research papers

Coherent X-ray diffractive imaging of protein crystals

Sébastien Boutet[a,b]* and Ian K. Robinson[a,c]

[a]Department of Physics, University of Illinois, Urbana, IL 61801, USA, [b]Linac Coherent Light Source, Stanford Linear Accelerator Center, Menlo Park, CA 94025, USA, and [c]London Centre for Nanotechnology, University College, London, England. E-mail: sboutet@slac.stanford.edu

The technique of coherent X-ray diffraction imaging (CXDI) has recently shown great promise for the study of inorganic nanocrystals. In this work the CXDI method has been applied to the study of micrometer-size protein crystals. Finely sampled diffraction patterns of single crystals were measured and iterative phase-retrieval algorithms were used to reconstruct the two-dimensional shape of the crystal. The density maps have limited reproducibility because of radiation damage, but show clear evidence for crystal facets. Qualitative analysis of a number of single-crystal diffraction peaks indicates the presence of inward surface contraction on 2 μm size crystals. A survey of several hundred diffraction patterns yielded a number of examples with dramatic single-sided streaks, for which a plausible model is constructed.

Keywords: coherent X-ray diffractive imaging; protein crystals; ferritin; phase retrieval; surface contraction.

1. Acronyms, abbreviations and conventions

f.c.c.: face-centered cubic.
CXD: coherent X-ray diffraction.
CXDI: coherent X-ray diffraction imaging.
(111): Miller indices corresponding to the body diagonal direction in the reciprocal cubic lattice.
{111}: family of all (111) Bragg peaks.
Gy: unit of absorbed dose (Gray = J kg^{-1}), deposited energy per kg.

2. Introduction

One of the great promises of present and future third-generation sources of synchrotron radiation is their thousand-fold increase in coherent flux provided by the strategic use of undulator technology. CXD is a new branch of science that has been opened up by this source development, with important applications in X-ray photon correlation spectroscopy as well as imaging. The prospect of CXDI being applied to a wide range of sample materials has motivated planning of new imaging beamlines at the latest sources.

It is important to establish how far the method of CXDI can be developed for more radiation-sensitive materials, such as biological samples. There are well established limits of radiation dose and the corresponding image resolution that would be expected (Howells *et al.*, 2005) as well as widespread speculation that the limits can be overcome by using (fourth-generation) short-pulse single-shot diffraction methods (Neutze *et al.*, 2000).

The preparation of protein crystals represents an important technological step along the way to solving the atomic structures of protein molecules. The limitations of radiation dose have been studied extensively (Nave & Garman, 2005; Nave & Hill, 2005; Ravelli *et al.*, 2002). In one earlier example the radiation was shown to cause an outward 'explosion' of the Bragg peaks that was thought to correspond to dehydration of the solvent channels within the crystal, which would explain why the lattice contracts (Boutet & Robinson, 2006). Crystallization of proteins still remains as one of the biggest limitations to the future advance of the field of macromolecular crystallography. Imaging of crystals during the initial nucleation phase would be an excellent way of learning about how to control them and prepare them better. Since the submicrometer sizes involved are beyond the limit of visible-light microscopy, shorter wavelengths are needed. In this paper we show that CXDI can indeed be used to image small protein crystals, but is subject to all the limitations of sample handling and radiation damage that arise along the way.

3. Coherent X-ray diffraction

In a typical X-ray diffraction experiment the size of the X-ray beam is often smaller than the sample. Also, the phase of the wavefield is generally not well defined over the entire wavefront. We define the coherence length of the beam as the distance in a certain direction over which the relative phase of the ensemble of electric fields varies by a value of π. There are three coherence lengths to consider, two of which are transverse and the other being longitudinal to the direction of

propagation of the X-rays. The transverse coherence lengths depend on the source size and the distance from the source while the longitudinal coherence length depends on the level of monochromaticity of the beam (Als-Nielsen & McMorrow, 2001). For the undulator beamline used at the Advanced Photon Source the typical coherence lengths are of the order of a few micrometers horizontally and a few tens of micrometers vertically. The longitudinal coherence length using a Si(111) double-crystal monochromator is roughly 1 µm.

The three orthogonal lengths define a coherence volume of the X-ray beam no more than a few micrometers across. This is much smaller than a typical protein crystal used in diffraction experiments and also much smaller than a typical X-ray beam. This implies that the diffraction pattern measured on a large crystal is representative of a structure averaged over multiple incoherently scattering volumes. The scattered intensity from a single coherence volume is given by

$$I(\mathbf{q}) \propto \left| \int \rho_{\mathrm{inf}}(\mathbf{r}) \exp(i\,\mathbf{q}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} \right|^2, \tag{1}$$

where $\rho_{\mathrm{inf}}(\mathbf{r})$ is the electron density of the infinite crystal, $\mathbf{q} = 2\pi/\lambda$ is the momentum transfer vector, λ is the wavelength of the radiation, $\mathbf{r}$ is the real-space position vector and the integral is performed over the coherence volume. If multiple coherence volumes are present in the beam, the intensities from each of these adds incoherently since the phases of the beams being scattered by different coherence volumes are uncorrelated.

In the case of a crystal smaller than the coherence volume of the beam the situation is quite different. The electron density of the small crystal is no longer an infinite sum of sharp localized functions as in the case of a large crystal. The electron density can be written as

$$\rho(\mathbf{r}) = s(\mathbf{r})\rho_{\mathrm{inf}}(\mathbf{r}), \tag{2}$$

where $s(\mathbf{r})$ is the shape function of the small crystal with a value of 1 inside and 0 outside the crystal. Recognizing that the scattered amplitude is a Fourier transform of the electron density, we can use the convolution theorem to calculate the intensity,

$$
\begin{aligned}
I(\mathbf{q}) &\propto \left| \int \rho(\mathbf{r}) \exp(i\,\mathbf{q}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} \right|^2 \\
&\propto \left| \int s(\mathbf{r})\rho_{\mathrm{inf}}(\mathbf{r}) \exp(i\,\mathbf{q}\cdot\mathbf{r})\,\mathrm{d}\mathbf{r} \right|^2 \\
&\propto \left| S(\mathbf{q}) \otimes F(\mathbf{q}) \right|^2.
\end{aligned}
\tag{3}
$$

In (3) the Fourier transform of $s(\mathbf{r})$ was written as $S(\mathbf{q})$ while the Fourier transform of the infinite crystal $\rho_{\mathrm{inf}}(r)$ was written as $F(\mathbf{q})$ and is known as the scattering form factor.

The physical implication of (3) is that the scattered intensity from a coherently illuminated small crystal is not simply a collection of delta-function-like Bragg peaks but contains information about the entire shape of the crystal. By measuring the function $S(\mathbf{q})$, which is repeating around each Bragg peak, one can in theory deduce the three-dimensional shape of the crystal.

Equation (3) was derived for a perfect crystal. A small protein crystal is unlikely to be perfect. In this situation it is possible to show that the density of a crystal can be written in terms of a complex function where the phase is related to the distortions of the local lattice while the amplitude can be interpreted as the density (Vartanyants & Robinson, 2001). Therefore, by allowing the shape function in (3) to be complex, one can use the technique to map the strain fields inside a nanocrystal (Pfeifer *et al.*, 2006).

The phase of the scattered waves is lost during the measurement because only the intensity is measured. Phase-retrieval algorithms based on the Gershberg–Saxton algorithm can be used to recover the phases (Gerchberg & Saxton, 1972; Robinson *et al.*, 2001; Fienup, 1982; Sayre *et al.*, 1998; Miao *et al.*, 1999; Marchesini, 2007). These algorithms rely on the assumption that the diffraction pattern is sufficiently oversampled so that all the spatial frequencies present in the object, from the size of the object down to the highest spatial resolution measured, are sampled at least at twice the Nyquist frequency. The oversampling by at least a factor of two implies that there are a sufficient number of data points to over-determine the unknown variables, *i.e.* the phases (Sayre, 1952; Miao *et al.*, 1999). With the phases known, calculating the inverse Fourier transform of $S(\mathbf{q})$ directly yields the complete structure of the crystal. The theoretical resolution limit of this technique is half the wavelength of the radiation, roughly 0.5 Å for hard X-rays. In practice, however, the resolution currently achievable is of the order of tens of nanometers and is limited by signal level and detector size. Nevertheless, this technique can be very useful in that it allows imaging in three dimensions of submicrometer crystals which cannot be imaged by optical methods.

4. Sample and experimental set-up

The coherent diffraction beamline at sector 34-ID-C at the Advanced Photon Source at Argonne National Laboratory was used to perform the measurements described here. The 9 keV radiation produced by the undulator was unfocused and a narrow energy band was selected using a double-crystal Si(111) monochromator. Two sets of slits at 3 m and 1 m upstream of the sample were used to collimate the beam and reduce it to a size of 200 µm × 200 µm.

Small crystals of the protein holoferritin from horse spleen were studied. The protein molecules in solution were purchased from the Sigma-Aldrich Company at a concentration of 100 mg ml^{-1} in 150 mM NaCl. The holoferritin molecule consists of a roughly spherical shell with inner diameter ~80 Å and outer diameter ~130 Å. The inside of this shell is filled with as many as 4500 iron atoms. *In vivo*, the iron atoms are stored inside the molecules until they are needed by another part of the host organism (Harrison & Arosio, 1996). Holoferritin was chosen for two primary reasons. First, it crystallizes very easily and it was therefore relatively easy to produce a shower of micrometer-sized crystals. The second reason is that the presence of the iron cores increases the diffracted intensity from the small nanocrystals.

Crystals were grown in solution containing 50 mg ml^{-1} of the protein in 150 mM NaCl and 22.5 mM CdCl$_2$. The intro-

duction of Cd^{2+} ions in solutions of ferritin has previously been found to lead to crystallization, the most common form being f.c.c. (Granier *et al.*, 1997). The large amount of cadmium added in this current case leads to rapid nucleation and therefore a very large number of small crystals in a short amount of time. It was necessary to produce a large amount of crystals owing to the inability to select and orient a single micrometer-size crystal in the X-ray beam. By increasing the density of the crystals, the probability of measuring a single Bragg peak from a single crystal was increased so the experiment could be performed in a reasonable amount of time.

The crystals in solution were illuminated by the beam and a direct-illumination CCD camera from Princeton Instruments with 1340 × 1300 pixels of 20 µm × 20 µm was placed 2.2 m away from the sample at an angle corresponding to the (111) Bragg peak of the ferritin crystal at the X-ray energy used. The (111) Bragg peak is the lowest angle reflection from an f.c.c. crystal. This angle is small ($\sim 0.7°$) and therefore the small camera chip (2.5 cm × 2.5 cm) still covers a fairly wide range of azimuthal angles. The scattering from a sample with crystals freely diffusing in solution illuminated for a long period of time or alternatively a sample containing a very large amount of crystals was a continuous ring of intensity or a powder pattern. The amount of cadmium was adjusted so that the number of crystals diffracting at any given time was small enough that these peaks were generally well separated and not overlapping.

5. Data

There are multiple technical difficulties in performing coherent diffractive imaging measurements on micrometer-size crystals of proteins which are suspended in solution. The thermal motions in the sample as well as the damage to the sample caused by the X-ray beam limit the accumulation time for a single crystal. Long exposures are required to measure the weak fine features near the intense Bragg peak. However, it was found that after a short period of time, which is generally shorter than the time that would be required to adequately perform the measurement, the sample had changed significantly and often to the point that a Bragg peak was no longer observed (Boutet & Robinson, 2006). Also, rapid tumbling or angular motions of the crystal led to an averaging of many orientations which would smear the coherent diffraction features. Despite all the technical difficulties limiting the effectiveness of the technique, some CXD patterns displaying expected features such as intensity fringes were measured and they are presented here.

A sample was prepared with 4 µl of holoferritin at 100 mg ml^{-1} into which 4 µl of 45 mM CdCl$_2$ were added. Crystals were formed immediately (faster than the time required to exit and close the experimental hutch) and data were collected immediately after the experimental hutch could be closed. Bragg peaks coming from individual small crystals were observed and the diffraction pattern of a single (111) peak was selected. These Bragg peaks were seen for a
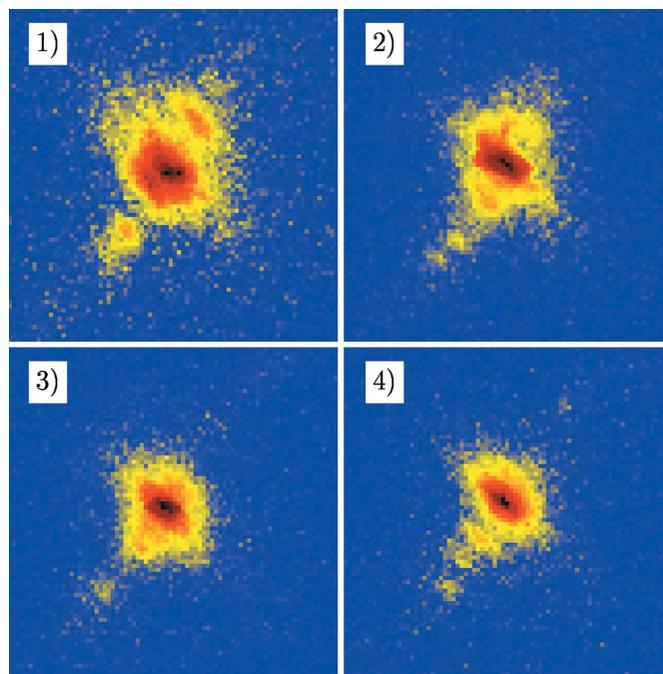
Figure 1
Intensity on a logarithmic scale of a single Bragg peak from a ferritin crystal over time. Each image corresponds to an exposure of 60 s.

few exposures of 60 s each, became progressively weaker and eventually disappeared completely. The first four diffraction images obtained from a particular crystal are shown in Fig. 1. Intensity flares with modulations, which are expected for faceted crystals, can be seen. The patterns are seen to change in time which is expected from the thermal motions of the crystal in solution and the radiation damage from the X-ray beam. These CXD patterns are not particularly well resolved when compared with the ones obtained on gold or lead nanocrystals (Robinson *et al.*, 2001; Williams *et al.*, 2003; Pfeifer *et al.*, 2006) due to the fact that the protein crystals were in solution and suffered from radiation damage. However, these patterns do represent, to our knowledge, the first coherent X-ray diffraction patterns from single nanocrystals of proteins.

The CXD pattern, which corresponds to the Fourier transform of the shape of the crystal, should be centrosymmetric if we assume the particle has a real electron density and to be a perfect crystal with exact translational symmetry. The patterns measured are clearly not centrosymmetric. There are two possible reasons for this. The first one is that the three-dimensional pattern is centrosymmetric but a random two-dimensional slice through the pattern does not necessarily have to be symmetric; only a slice through the exact center of the pattern would be symmetric. There is no reason here to expect the slice seen in a given exposure to pass through the center since the crystal has a random orientation. This is because the two-dimensional slice seen is where the diffraction pattern intersects the Ewald sphere. The random orientation of the Ewald sphere means that the center of any given Bragg peak does not necessarily lie on the surface of it.

Therefore, each of these slices is the Fourier transform of a complex projection of the electron density onto the detector plane.

Another possible reason for the absence of centrosymmetry could be the presence of strain in the crystal. Strain would show up as an imaginary part to the electron density in a reconstruction (Pfeifer *et al.*, 2006). The quality of the data does not allow us to accurately reconstruct a strain field for the crystal. This was attempted but to reconstruct the phase as well as the amplitude of the electron density requires twice the number of undetermined variables and led to unstable phase retrieval and multiple solutions that were not reproducible. Whichever origin of the asymmetry applies, to obtain a reconstruction of the projection of the shape of the crystal the patterns were symmetrized about the center. This assumes that the diffraction pattern does not represent a slice through the middle of the Bragg peak and that the measured fringes on one side are missing from the other side only due to misalignment. This assumption is questionable in this present case since the asymmetries appear to be real as will be discussed below. The symmetrization was carried out by determining the most symmetric point in the diffraction pattern in the following manner. We picked a center point in the pattern and then inverted the pattern around this point. We then compared the original pattern with the inverted one and repeated the operation for many center points. The ideal center of the pattern was selected by maximizing the correlation between the two patterns. This has the effect of removing the imaginary part of the electron density and it allowed the algorithms to converge to a reproducible solution. This is shown in Fig. 2 where the same four images have been symmetrized. The subject of strain will be revisited in §7.

The symmetrized patterns were then put through a series of algorithms to retrieve the phases of the scattered waves in order to reconstruct the real two-dimensional projection of the shape of the crystal.

6. Reconstruction

A series of programs to treat the data obtained with the CCD detector has been developed and was used directly to convert the data into a format that the phase-retrieval routines could handle.

The scattered amplitude is calculated by taking the square root of the intensity. A background is subtracted from the array. The data are then padded to avoid aliasing and to decrease the real-space steps of the reconstruction. This is done by embedding the array at the center of a larger array of zeroes. This only makes the real-space pixels smaller but does not increase the resolution of the reconstructed image since this is determined by the highest angle at which signal is observed higher than the noise.

The phase-retrieval program is based on the iterative method of Gershberg & Saxton (1972). The method iterates between real and reciprocal space using fast Fourier transforms. On every iteration in real space the object is cut off to a well defined rectangular area, called the support. Along with
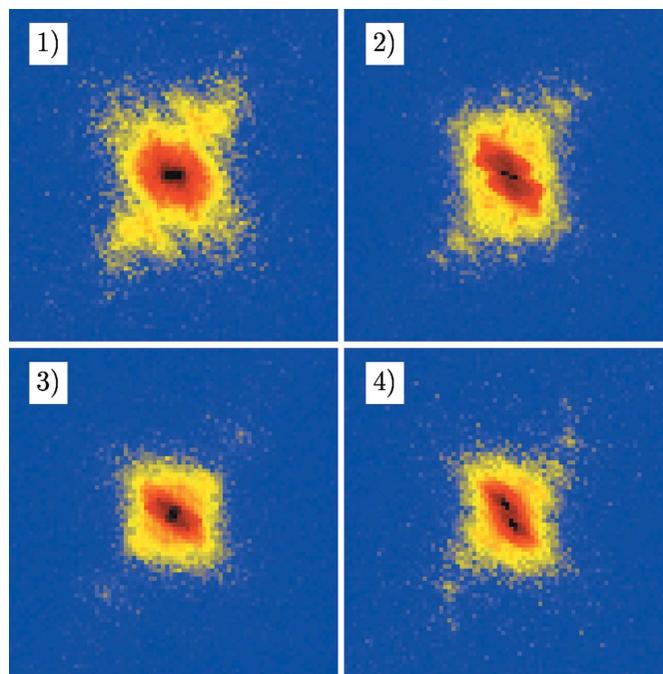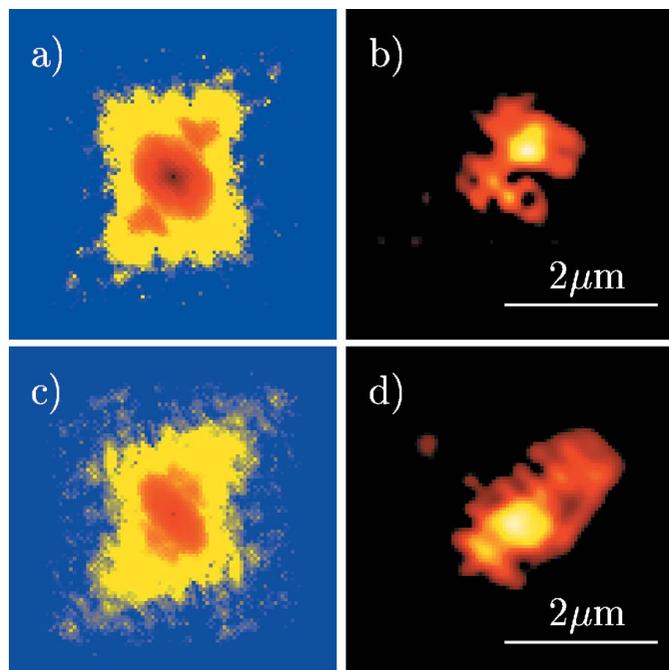

Figure 2
Symmetrized intensity on a logarithmic scale of CXD patterns from a single crystal. The patterns of Fig. 1 were symmetrized about the center of the pattern.

the reciprocal-space data, the program requires a real-space starting array. This array needs to be of the same size and contains the real-space support. The size and shape of the support is determined beforehand by looking at the spacing of the fringes in the data or at the autocorrelation of the object obtained by a Fourier transform of the intensity. The support has a value of 1 where the electron density can be non-zero and 0 everywhere else. For these data, owing to the orientation of the flares, the support used was a square of 60×60 pixels rotated by $45°$. In the present case we used a combination of hybrid input–output and error reduction. Both of these represent different kinds of real-space constraints (Marchesini, 2007). A total of 50 different fits were performed with different random starting phases for each of the images. The very best fit, determined to be the fit where the recovered diffracted intensity most accurately matched the symmetrized intensity, is shown for two images in Fig. 3.

Both of the reconstructions obtained in Fig. 3 are quite good considering all the difficulties of the experiment and the low quality of the data. The reconstructed projection of the crystal shape displays a compact object with some well defined facets on both reconstructions shown on Fig. 3. Even though the same object is illuminated in both cases, the changes in time owing to sample rotation and radiation damage explain why the two reconstructions look different. The rotation of the crystal means a projection of the crystal shape on two different planes is measured. In the second reconstruction a slightly more elongated object is recovered with again fairly sharp interfaces. The size of the objects reconstructed is around $2 \mu m$ as shown. Given the source properties and the Si(111) monochromator used, this object was at the very limit and

Figure 3
Reconstructed diffraction patterns and real-space images. (*a*) Reconstructed scattered amplitude of image 1 of Fig. 2. Note this is not on the same scale as the measurement. (*b*) Reconstructed crystal shape of image 1. (*c*) and (*d*) Same for image 2.

ence for the second slice shown in Fig. 3(*d*). Comparing the other reconstructions reveals that the other solutions found with higher χ^2 rapidly become very different. The low quality of the data leads to a low reproducibility of the reconstructions and limits the resolution at which one can be confident in the image obtained. Nevertheless, the overall shape of the images and the facets are reproducible while the internal structure of the particle is not reliably imaged.

7. Lattice distortions

The single slices of available data make it impossible to obtain more than a two-dimensional projection of the crystal shape of micrometer-sized protein crystals. However, sense of the third dimension can be deduced from the overall characteristics of the many time series that were measured. Every so often in a time series some of the Bragg peaks are seen to display sharp streaks of intensity in a direction which is not radial, *i.e.* not in the (111) direction for the (111) Bragg peak measured. Some examples of streaks are shown in Fig. 4.

These streaks appear and disappear with a corresponding Bragg peak owing to the tumbling motions discussed above. The streaks are therefore part of the same Bragg peak. Furthermore, in some cases the streak and the central part of the peak can be seen to change over a few images. Sometimes the streaks are seen first as the crystal rotates into the Bragg condition and the Bragg peak then appears with a corresponding change in the intensity of the streak. This is the signature of a rotating crystal with facets. Intensity streaks which are not radial can only come from a facet of the crystal in a direction different from the **q** vector (Robinson, 1986; Boutet *et al.*, 2002). These are known as crystal truncation rods

even beyond the limit for coherent illumination. This was reflected in the difficulty to observe any fringes around the Bragg peaks for these experiments. Most crystals could not be considered to be coherently illuminated for the entire exposure and it made it very difficult to observe coherent diffraction features.

The left side of Fig. 3 shows the reconstructed scattered amplitude obtained from the reconstructed crystal shape. Comparing the fitted amplitude with the data, we obtain a value of χ^2 of 0.021 for the first image and 0.027 for the second. This means the fitted amplitude is on average 14% and 16%, respectively, different from the measured amplitude.
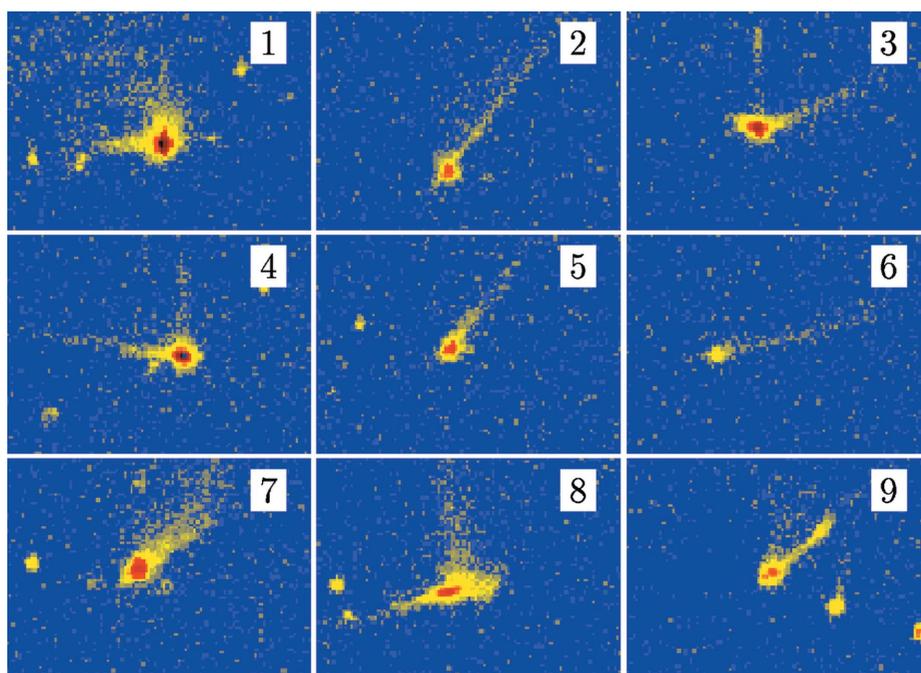
We can estimate the reproducibility of the reconstructed crystal shape using the multiple reconstructions performed on a single diffraction pattern. A computer program calculates a value similar to χ^2, that is the sum of the squared differences between the two solutions, normalized by the sum of the squared density of one of the two solutions (Williams *et al.*, 2003). A comparison of the two best fits yields an average 40% difference per pixel for the first image shown in Fig. 3(*b*) and a 30% differ-

Figure 4
Multiple examples of single Bragg peaks with intensity streaks. The measurement **q** vector is pointing up the page. The direction of these streaks is neither radial nor in the direction of the CCD readout and is always consistent with either a (111) or (200) direction.

(Robinson, 1986). The presence of these streaks is not surprising since faceting of the crystal is already seen in the images of Fig. 3. What is unexpected is that most of the streaking is towards a higher q value than that of the Bragg peak. The streaks are almost always on the high q side of the Bragg ring. Two of the eight {111} directions in the f.c.c. crystal are radial on the CCD detector. Any rotation around this vector is allowed. There is no preferred orientation of the sample since the crystals are free to rotate in solution and therefore there is no reason to expect facets to be only pointing in directions leading to streaking to higher q. As explained above, for a real object, the diffraction pattern is supposed to be centrosymmetric and streaks should be seen on both sides with equal probability but this is not found to be the case here.

The expected shape of a f.c.c. crystal is an octahedron with the eight facets corresponding to the {111} directions. This is expected because the {111} directions are the close-packed directions in the crystal, *i.e.* the directions where the fewest nearest-neighbor contacts are left dangling in order to expose the surface. Even though the {111} facets are the most stable, they are not expected to be the only surface present. However, any large facet on the crystal giving rise to a strong streak in intensity is expected to be a low-order surface with fairly tight packing.

The angles between all eight {111} directions and all six {200} directions are fixed by the lattice. The angle between two {111} facets has to be either 71°, 109° or 180°, while the angle between two (200) directions is always a multiple of 90°. The angle between a {111} facet and a {200} facet can take on a value of 35°, 55°, 126° or 144°. Therefore, by measuring the angle between the q vector of the Bragg peak, which is always the (111) direction (except for some cases where Bragg peaks were observed on the {200} ring since the detector was large enough to cover both Bragg rings), and the direction of the streak in intensity, we can determine whether the facets producing the streaks are {111}, {200} or some other direction. These angles are shown in Fig. 5 and they are found to always correspond, within the accuracy of the measurement, to one of the allowed angles between the (111) and either another {111} or a {200} direction. We can therefore conclude that the small crystals observed possess primarily {111} and {200} facets.

We propose a model to explain why the streaks almost always point towards higher q. We first assume that the crystal has an octahedral shape with {111} facets, as observed for macroscopic crystals of the same protein (Boutet *et al.*, 2002). The Fourier transform of this shape produces streaks of intensity perpendicular to each facet and intensity modulations along these fringes owing to the interference between two parallel facets. There are eight streaks and choosing one of them as the radial direction determines that four of them point toward a lower q value while the remaining four point towards higher q.

A simulated cut through the reciprocal space is shown in Fig. 6(*a*). The cut is along the plane spanned by the (110) and the (002) vectors, showing only the left half of the plane. For an infinite crystal the Bragg peaks would be δ functions.
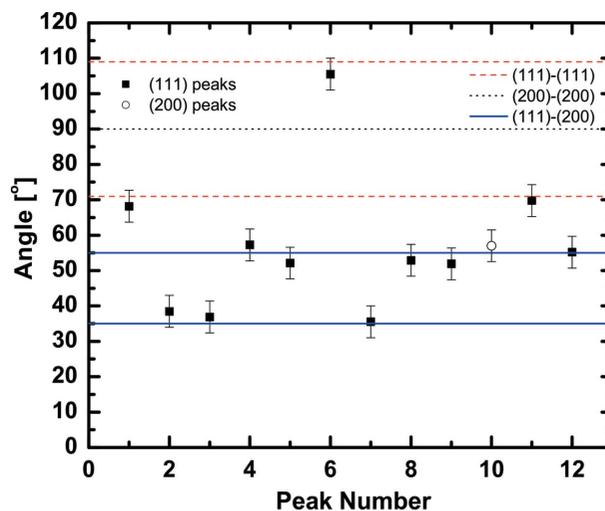
Figure 5
Measured angles (symbols) between the (111) Bragg peak and the crystal truncation rods. The lines indicate the allowed values between the (111) and the other {111} and {200} directions. The measured angles are consistent with the expected angles within the accuracy of the measurement.

However, the reciprocal space is shown for a finite crystal of octahedral shape. The scattered amplitude is shown as an isosurface for each Bragg peak. The CXD patterns shown correspond to a crystal size of 0.16 μm. The different size of the simulated crystal and the measured ones does not affect the conclusions below. The size for the simulation here was chosen to facilitate and speed up the computation and also because the smaller aspect ratio between the crystal size and the lattice spacing allows the fringes of the diffraction pattern to be more visible. The fringe spacing is 15 times smaller than the reciprocal-lattice spacing in the (111) direction. Therefore
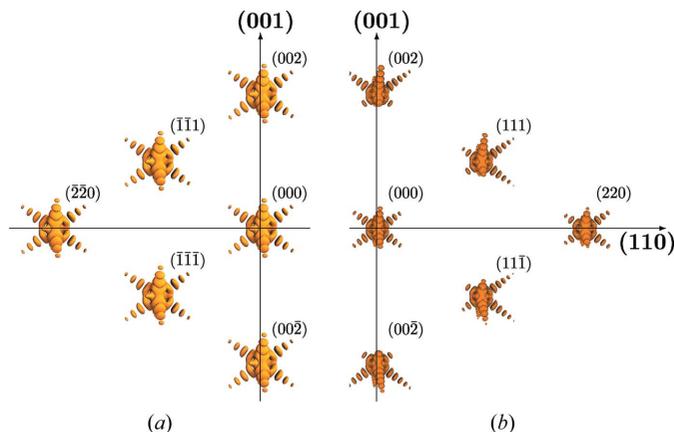
Figure 6
(*a*) Reciprocal space of a perfect micrometer-sized f.c.c. crystal with octahedral habit. The reciprocal space is shown as a cut in the plane spanned by the (110) and (002) vectors. Only the left half of the plane is shown. The streaks of intensity and fringes are perpendicular to the facets of the crystal and are what CXD aims to measure. The coherent pattern is identically reproduced at each Bragg peak, including the origin. (*b*) Cut through reciprocal space of an octahedral f.c.c. crystal with a 16% contraction of the spacing of the top layer in the crystal. This surface strain causes the Bragg peaks to be asymmetric around their center. The surface contraction leads to stronger intensities on the high q side of the Bragg peak.

the size of the crystal is 15 times larger than the direct lattice spacing in the (111) direction, which is 106 Å.

If the crystal lattice was distorted it would introduce asymmetry in the intensity. The distortion is modeled as an imaginary part of the shape function and the Fourier transform of a complex object is generally not centrosymmetric (Pfeifer *et al.*, 2006). The intensity flares from the presence of facets are still present but the intensity of the flare on one side of the peak is no longer the same as the corresponding flare on the other side.

The question is then what kind of lattice distortion will lead to enhancement of the intensity on the high q side with strong cancellation on the low side. One possible answer is a surface contraction. If the top layer of the crystal is shifted inward from the bulk value, the asymmetry to the Bragg peak will be towards higher q. This is simulated in Fig. 6(*b*) with 16% contraction of the top crystal layer. The bulk spacing of the (111) direction in ferritin is 106 Å and the strained spacing in the simulation is 89 Å. The actual amount of surface strain cannot be determined accurately with the available experimental data. The presence of the surface strain has the effect of distorting the diffraction pattern radially from the origin. The zeroth-order peak at the origin is unchanged by the presence of strain and the entire reciprocal space is still centrosymmetric about the origin but no longer locally symmetric around each Bragg peak.

The presence of lattice contraction near the surface giving rise to the observed asymmetry towards higher q is a reasonable explanation in light of previous published work (Boutet *et al.*, 2002). Such a surface contraction was previously measured on large millimeter-size ferritin crystals. The crystal truncation rod was measured and also showed a strong asymmetry toward the high q side. This was fitted to a 4% contraction of the top layer of the crystal. From general considerations it is expected that a small crystal will have bigger effects of strain than a large one, because of the contribution of its edges and corners.

It is worth pointing out that it is possible that the lattice contraction of the crystal layers near the surface is not an intrinsic property of holoferritin nanocrystals. The presence of the iron core with the molecule which helped the feasibility of these experiments could also be the cause of the observed effect *via* an enhancement of radiation damage. The photoionization of the iron in the core at the photon energy used is very high and leads to highly charged iron species which interact with the protein shells and lead to an accelerated breakdown of the molecule. This could lead to a more rapid destruction of the protein shell *via* Fenton reactions, leaving behind a mostly intact iron core. Then one would expect the lattice spacing to shrink until enough of the protein shell was destroyed and the crystal would break down completely, leading to no observed Bragg peaks. Radiation damage is expected to be seen at some level during the exposure of 60 s depositing a total dose of roughly 10^6 Gy, but this dose is not expected to fully destroy the features at the level of the resolution observed here, which was roughly 100 nm. The classical damage limit beyond which imaging of biological materials is not possible for a time-integrating measurement is of the order of 10 nm (Howells *et al.*, 2005).

While this provides an explanation for the shrinking lattice spacing, it does not explain in itself the asymmetry in the Bragg peaks. It is possible that the surface layers of the crystals are more sensitive to radiation damage but this would be unexpected at the photon energy used here where the X-rays penetrate through the entire crystal and one would expect all proteins within the crystals to become damaged the same way, regardless of their position within the crystal. The penetration depth of 9 keV photons through iron is larger than 5 µm.

We clearly do not observe here a collapse of the lattice spacing throughout the crystal. That would lead to a change in the position of the Bragg peak but not to asymmetric Bragg peaks. If radiation damage is the source of the surface contraction observed, by destroying the crystal contacts and removal of the water separating neighboring proteins, then the data presented here indicate that the outer layers of the nanocrystal collapse more rapidly than the inner layers.

8. Conclusion

The study of nanocrystals of proteins is expected to be an important field of research at the soon to be available hard X-ray free-electron lasers (FELs), *i.e.* the Linac Coherent Light Source (LCLS), the European X-ray FEL (XFEL) and the SPring-8 Compact SASE Source (SCSS) in California, Germany and Japan, respectively. The short pulse duration and extremely high peak power of these sources will allow instantaneous snapshots of nanocrystals to be taken with limited damage during the pulse (Neutze *et al.*, 2000). All the experimental difficulties encountered during the measurements described here can be avoided by performing the measurements in a few femtoseconds. On this time scale no tumbling of the sample would be observed and the damage to the sample would be minimal during the pulse. Results in the VUV energy range at the free-electon laser in Hamburg (FLASH) recently supported the concept of flash imaging in which the sample is completely destroyed but not before the diffraction pattern of the undamaged sample is recorded (Chapman *et al.*, 2006, 2007; Hau-Riege *et al.*, 2007). The new hard X-ray FEL sources could make it possible to image crystal nuclei as they form since a single FEL shot would provide sufficient signal for imaging with nanometer or better resolution in two dimensions.

Nanocrystals of biological macromolecules could also be used directly for structure determination using hard X-ray FELs. Single nanocrystalline particles can be delivered to the X-ray beam and a two-dimensional diffraction pattern can be collected before the onset of radiation damage using new advances in single-particle delivery techniques (Bogan *et al.*, 2008; Weierstall *et al.*, 2008). Each two-dimensional diffraction pattern of these particles would show Bragg peaks owing to the crystalline structure of the nanoparticles. Assuming all the particles have the same crystal structure, it would be possible to build the three-dimensional diffraction pattern using the Bragg peaks to determine the relative orientation of each

crystalline nanoparticle. A good signal-to-noise ratio could be achieved by adding the diffraction patterns from many nanocrystals once their orientations have been classified. The coherent illumination of the particle would cause each Bragg peak to have structure representative of the shape of the particle as discussed in the results shown above. Since every nanocrystal would be expected to have a different shape, all of the fine structure around its Bragg peaks would be different. This would make it impossible to reconstruct in three-dimensions the entire shape of a nanocrystal. However, the common lattice structure of all the nanocrystals would make it possible to sum all the intensities at the Bragg peaks and a data set identical to standard macromolecular crystallography could be obtained with a large collection of nanocrystals. The atomic structure could then be recovered using the standard crystallography techniques.

The results presented in this paper demonstrate that it is possible to measure the coherent diffraction pattern from a single micrometer-sized crystal of protein. While such a measurement is difficult and the resolution of the images that can be obtained is limited by radiation damage and the thermal motions of the small crystals, it is expected that both of these problems can be overcome using hard X-ray free-electron lasers. The study of nanocrystals to understand better the crystal formation process as well as for macromolecular structure determination could become a very important field of research at the new fourth-generation X-ray sources in the future.

References

Als-Nielsen, J. & McMorrow, D. (2001). *Elements of Modern X-ray Physics*. New York: Wiley.

Bogan, M. J. *et al.* (2008). *Nanoletters*, **8**, 310–316.

Boutet, S. & Robinson, I. K. (2006). *J. Synchrotron Rad.* **13**, 1–7.

Boutet, S., Robinson, I., Hu, Z., Thomas, B. & Chernov, A. (2002). *Phys. Rev. E*, **66**, 061914.

Chapman, H. N. *et al.* (2006). *Nat. Phys.* **2**, 839–843.

Chapman, H. N. *et al.* (2007). *Nature (London)*, **448**, 676–680.

Fienup, J. (1982). *Appl. Opt.* **21**, 2758–2769.

Gerchberg, R. & Saxton, W. (1972). *Optik*, **35**, 237–246.

Granier, T., Gallois, B., Dautant, A., Langlois d'Estaintot, B. & Précigoux, G. (1997). *Acta Cryst.* D**53**, 580–587.

Harrison, P. M. & Arosio, P. (1996). *Biochim. Biophys. Acta*, **1275**, 161–203.

Hau-Riege, S. P. *et al.* (2007). *Phys. Rev. Lett.* **98**, 145502.

Howells, M., Beetz, T., Chapman, H. N., Cui, C., Holton, J. M., Jacobsen, C. J., Kirz, J., Lima, E., Marchesini, S., Miao, H., Sayre, D., Shapiro, D. A. & Spence, J. C. H. (2005). Preprint at http://arxiv.org/abs/physics/0502059.

Marchesini, S. (2007). *Rev. Sci. Instrum.* **78**, 011301.

Miao, J., Charalambous, P., Kirz, J. & Sayre, D. (1999). *Nature (London)*, **400**, 342–344.

Nave, C. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 257–260.

Nave, C. & Hill, M. A. (2005). *J. Synchrotron Rad.* **12**, 299–303.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature (London)*, **406**, 752–757.

Pfeifer, M., Vartanyants, I., Vartanyants, I., Harder, R., Williams, G. & Robinson, I. (2006). *Nature (London)*, **442**, 63–66.

Ravelli, R. B. G., Theveneau, P., McSweeney, S. & Caffrey, M. (2002). *J. Synchrotron Rad.* **9**, 355–360.

Robinson, I. K. (1986). *Phys. Rev. B*, **33**, 3830–3836.

Robinson, I., Vartanyants, I., Williams, G., Pfeifer, M. & Pitney, J. (2001). *Phys. Rev. Lett.* **87**, 195505.

Sayre, D. (1952). *Acta Cryst.* **5**, 843.

Sayre, D., Chapman, H. N. & Miao, J. (1998). *Acta Cryst.* A**54**, 232–239.

Vartanyants, I. A. & Robinson, I. K. (2001). *J. Phys. Condens. Matter*, **13**, 10593–10611.

Weierstall, U., Doak, R., Spence, J., Starodub, D., Shapiro, D., Kennedy, P., Warner, J., Hembree, G., Fromme, P. & Chapman, H. (2008). *Exp. Fluids*, **44**, 675–689.

Williams, G., Pfeifer, M., Vartanyants, I. & Robinson, I. (2003). *Phys. Rev. Lett.* **90**, 175501.