{jk}^m(\mathbf{R}) = p^m(\mathbf{R}_j)p^m(\mathbf{R}_k)$ is the crystal form distribution function, which is the self-convolution of the crystal shape function (§4).

Consequently, equation (27) can be written in the form (Vainshtein, 1966)

$$A_{jk}(\mathbf{q}) = \langle W(\mathbf{q}) \rangle \otimes g_{jk}(\mathbf{q}), \quad (28)$$

where $\langle W(\mathbf{q}) \rangle = \hat{F}[\langle P(\mathbf{R}) \rangle]$ is the average CFF, \hat{F} denotes the Fourier transform operator and $g_{jk}(\mathbf{q})$ is defined by equation (12b). The mean-square displacement is defined in §3.

In order to acquire information about the average shape and dimensions of the crystal, we first applied the conventional CDI algorithm to reconstruct its image assuming that $g_{jk}(\mathbf{q}) = 1$ for all j and k . Fig. 7 shows the reconstructed images in the (001) planes for both the ideal (Fig. 7a) and disordered protein nanocrystals (Figs. 7b, 7c), with a resolution of 12 Å.

Since the conventional CDI algorithm does not take into account the imperfections of the crystal structure, an excellent quality image of the ideal crystal is reconstructed, while the image of the imperfect crystal shows a pronounced blurring effect. Nevertheless, the shape and the dimensions of the crystal projection can be correctly identified from Figs. 7(b), 7(c). Considering the diversity of individual nanocrystals, however, the average form of the nanocrystal defines the $\langle W(\mathbf{q}) \rangle$ function for further CDI reconstructions.

We also generated two data sets of integrated intensities of the Bragg reflections for ideal, $I_i(hkl)$, and disordered, $I_d(hkl)$, protein nanocrystals, using the intensity integration approach presented in Kirian *et al.* (2010), to analyse them using conventional crystallography. The crystal disorder was modelled by the static Debye–Waller factor, $\exp[-B(\sin\theta/\lambda)^2]$, where $B = 40.0 \text{ \AA}^2$. We first conducted a molecular replacement with PHASER (McCoy *et al.*, 2007) using the published photosystem I structure (Protein Data Bank ID 1jb0) as an initial model against the generated structure factors, $F_i(hkl) = [I_i(hkl)]^{1/2}$ and $F_d(hkl) = [I_d(hkl)]^{1/2}$, respectively.

In both cases, the analysis showed the correct position and orientation of the photosystem I molecule in the unit cell. This is the expected result, since the position, orientation and shape of the molecule can be determined solely from the low- q data because the high- q data are most sensitive to the structural imperfections. Subsequent constrained structural refinement with REFMAC (Murshudov *et al.*, 1997) failed, however, with extremely high thermal parameters in the range of 40–120 Å² for most of the atoms and over 100% R/R_{free} factors for the disordered data set (Table 3). This is also in good agreement with the results obtained using the conventional CDI algorithm.

Finally, we reconstructed the electron density of the molecular clusters by incorporating models of structural imperfections into the reconstruction process. The common ‘recipe’ adopted for the CDI reconstruction using the partially coherent models has been described in several publications

Table 3

Refinement statistics.

	Ideal	Disordered
Resolution range (Å)	243.35–2.96	
Number of reflections	25338	
Data cutoff, $\sigma(F)$	none	none
R/R_{free} † factor (%)	18.8/20.2	>100
Average B value for all atoms (Å ²)	6.1 (3)	64.5 (2)

† Represents 5% of the data.

(Whitehead *et al.*, 2009; Chen *et al.*, 2009; Abbey *et al.*, 2011; Quiney & Nugent, 2011). For structure analysis of the partially disordered protein nanocrystals we devised the following iterative procedure:

(i) The reconstruction process starts from the initial guess of the uniform distribution of the electron density in the unit cell, $\rho_M(\mathbf{r}) = Z/V$, where V is the unit-cell volume and Z is the total number of electrons in the unit cell. We use the unit cell of the crystal as the initial support function, $Su(\mathbf{r})$. Later, during the reconstruction process, the low-resolution envelope of the molecular cluster is used as a support.

(ii) The MFF function is calculated using $\Phi(\mathbf{q}) = \hat{F}[\rho_M(\mathbf{r})]$.

(iii) The set of complex functions $T_j(\mathbf{q})$ is calculated using $T_j(\mathbf{q}) = \Phi(\mathbf{q}) \exp(i\mathbf{q} \cdot \mathbf{R}_j)$ for all $\mathbf{R}_j \in \Omega$, and the $A(\mathbf{q})$ matrix is calculated using equations (28) and (12b).

(iv) The integrated intensity, $I_T(\mathbf{q})$, is calculated using equation (19a).

(v) Parameters $\alpha_{1,\dots,3}$ are optimized by minimizing the error cost function, $E = \sum [I_{\text{exp}}(\mathbf{q}) - I_T(\mathbf{q})]^2$, using the steepest descent method, where $I_{\text{exp}}(\mathbf{q})$ is the measured (simulated) intensity. We used $\alpha_1 = \alpha_2 = 0.002$ and $\alpha_3 = 0.001$ as the initial values for $\alpha_{1,\dots,3}$.

(vi) The modulus constraint is imposed using $T'_j(\mathbf{q}) = T_j(\mathbf{q})[I_{\text{exp}}(\mathbf{q})]^{1/2}/[I_T(\mathbf{q})]^{1/2}$.

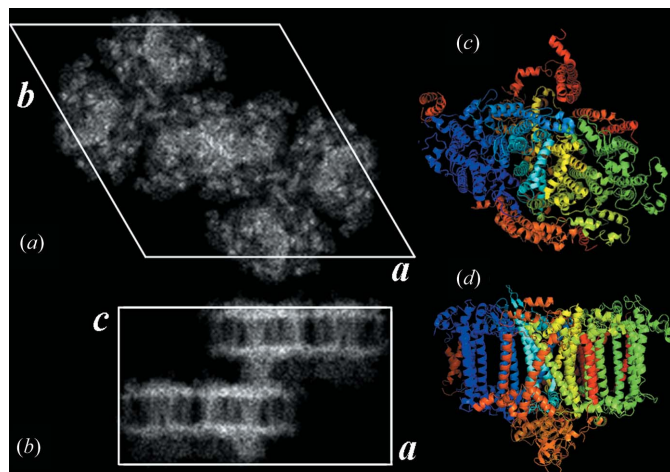


Figure 8

The projections of the electron density of the photosystem I molecular cluster, reconstructed with a resolution of 4.1 Å by incorporating models of the partial coherency: (a) (ab) crystallographic plane, (b) (ac) crystallographic plane. (c) and (d) models of the photosystem I molecule, (ab) and (ac) projections, respectively.

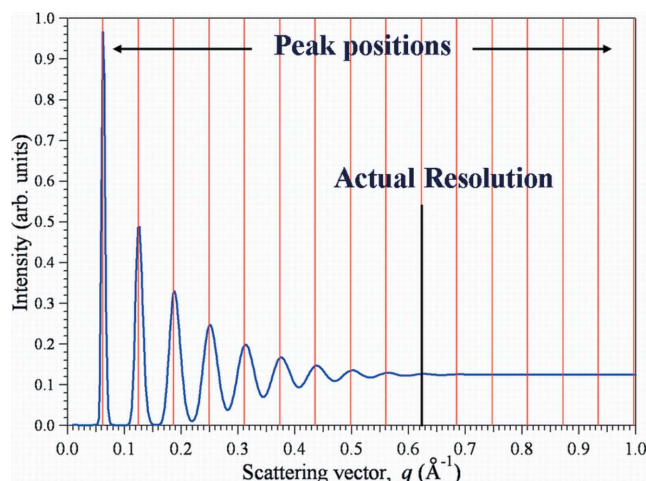


Figure 9

The one-dimensional interference function calculated for the nanocrystal with structural disordering.

(vii) With the amplitude updated, the complex functions, $T'_{j=0}(\mathbf{q})$, are transformed into real space and the support function $Su(\mathbf{r})$ is applied to obtain the updated electron density of the molecular cluster distributed in the unit cell of the crystal, $\rho_M(\mathbf{r}) = \hat{F}^{-1}[T'_{j=0}(\mathbf{q})]Su(\mathbf{r})$.

Steps (ii)–(vii) are repeated and the progress of the reconstruction monitored using an error metric defined as

$$R = \frac{\sum [I_{\text{exp}}(\mathbf{q}) - KI_T(\mathbf{q})]^2}{\sum [I_{\text{exp}}(\mathbf{q})]^2}, \quad (29)$$

where K is the scale factor.

Fig. 8 shows the electron density of the photosystem I molecular clusters reconstructed at a resolution of 4.1 Å. The resulting images clearly show structural details, including the envelopes of photosystem I protein molecules and transmembrane α -helices. As one can see from Fig. 8, the method enables reconstruction to be obtained from data that are intractable to conventional crystallographic analysis.

7. The resolution limit

The broadening of diffraction peaks due to imperfections of protein nanocrystals determines the resolution achievable in the reconstructed image. The resolution limit is defined by the q_{max} value corresponding to the largest measured angle for which the diffraction pattern provides prominent reflections. If the Bragg peak becomes so broad that its half-width is greater than $\Delta\mathbf{q} = D$, where D is the distance between the pair of nearest-neighbour Bragg reflections, and overlaps significantly with its neighbours, then the interference function $\hat{\Lambda}(\mathbf{q})$ becomes almost constant (Fig. 9). This defines the coherence length of the partially disordered nanocrystal that is the boundary between the crystal-type scattering (molecular clusters are scattering coherently) and ‘gas’-type scattering (molecular clusters are scattering incoherently). For the

classes of structural imperfection considered here, the coherence length, L_C , is defined by Vainshtein (1966),

$$L_C = m_{\max} d, \quad (30)$$

where d is the distance between two nearest molecular clusters and $m_{\max} = 1/(2.5\Delta_1/d)^2$. According to equation (30), L_C is governed by the degree of disorder, Δ_1/d . The 'gas'-type scattering for which the $\tilde{\Lambda}(\mathbf{q})$ function exhibits no maximum occurs if the degree of disorder is given by $\Delta_1/d > 0.25$ (Vainshtein, 1966).

8. Conclusion

Considering the diffraction pattern of the protein nanocrystal as a continuous function of the scattering vector, \mathbf{q} , rather than as a discrete set of Bragg reflections, yields a more complete extraction of structural information from diffraction data in the structural analysis of partially disordered nanocrystals. Our approach is particularly suited to crystals obtained by *in vivo* crystallization as this intrinsically leads to the formation of partially disordered crystals (Hempelmann & Marques, 1994; Wolf *et al.*, 1999; Frenkiel-Krispin & Minsky, 2002). The boundary between the continuous and discrete approaches to structural analysis of diffraction by protein nanocrystals is determined primarily by the ratio N_S/N_V (§1). The conventional approach can be applied if the fraction of unit cells adjacent to the surface of the crystal does not exceed 10% of the total number of unit cells, suggesting an effective size limit of 1 μm for a protein crystal with an average unit-cell parameter of 100 Å.

We have shown that structural imperfections, such as mutual displacement or misalignment of proteins, particularly at the surface of the nanocrystals, as well as the shape and dimensions of the nanocrystals, play an important role in the formation of a diffraction pattern. It has also been shown that an algorithm that accommodates these structural imperfections is able to extract information in cases for which existing conventional methods fail. The analysis proposed here, based on continuous diffraction patterns, offers a unique solution for the structure without the need for additional assumptions or additional data. This may lead to the determination of molecular structures without the need to introduce molecular replacement strategies which introduce a natural bias in favour of existing trial structures. Such an approach is in the direction of truly *ab initio* structure determination, which may be made possible by XFEL sources.

The authors acknowledge the support of the Australian Research Council through its Centres of Excellence and Federation Fellowship programmes, and from the CSIRO Preventative Health Flagship, Neurodegenerative Diseases Theme.

References

Abbey, B., Lachlan, W., Whitehead, H. M., Quiney, D. J., Vine, G. A., Cadenazzi, C. A., Henderson, K. A., Nugent, E., Balaur, C. T.,

- Putkunz, A. G., Peele, G., Williams, J. & McNulty, I. (2011). *Nat. Photonics*, **5**, 420–424.
- Baker, M. (2010). *Nat. Methods*, **7**, 429–434.
- Bates, R. H. T. (1982). *Optik*, **61**, 247–262.
- Boutet, S. *et al.* (2012). *Science*, **337**, 362–364.
- Caffrey, M. (2003). *J. Struct. Biol.* **142**, 108–132.
- Chapman, H. N., Barty, A., Marchesini, S., Noy, A., Hau-Riege, S. P., Cui, C., Howells, M. R., Rosen, R., He, H., Spence, J. C. H., Weierstall, U., Beetz, T., Jacobsen, C. & Shapiro, D. (2006). *J. Opt. Soc. Am. A*, **23**, 1179–1200.
- Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
- Chen, B., Dilanian, R. A., Teichmann, S., Abbey, B., Peele, A. G., Williams, G. J., Hannaford, P., Dao, L. V., Quiney, H. M. & Nugent, K. A. (2009). *Phys. Rev. A*, **79**, 023809.
- Eyal, E., Gerzon, S., Potapov, V., Edelman, M. & Sobolev, V. (2005). *J. Mol. Biol.* **351**, 431–442.
- Faure, P. *et al.* (1994). *Nat. Struct. Biol.* **1**, 124–128.
- Feher, G. & Kam, Z. (1985). *Methods Enzymol.* **114**, 77–112.
- Fienup, J. R. (1982). *Appl. Opt.* **21**, 2758–2769.
- Frenkiel-Krispin, D. & Minsky, A. (2002). *ASM News*, **68**, 277–283.
- Fromme, P. & Spence, J. C. H. (2011). *Curr. Opin. Struct. Biol.* **21**, 509–516.
- Fung, R., Shneerson, V., Saldin, D. K. & Ourmazd, A. (2009). *Nat. Phys.* **5**, 64–67.
- Garcia-Ruiz, J. M. (2003). *J. Struct. Biol.* **142**, 22–31.
- Gerchberg, R. W. & Saxton, W. O. (1972). *Optik*, **34**, 275–284.
- Giacovazzo, C. (2011). Editor. *Fundamentals of Crystallography*, 3rd ed. Oxford University Press.
- Grant, M. L. & Saville, D. A. (1994). *J. Phys. Chem.* **98**, 10358–10367.
- Hempelmann, E. & Marques, H. M. (1994). *J. Pharmacol. Toxicol. Methods*, **32**, 25–30.
- Johansson, L. C. *et al.* (2012). *Nat. Methods*, **9**, 263–265.
- Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W. & Krauss, N. (2001). *Nature (London)*, **411**, 909–917.
- Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., Spence, J. C. H., Hunter, M., Fromme, P., White, T., Chapman, H. N. & Holton, J. (2010). *Opt. Express*, **18**, 5713–5723.
- Kirian, R. A., White, T. A., Holton, J. M., Chapman, H. N., Fromme, P., Barty, A., Lomb, L., Aquila, A., Maia, F. R. N. C., Martin, A. V., Fromme, R., Wang, X., Hunter, M. S., Schmidt, K. E. & Spence, J. C. H. (2011). *Acta Cryst. A* **67**, 131–140.
- Koopmann, R. *et al.* (2012). *Nat. Methods*, **9**, 259–262.
- Loh, N. D. & Elser, V. (2009). *Phys. Rev. E*, **80**, 026705.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Malkin, A. J. & Thorne, R. E. (2004). *Methods*, **34**, 273–299.
- Miao, J., Sayre, D. & Chapman, H. M. (1998). *J. Opt. Soc. Am. A*, **15**, 1662–1669.
- Mizuguchi, K., Kidera, A. & Go, N. (1994). *Proteins*, **18**, 34–48.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst. D* **53**, 240–255.
- Quiney, H. M. (2010). *J. Mod. Opt.* **57**, 1109–1149.
- Quiney, H. M. & Nugent, K. A. (2011). *Nat. Phys.* **7**, 142–146.
- Rafaja, D., Klemm, V., Schriber, G., Knapp, M. & Kužel, R. (2000). *Phys. Rev. B*, **61**, 16144–16153.
- Rafaja, D., Klemm, V., Schreiber, G., Knapp, M. & Kužel, R. (2004). *J. Appl. Cryst.* **37**, 613–620.
- Schwander, P., Fung, R., Phillips, G. N. Jr & Ourmazd, A. (2010). *New J. Phys.* **12**, 035007.
- Shneerson, V. L., Ourmazd, A. & Saldin, D. K. (2008). *Acta Cryst. A* **64**, 303–315.
- Spence, J. C. H. (2004). *Science of Microscopy*, edited by P. W. Hawkes & J. C. H. Spence. New York: Springer.
- Spence, J. C. H., Kirian, R. A., Wang, X., Weierstall, U., Schmidt, K. E., White, T., Barty, A., Chapman, H. N., Marchesini, S. & Holton, J. (2011). *Opt. Express*, **19**, 2866–2873.

- Tokuhisa, A., Taka, J., Kono, H. & Go, N. (2012). *Acta Cryst.* **A68**, 366–381.
- Vainshtein, B. K. (1966). *Diffraction of X-rays by Chain Molecules*. New York: American Elsevier Publishing Co., Inc.
- Welberry, T. R. (2004). *Diffuse X-ray Scattering and Models of Disorder*. Oxford University Press.
- Welberry, T. R., Heerdegen, A. P., Goldstone, D. C. & Taylor, I. A. (2011). *Acta Cryst.* **B67**, 516–524.
- Whitehead, L. W., Williams, G. J., Quiney, H. M., Vine, D. J., Dilanian, R. A., Flewett, S., Nugent, K. A., Peele, A. G., Balaur, E. & McNulty, I. (2009). *Phys. Rev. Lett.* **103**, 243902.
- Williams, G., Pfeifer, M., Vartanyants, I. & Robinson, I. (2007). *Acta Cryst.* **A63**, 36–42.
- Wolf, E. (1982). *J. Opt. Soc. Am.* **72**, 343–351.
- Wolf, S. G., Frenkiel, D., Arad, T., Finkel, S. E., Kolter, R. & Minsky, A. (1999). *Nature (London)*, **400**, 83–85.