Macromolecular diffractive imaging using imperfect crystals

Kartik Ayyer[1], **Oleksandr Yefanov**[1], **Dominik Oberthür**[2], **Shatabdi Roy-Chowdhury**[3,4], **Lorenzo Galli**[1,2], **Valerio Mariani**[1], **Shibom Basu**[3,4], **Jesse Coe**[3,4], **Chelsie E. Conrad**[3,4], **Raimund Fromme**[3,4], **Alexander Schaffer**[3,4], **Katerina Dörner**[1,3], **Daniel James**[5], **Christopher Kupitz**[3,6], **Markus Metz**[2], **Garrett Nelson**[5], **Paulraj Lourdu Xavier**[1,2], **Kenneth R. Beyerlein**[1], **Marius Schmidt**[6], **Iosifina Sarrou**[7], **John C. H. Spence**[5], **Uwe Weierstall**[5], **Thomas A. White**[1], **Jay-How Yang**[3,4], **Yun Zhao**[3,4], **Mengning Liang**[8], **Andrew Aquila**[8], **Mark S. Hunter**[8], **Joseph S. Robinson**[8], **Jason E. Koglin**[8], **Sébastien Boutet**[8], **Petra Fromme**[3,4], **Anton Barty**[1], and **Henry N. Chapman**[1,2,9]

[1]Center for Free-Electron Laser Science, DESY, 22607 Hamburg, Germany

[2]Department of Physics, University of Hamburg, 22761 Hamburg, Germany

[3]Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA

[4]Center for Applied Structural Discovery, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, USA

[5]Department of Physics, Arizona State University, Tempe, Arizona 85287, USA

[6]Physics Department, University of Wisconsin, Milwaukee, WI 53211, USA

[7]Institute of Molecular Biology and Biotechnology, Foundation for Research and Technology, Hellas, Crete, Greece

[8]Linac Coherent Light Source, Stanford Linear Accelerator Center (SLAC) National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA

[9]Centre for Ultrafast Imaging, 22607 Hamburg, Germany

Abstract

The three-dimensional structures of macromolecules and their complexes are predominantly elucidated by X-ray protein crystallography. A major limitation is access to high-quality crystals, to ensure X-ray diffraction extends to sufficiently large scattering angles and hence yields sufficiently high-resolution information that the crystal structure can be solved. The observation that crystals with shrunken unit-cell volumes and tighter macromolecular packing often produce higher-resolution Bragg peaks[1,2] hints that crystallographic resolution for some macromolecules may be limited not by their heterogeneity but rather by a deviation of strict positional ordering of the crystalline lattice. Such displacements of molecules from the ideal lattice give rise to a continuous diffraction pattern, equal to the incoherent sum of diffraction from rigid single molecular complexes aligned along several discrete crystallographic orientations and hence with an increased information content[3]. Although such continuous diffraction patterns have long been observed—and are of interest as a source of information about the dynamics of proteins[4]—they have not been used for structure determination. Here we show for crystals of the integral membrane protein complex photosystem II that lattice disorder increases the information content and the resolution of the diffraction pattern well beyond the 4.5 Å limit of measurable Bragg peaks, which allows us to directly phase[5] the pattern. With the molecular envelope conventionally determined at 4.5 Å as a constraint, we then obtain a static image of the photosystem II dimer at 3.5 Å resolution. This result shows that continuous diffraction can be used to overcome long-supposed resolution limits of macromolecular crystallography, with a method that puts great value in commonly encountered imperfect crystals and opens up the possibility for model-free phasing[6,7].

High-resolution Bragg diffraction from a crystal requires a high degree of regularity, which is often not obtained in practice. If one of the structural units that make up a crystal (such as a molecule) is displaced from the ideal lattice by an amount s then the phase of the diffracted wave from this unit is changed by $2\pi s/d$ at a scattering angle 2θ corresponding to the resolution d, leading to destructive interference at that resolution when $s = d/2$. If all units are displaced randomly with a root mean square (RMS) displacement σ along a given coordinate then Bragg intensities will accordingly diminish at a resolution of $d = 2\pi\sigma$ according to the well-known Debye-Waller factor $\exp(-4\pi^2 q^2 \sigma^2)$ where $q = 1/d = 2\sin\theta/\lambda$ for an X-ray wavelength λ (ref. 8). The disorder length σ is easily estimated from a so-called Wilson plot[9] of the diffraction data. Less widely appreciated is the fact that while the Bragg peaks diminish exponentially with the square of scattering angle, an incoherent sum of the diffraction of individual structural units arises contrariwise[8,10,11] (see Fig. 1 and Methods Sec. 2) in order to conserve the energy lost from the Bragg peaks. This incoherent sum is equal to the Fraunhofer diffraction pattern of the structural unit multiplied by the number of units, without any modulation by Bragg peaks.

Continuously modulated diffraction intensities have frequently been observed in diffraction patterns of protein crystals[10-15], and have even been ascribed to the disorder in the arrangement of rigid structural units as large as single macromolecules[10-13]. Although the continuous diffraction was seen to follow the point-group symmetry of the crystal[14] previous studies did not consider the use of these patterns for structure determination. Protein crystals usually contain several molecules or molecular assemblies per unit cell as well as a solvent. The packing of molecules is mediated by non-covalent contacts between them (H-bonds as

well as electrostatic and hydrophobic interactions). It is therefore reasonable to propose a model where the macromolecule is the rigid structural unit. In this case, with uncorrelated molecular displacements, the continuous diffraction pattern will consist of the incoherent sum of the continuous diffraction patterns of the macromolecule aligned in each of its various orientations within the crystal (given by Eqn. 9 in Methods).

Accessing the "single molecule" continuous diffraction overcomes the phase problem of crystallography, by increasing the number of observable coefficients to exceed the number required to describe the molecule, independent of resolution[16,17] (see Methods Sec. 7). Bragg intensities can be equivalently represented in real space, by Fourier transformation, as the three-dimensional (3D) pair-correlation function (or Patterson function[9]) of the crystal. This cannot always distinguish between object pairs existing only in one unit cell or straddling neighbours, whereas such an ambiguity is removed if the object is non-repeating. The highly-sampled single molecule diffraction data of an imperfect crystal (averaged over the orientational symmetries of the point group) could be used to obtain a real-space image of the macromolecule at a resolution beyond the resolution of the Bragg peaks, using iterative phasing algorithms[18-22] of coherent diffractive imaging and aligned molecule diffraction[16].

Measurements from crystals of the large integral membrane protein-complex photosystem II (PSII) support the notion that the inherent resolution of the single asymmetric unit in the crystal can be better than that of the entire crystal. Recent X-ray free-electron laser experiments yielded Bragg diffraction from crystals of 1 μm^3 volume to a resolution of 5 Å (ref. 23). The protein-cofactor complex consists of a dimer of approximately 700 kDa. There are only four crystal contacts per dimer, and 64% solvent by volume in the crystal. If ascribed to disorder of the lattice alone, the Bragg limit of 5 Å could be caused by a RMS displacement of molecules by 0.8 Å in each dimension (1.4 Å RMS 3D displacement, see Methods Sec. 2). Shen and co-workers recently reported Bragg diffraction from large PSII crystals (1 mm^3 in volume) extending to 1.9 Å, achieved by dehydrating the crystal in the presence of small amphiphiles, leading to an increase in crystal contacts and a tighter crystal packing[2]. We set out to determine whether the limited crystallographic resolution of 5 Å is caused by random displacements of rigid-body structural units.

Still snapshot diffraction patterns, such as shown in Fig. 2a, were recorded from room-temperature microcrystals of PSII in random orientations, as per the method of serial femtosecond crystallography[23-25] (Methods, Sec. 1). The orientation of each individual crystal was determined from its diffraction pattern by indexing the Bragg peaks[26] allowing it to be inserted into a 3D array to build up the full intensity distribution[27] in the crystal coordinate frame (Fig. 2c,d). The accumulation of continuously modulated features or "speckles" in the diffraction volume implies that they are associated with the crystal rather than other scattering sources. Further, consistent with the hypothesis of rigid-body random displacements, the features are absent at lowest resolutions and become dominant at scattering angles beyond the highest-angle Bragg peaks. The size of the structural unit can be estimated in an unbiased fashion from the autocorrelation of the diffracting object (the 3D pair distribution function), obtained by a Fourier transform of the continuous diffraction

intensities[22]. We found this to be in agreement with the size of a PSII dimer, as shown in Fig. 3.

One can consider several ways to obtain a high-resolution structure from the data shown in Fig. 2. We found that a particularly robust method was to treat the Bragg and continuous diffraction as two distinct sources of data of the same structure, the first arising from a coherent interference of molecules in the unit cell, the second due to the incoherent addition of aligned single molecule diffraction (Methods, Sec. 3). Starting with a known search model for PSII, we first produced an electron density map of the PSII dimer by molecular replacement phasing to 4.5 Å resolution, limited by the angular extent of the Bragg peaks. This map was then used to generate a low-resolution binary mask of the smoothed molecular envelope of a single photosystem II dimer shown in Fig. 3d. This mask forms the real-space constraint[4] used to generate a 3D image of the electron density of the PSII dimer by iterative phasing of the continuous diffraction.

The iterative phasing of the continuous diffraction, covering a resolution range of 4.5 Å to 3.3 Å, was carried out using the difference map algorithm[19,16] (Methods, Sec. 4). As with other iterative phasing algorithms, an initial guess of the electron density obtained from random phases is constrained on each iteration to be consistent with the measured diffraction and to fit within a certain finite (but not necessarily precisely known) real-space extent called the "support"—the smoothed envelope generated from the 4.5 Å resolution electron density map in this case.

After convergence of the iterate, the Fourier amplitudes and phases over the entire resolution range were combined to synthesise a 3.3 Å resolution structure. Averaging solutions obtained over multiple random starts produced a self-consistent electron density to 3.5 Å resolution, as determined by the Fourier shell correlation (FSC) and the phase retrieval transfer function (PRTF)[28] (Fig. 3e). Finally, pseudo-crystallographic refinement[29] of this electron density was carried out, following a similar procedure as for single-particle cryo-electron microscopy data (Methods, Sec. 5). Regions of the final electron density map are noticeably superior to the same regions generated using the Bragg peaks alone (Fig. 4). In particular, helices show better definition in side-chains, and the model better follows the electron density. The benefit of including the continuous diffraction data is quantified in the improvement of the R_{free} factor at low resolution (Extended Data Fig. 1). Although the assembled continuous diffraction data is quite noisy, due to the limited number of 2,848 diffraction patterns, the structure clearly benefits from its inclusion. Control experiments verified these benefits (see Methods Sec. 6). Further improvements could be expected with more measurements.

Our approach, demonstrated using serial diffraction data collected at an XFEL may also be applicable to larger crystals measured with conventional X-ray sources. The advantages of data collection with an XFEL include a higher tolerable dose and the capacity to measure data from hundreds of thousands of crystals from which the most suitable can be selected based solely on their diffraction patterns. The improved structural refinement at a resolution beyond measurable Bragg peaks establishes the origin of the continuous diffraction and demonstrates its utility in obtaining significant increases in resolution without having to

improve crystal lattice order. The crystal is simply a means to obtain an ensemble of oriented macromolecules to carry out aligned-molecule diffractive imaging[30]. We anticipate that the approach will find application to a range of macromolecular systems and may allow *ab initio* phasing of crystals. After many decades of being discarded as of little use, all too common "poorly-diffracting" protein crystals might now be exploited for high-resolution structure determination.

Methods

1. Crystallography

Microcrystals were grown by a free interface diffusion centrifugation technique, as described previously[23]. The size of the crystals ranged from 0.5 to 5 μm width, which was determined by dynamic light scattering. The space group of the crystals is $P2_12_12_1$ (ref 23).

The experiments were carried out at the Coherent X-ray Imaging (CXI) end station[31] of the Linac Coherent Light Source (LCLS)[32]. X-ray pulses of 40 fs duration and 9.48 keV photon energy were focused using X-ray optics in a Kirkpatrick-Baez geometry to a beam size of about 1×1 μm^2 area. A suspension of photosystem-II crystals in their crystallization buffer[23] was pre-filtered using 20 μm stainless steel filters and delivered at room-temperature to the X-ray interaction region as a 3–4 μm diameter jet at a flow rate of 10–15 μl/min, using a gas dynamic virtual nozzle (GDVN)[33,34] as previously described[23]. The average X-ray pulse energy at the sample was about 1.0 mJ, assuming a beamline transmission of 60%. The dose to each crystal was at most 275 MGy, and diffraction patterns were acquired on the basis of "diffraction before destruction"[35,36]. Diffraction frames were collected on a Cornell-SLAC Pixel Array Detector (CSPAD) at a rate of one frame per X-ray pulse, at 120 Hz (ref. 37). 1.24 million frames were collected, during an integrated collection time of 2.9 hours.

The diffraction frames were initially processed following a standard data evaluation pipeline established for serial femtosecond crystallography. Frames containing crystal diffraction were found using *Cheetah*[38] based on the identification of Bragg peaks. These "hits," accounting for about 5.0 % of the frames, were transferred to *CrystFEL* (version 0.6.0)[26,39] along with the peak positions, for automated indexing and to obtain a merged set of integrated Bragg intensities. Indexing the lattice determines the crystal orientation and crystal lattice parameters for each diffraction pattern, information that was later used to assemble the full 3D pattern of Fig. 2c. The relative coordinates of all pixels in the segmented CSPAD detector were determined by a procedure using a strongly diffracting standard (lysozyme)[40]. Using these calibrated pixel positions, a total of 25,585 patterns out of 61,946 were successfully indexed to create a final list of Bragg reflections and their integrated intensities. The resulting data set of Bragg intensities extended to a resolution of 4.5 Å, as indicated by the modified Pearson correlation coefficient CC* (ref. 41) (Extended Data Fig. 2).

The crystal orientations obtained by indexing the lattice were subsequently used to assemble the individual diffraction pattern snapshots into a 3D reciprocal space grid representing the full-pattern intensity distribution of the average disordered crystal. The volume consisted of

501×501×501 cubic voxels with a width of 1.364×10^{-3} Å^{-1} (where $q = 1/d$). Thus, the magnitude of the wave-vector transfer q at the face centre of the voxel array was 0.341 Å^{-1} corresponding to a resolution of 2.93 Å. Most of the individual snapshot patterns did not show noticeable continuous diffraction above a significant featureless background caused by scattering from the water. The strength of the background depends linearly on the thickness of the liquid jet, and was removed by subtracting the average computed by summing counts along circles of constant q after correcting for the linear polarisation of the X-ray beam. This radial average excluded very high and very low values to avoid the influence of Bragg peaks. The strength of the remaining crystal-dependent continuous diffraction signal for each frame was estimated by first replacing Bragg peak values with a local fit of the slowly varying continuous signal. This was achieved by using a thresholded median filter: pixel values that differed from a 5×5 pixel median-filtered copy of the pattern by more than one standard deviation above the mean were selected as bright pixels. The values of these, and pixels within a 3 pixel distance of them, were replaced with the values from the 5×5 pixel median-filtered copy. The strength of the continuous diffraction signal for each frame was best estimated from pixels in the range of 0.1 Å^{-1} to 0.2 Å^{-1} after excluding the Bragg signal, and is plotted as a function of the total integrated Bragg-peak signal in Extended Data Fig. 3 for all of the 25,585 indexed patterns. The plot shows a strong linear correlation between these two signals. The incident intensity for each snapshot diffraction pattern fluctuates due to the random targeting of crystals to the X-ray focus profile and their variation in size. The trend shows that increased intensity or crystal size leads to an increase in both the Bragg and continuous diffraction. A single linear trend indicates little variation in the degree of lattice disorder of the crystals. We found that selecting the 2,848 patterns with the strongest continuous diffraction to create the full 3D pattern gave higher signal to noise and continuous scattering contrast at high resolution than including the entire 25,585 patterns.

2. The effect of rigid body disorder on Bragg and continuous diffraction

A deviation from strict translational correlation in a crystal gives rise to a component of the diffraction pattern that is continuous, which is usually studied in the context of protein dynamics[4,10-14]. The time-averaged diffraction of a moving object can be considered as the incoherent sum of static snapshots over the time of the exposure, and for uncorrelated motions of structural units in a crystal the time-averaged diffraction is equal to any particular snapshot in the limit of a large number of unit cells. Thus, time-integrated diffraction measurements cannot distinguish static disorder from dynamic. In the case of X-ray FEL exposures the positions of structural units such as entire macromolecules are certainly static. Following the approach of Cowley[42] we first derive the diffraction of a translationally disordered crystal containing only one rigid object per unit cell (i.e. P1 symmetry). The real-space density $\rho_c(\mathbf{x})$ of this crystal can be described by

$$\rho_c(\mathbf{x}) = \rho_A(\mathbf{x}) \otimes \sum_{n=1}^{N} \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n) \quad (1)$$

where $\rho_A(\mathbf{x})$ is the density of the asymmetric unit, repeated over N positions in the crystal given by random displacement vectors Δ_n from ideal lattice positions $\mathbf{a}_n$. In Eqn. (1), $\otimes$

represents a convolution and $\delta(\mathbf{x})$ is the Dirac delta function. Without loss of generality, the mean displacement $\langle \Delta_n \rangle = 0$ and $\langle \Delta_n^2 \rangle = \sigma^2$.

The 3D diffraction intensity distribution is the squared modulus of the Fourier transform of the crystal electron density $\rho_c(\mathbf{x})$. This can also be obtained by calculating the Fourier transform of the electron density auto-correlation function (the Patterson function), given by $P(\mathbf{x}) = \rho_c(\mathbf{x}) \otimes \rho_c^*(-\mathbf{x})$. For the disordered crystal,

$$
\begin{aligned}
P(\mathbf{x}) &= \rho_A(\mathbf{x}) \otimes \sum_{n=1}^{N} \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n) \otimes \rho_A^*(-\mathbf{x}) \otimes \sum_{k=1}^{N} \delta(\mathbf{x} + \mathbf{a}_k + \Delta_k) \\
&= \rho_A(\mathbf{x}) \otimes \rho_A^*(-\mathbf{x}) \otimes \left(\sum_{n=1}^{N} \sum_{k=1}^{N} \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n) \otimes \delta(\mathbf{x} + \mathbf{a}_k + \Delta_k) \right) \\
&= \rho_A(\mathbf{x}) \otimes \rho_A^*(-\mathbf{x}) \otimes \left(\sum_{n=1}^{N} \sum_{k=1}^{N} \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n + \mathbf{a}_k + \Delta_k) \right) \\
&= \rho_A(\mathbf{x}) \otimes \rho_A^*(-\mathbf{x}) \otimes \left(N\delta(\mathbf{x}) + \sum_{n=1}^{N} \sum_{\substack{k=1 \\ k \neq n}}^{N} \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n + \mathbf{a}_k + \Delta_k) \right).
\end{aligned}
\tag{2}
$$

The double sum in Eqn. (2) can be understood for large crystals in the following manner. A constant lattice difference vector $\mathbf{a}_n - \mathbf{a}_k$ is obtained for many pairs of n and k, as illustrated in Extended Data Fig. 4. For this difference, the sum samples all possible displacement differences, $\Delta_n - \Delta_k$, except when $n = k$. Thus, the delta function at that lattice difference vector is broadened by the probability distribution of the displacement difference. If all unit cells are displaced independently, this probability distribution of the difference vectors will be the autocorrelation of the probability distribution of a single displacement. For a Gaussian with standard deviation σ, this is just another Gaussian with a standard deviation of $\sqrt{2}\sigma$. Thus,

$$
\sum_{n=1}^{N} \sum_{\substack{k=1 \\ k \neq n}}^{N} \delta(\mathbf{x} - \mathbf{a}_n - \Delta_n + \mathbf{a}_k + \Delta_k) = \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \otimes \sum_{n=1}^{N} \sum_{\substack{k=1 \\ k \neq n}}^{N} \delta(\mathbf{x} - \mathbf{a}_n + \mathbf{a}_k)
\tag{3}
$$

Keeping in mind that the Fourier transform $\mathscr{F}$ maps convolutions to multiplications, the intensity can be calculated from Eqns. (2) and (3) as

$$I(\vec{q})=|F_A(\mathbf{q})|^2 \mathscr{F}\left[N\delta(\mathbf{x}) + \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \otimes \sum_{n=1}^{N} \sum_{\substack{k=1 \\ k \neq n}}^{N} \delta(\mathbf{x} - \mathbf{a}_n + \mathbf{a}_k) \right]$$

$$=|F_A(\mathbf{q})|^2 \mathscr{F}\left[N\delta(\mathbf{x}) \otimes \left(\delta(\mathbf{x}) - \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}}\right) + \frac{\exp(-x^2/4\sigma^2)}{\sqrt{4\pi\sigma^2}} \otimes \sum_{n=1}^{N}\sum_{k=1}^{N} \delta(\mathbf{x}-\mathbf{a}_n+\mathbf{a}_k) \right]$$

$$=|F_A(\mathbf{q})|^2 \left(N(1-e^{-4\pi^2\sigma^2 q^2}) + e^{-4\pi^2\sigma^2 q^2} \sum_{n=1}^{N}\sum_{k=1}^{N} \exp(2\pi i(\mathbf{a}_n - \mathbf{a}_k)\cdot\mathbf{q}) \right)$$

(4)

where $F_A(\mathbf{q})$ is the (complex-valued) Fourier transform of $\rho_A(\mathbf{x})$ at a resolution $|\mathbf{q}| = 1/d$. Equation (4) shows that the diffraction pattern of a disordered crystal with a single asymmetric unit consists of the squared modulus of the molecular transform F_A modulated by two terms. The second term is the reciprocal lattice, which gives the Bragg peaks, further modulated by the Debye-Waller factor $\exp(-4\pi^2\sigma^2 q^2)$. The first term has no lattice associated with it, but is a monotonically increasing function with a value of 0 at $q = 0$ and asymptoting to a constant for values of q larger than $1/(2\pi\sigma)$ Thus, at high q values, the diffraction is proportional to the continuous diffraction pattern of the rigid unit, $|F_A(\mathbf{q})|^2$.

Generalising further to the case of M asymmetric units in a unit cell, one can write the unit cell electron density as

$$\sum_{m=1}^{M} \rho_A(\mathbf{R}_m\mathbf{x}+\mathbf{t}_m) \quad (5)$$

where $\mathbf{R}_m$ and $\mathbf{t}_m$ are the rotation and translation operators for the m-th asymmetric unit. As in Eqn. (1), the electron density for a disordered crystal is

$$\rho_c(\mathbf{x})=\sum_{m=1}^{M}\sum_{n=1}^{N}\rho_A(\mathbf{R}_m\mathbf{x}+\mathbf{t}_m)\otimes\delta(\mathbf{x}-\mathbf{a}_n-\Delta_{m,n}) \quad (6)$$

where $\Delta_{m,n}$ is the random displacement of the m-th asymmetric unit in the n-th unit cell.

The Patterson function of $\rho_C(\mathbf{x})$ can be split into two terms, the first a sum of like terms connecting corresponding asymmetric units in different unit cells and the second consisting of the cross terms. Thus,

$$P(\mathbf{x})=\sum_{m=1}^{M}\sum_{n=1}^{N}\rho_A(\mathbf{R}_m\mathbf{x}+\mathbf{t}_m)\otimes\delta(\mathbf{x}-\mathbf{a}_n-\Delta_{m,n})\otimes\sum_{j=1}^{M}\sum_{k=1}^{N}\rho_A(\mathbf{R}_j\mathbf{x}+\mathbf{t}_j)\otimes\delta(\mathbf{x}-\mathbf{a}_k-\Delta_{j,k})$$

$$=\sum_{m=1}^{M}\left(\rho_A(\mathbf{R}_m\mathbf{x}+\mathbf{t}_m)\otimes\rho_A^*(-\mathbf{R}_m\mathbf{x}+\mathbf{t}_m)\otimes\sum_{n}\delta(\mathbf{x}-\mathbf{a}_n-\Delta_{m,n})\otimes\sum_{k}\delta(\mathbf{x}+\mathbf{a}_k+\Delta_{m,k})\right)$$

$$+\sum_{m=1}^{M}\sum_{\substack{j=1 \\ j \neq m}}^{M}\left(\rho_A(\mathbf{R}_m\mathbf{x}+\mathbf{t}_m)\otimes\rho_A^*(-\mathbf{R}_j\mathbf{x}+\mathbf{t}_j)\otimes\sum_{n}\delta(\mathbf{x}-\mathbf{a}_n-\Delta_{m,n})\otimes\sum_{k}\delta(\mathbf{x}+\mathbf{a}_k+\Delta_{j,k})\right)$$

(7)

For the first term in Eqn. (7), the convolution of the two delta function sums can be evaluated similar to the single asymmetric unit case in Eqn. (3). This yields a term $N\delta(\mathbf{x})$ for $n = k$ and the blurred delta functions for all other cases. However, for the cross term, since different asymmetric units are assumed to move independently, this convolution does not have the $N\delta(\mathbf{x})$ term. Taking the inverse Fourier transform of $P(\mathbf{x})$ and simplifying,

$$
I(\mathbf{q}) = \left(\sum_{m=1}^{M} |F_A(\mathbf{R}_m\mathbf{q})|^2 \right) \left(N(1 - e^{-4\pi^2\sigma^2q^2}) + e^{-4\pi^2\sigma^2q^2} \sum_{n=1}^{N}\sum_{k=1}^{N} \exp(2\pi i(\mathbf{a}_n - \mathbf{a}_k)\cdot\mathbf{q}) \right)
$$
$$
+ \sum_{m=1}^{M}\sum_{\substack{j=1 \\ j \neq m}}^{M} F_A(\mathbf{R}_m\mathbf{q})F_A(\mathbf{R}_j\mathbf{q})e^{2\pi i\mathbf{q}\cdot(\mathbf{t}_m - \mathbf{t}_j)} \left(e^{-4\pi^2\sigma^2q^2} \sum_{n=1}^{N}\sum_{k=1}^{N} \exp(2\pi i(\mathbf{a}_n - \mathbf{a}_k)\cdot\mathbf{q}) \right). \quad (8)
$$

On gathering the Bragg terms and factorising, the intensity distribution simplifies to

$$
I(\vec{q}) = N\left[\sum_{m=1}^{M} |F_A(\mathbf{R}_m\mathbf{q})|^2 \right] (1 - e^{-4\pi^2\sigma^2q^2}) + \left| \sum_{m=1}^{M} F_A(\mathbf{R}_m\mathbf{q})e^{2\pi i\mathbf{q}\cdot\mathbf{t}_m} \right|^2 e^{-4\pi^2\sigma^2q^2} \sum_{n=1}^{N}\sum_{k=1}^{N} \exp(2\pi i(\mathbf{a}_n - \mathbf{a}_k)\cdot\mathbf{q}). \quad (9)
$$

This equation has a similar form to Eqn. (4), in that there are two terms, representing the continuous and Bragg diffraction. The difference is that each term is modulated by different combinations of the asymmetric unit transform. The continuous diffraction is the sum of the *intensities* of each asymmetric unit transform, while the Bragg peaks depend on the coherent sum of the asymmetric unit transforms (the unit cell transform). This means that only the Bragg peaks are affected by the systematic absences of the space group, while the continuous diffraction only depends on the point group of the crystal.

This derivation started with the assumption of normally distributed displacements of rigid body asymmetric units. The conclusions hold for other displacement distributions as well, whereby the term $e^{-4\pi^2\sigma^2q^2}$ is replaced by the Fourier transform of the autocorrelation of the probability distribution. For a smooth (band-limited) distribution, the Fourier transform will reduce to zero at high q, and the continuous diffraction term will dominate. Furthermore, the derivation can be generalised to a number of non-identical rigid structural units in the unit cell.

3. Two data sources from one experiment

Structure determination could be considered in terms of a whole-pattern analysis, possibly based on Eqn. (9). An alternative employed here was to treat the Bragg diffraction and continuous diffraction as two separate data sources that both encode structural information of the macromolecular system. The observable Bragg diffraction covers the range from 0.03 Å^{-1} to 0.22 Å^{-1} (4.5 Å resolution). The continuous diffraction overlaps this range but extends to a resolution of about 2.2 Å (seen in the corners of Fig. 2d). However, to avoid the influence of the large pixel values of strong Bragg peaks when separating the continuous diffraction we consider just the range from 0.22 Å^{-1} to 0.35 Å^{-1} for this dataset. Our strategy for utilising both forms of diffraction data was to use the continuous diffraction to extend the

resolution of a structure determined from the Bragg data alone and to add strong phasing constraints by directly accessing the molecular transform of the rigid structural unit.

Given Eqn. (9), predicting that the continuous diffraction is proportional to the incoherent sum of transforms of single rigid units aligned in their various crystallographic orientations, we expect the statistics of the continuous diffraction intensities to follow that of molecular diffraction, which should be exponentially distributed[43]. This can be seen in Extended Data Fig. 1, which shows a linear trend in the log-histogram above a background level of around 1 count per voxel.

The size of the rigid structural unit that gives rise to the observed continuous diffraction pattern was estimated from the extent of the autocorrelation function, obtained by a Fourier transform of the continuous diffraction intensities. From Eqn. (9), this yields the incoherent sum of autocorrelation functions of the structural unit in its various crystallographic orientations. To avoid the influence of the Bragg peaks on this calculation, this symmetrised autocorrelation function was generated from windowed regions of the full 3D pattern above a resolution $1/q = 4.5$ Å, following a procedure previously described[22]. Using a Gaussian window function with a 15 pixel radius, 30 auto-correlation functions were generated in different locations in an orthogonal central section of the full 3D pattern. The functions produced from each sub-region of reciprocal space were then averaged to obtain a single estimate of the projected symmetrised autocorrelation in the corresponding orthogonal direction (Fig. 3a). While the windowing in reciprocal space reduces the spatial resolution of the determined autocorrelation map, this is still sufficient to identify the structural unit as an object of similar size to the photosystem II dimer. The assumption of the PSII dimer as the rigid structural unit was ultimately validated by the final electron density map obtained by phasing the continuous diffraction and improvements obtained in structural refinement (Extended Data Fig. 5), discussed below (another choice was also tested as described below). This result is not surprising as the dimer of PSII is the asymmetric unit of the crystals and is also the native biological oligomeric form, present in the photosynthetic membrane, maintained during the isolation procedure and in the process of crystallization.

The low-resolution structure from the Bragg data was obtained by molecular replacement phasing. The program F2MTZ was used to obtain an MTZ-file from the integrated intensities produced by *CrystFEL*. Structure-factor amplitudes were generated using TRUNCATE[44]. Molecular replacement phasing was performed using PHASER[45] in PHENIX[46] with the starting model being the photosystem-II monomer from a published 1.9 Å resolution structure[2] (PDB: 3WU2). All ligands and co-factors except lipids, detergents, water, and cryo-protectant molecules were kept in the input model. A composition containing the input model and its repeat was searched for and PHASER found a single solution. Three rounds of simple refinement were performed using *phenix.refine* (with the program options "rigid body", "XYZ coordinates" and "occupancies"). In the refinement the 3WU2 structure was used as a reference model to restrain relationships between atoms to avoid overfitting. No further model building was carried out after this step. This process generated R and R_{free} values of 24.8% and 27.2% respectively. A rendering of representative structural elements of PSII in the electron density map of the 4.5 Å resolution structure derived from Bragg peaks only is shown in Fig. 4 (with the model rendered in green).

The 4.5 Å resolution electron density map was used to generate a binary map of the support of the photosystem II dimer, shown in Fig. 3d, to be used for iterative phasing of the continuous diffraction. The inverse Fourier transform of the phased Bragg structure factors yields the unit cell electron density map. To get the electron density of a single PSII dimer from the unit cell, *phenix.cut_out_density* was used to select only that density which was at most 2 Å from any atom in the refined structural model of the PSII dimer. This density was Fourier transformed to obtain the squared modulus of the molecular transform up to $1/q =$ 4.5 Å, which was then symmetrised by the point group operations consisting of 180° rotations about each orthogonal axis and inversion through the origin (Extended Data Fig. 6). The similarity between this low-resolution computed continuous diffraction pattern and the measurements gives further confidence in the choice of the PSII dimer as the rigid structural unit. To avoid over-constraining the phasing of the continuous diffraction data (described below) a dilated binary support mask was generated from the PSII dimer electron density by convolving it with a Gaussian of width 4.4 Å (3 voxels) and thresholding the blurred density. Orthogonal slices through the support can be compared with the PSII dimer electron density image in Extended Data Fig. 7.

4. Iterative phasing

The 3D continuous diffraction was phased using the difference map algorithm[19]. This algorithm iteratively refines the electron density image using an update rule expressed in terms of so-called projection operators operating in a finite-dimensional vector space. This *N*-dimensional space spans all possible complex-valued electron density images in a 3D array of *N* voxels in total. A vector **x** in this space represents a particular electron density image. A projection of **x** on to a constraint set C, $\mathscr{P}_C(\mathbf{x})$, is the nearest point to **x** which satisfies the constraint C. The difference map update rule is

$$\mathbf{x} \rightarrow \mathbf{x} + \mathscr{P}_D\big(2\mathscr{P}_S(\mathbf{x}) - \mathbf{x}\big) - \mathscr{P}_S(\mathbf{x}), \quad (10)$$

where $\mathscr{P}_S(\mathbf{x})$ is the support projection operator and $\mathscr{P}_D(\mathbf{x})$ the data projection operator.

The support projection operator enforces the electron density to be zero outside the support volume (or low-resolution molecular envelope) of the PSII dimer described above. The data (or Fourier) projection operator enforces the electron density to be consistent with the measured Fourier magnitudes. In iterative phase retrieval of continuous coherent diffraction patterns this projection operation is usually carried out by Fourier transforming the current iterate *x* and setting the magnitude to the measured magnitude while keeping the phases from the iterate. Here, this projection was modified in two ways. First, the phases at low-resolution are provided by molecular replacement phasing, so for these voxels in reciprocal space the projection operator was set to simply replace the complex Fourier amplitude by these amplitudes and phases. Second, the point-group symmetries were applied to the squared Fourier magnitudes. This point group symmetrisation has been discussed by Elser & Millane[16]. In the absence of symmetry, the multiplicative factor applied to the Fourier amplitudes is $\sqrt{I_{\mathrm{obs}}}/|F_x|$ where F_x is the Fourier transform of the iterate and I_{obs} is the observed intensity. In the case of a symmetrised intensity, a similar factor is used, except that $|F_x|$ is replaced by the point group symmetrised version of itself.

Before applying the phasing algorithm, the molecular transform calculated from the phased Bragg intensities was scaled in magnitude and interpolated to match the merged continuous diffraction data. First, the low-resolution electron density grid was padded with zeros by an amount to approximately match the reciprocal space voxel size to that of the continuous data. The zero-padded volume was then Fourier transformed and stretched using linear interpolation to make the voxel sizes match exactly. The relative scaling of the magnitude of continuous diffraction to Bragg diffraction depends on the lattice disorder length and detector pixel size. This scaling was determined empirically at an intermediate q range (corresponding to the resolution range 5.23 Å – 4.58 Å) where both data types were non-zero and could be compared. The scale factor was chosen to minimise the sum of the squares of the intensity differences of the two data sets over the voxels in the resolution range.

Starting from a random white-noise guess for the voxel values of the electron density, the difference map algorithm converged within 100 iterations, retrieving over 30 million phases. The convergence was achieved in far fewer iterations than is usual in 3D coherent diffractive imaging problems of similar size[22], probably due to the initial well-estimated support and foreknowledge of the low-resolution phases. The final 3D electron density image of the PSII dimer was calculated by averaging the complex amplitudes over another 100 iterations. Any variations in retrieved phases at a particular voxel will reduce the sum in that voxel as compared with the sum of the moduli. The ratio of these sums, known as the phase retrieval transfer function (PRTF)[28], thus gives a measure of the consistency of the retrieved phases, and the complex sum naturally truncates the data where the phases are not known. In addition, this process was repeated with multiple random initial guesses. The solutions obtained were compared using a Fourier Shell Correlation (FSC) metric to confirm that there was no bias due to the choice of initial guess. Resolution-dependent plots of these metrics are shown in Fig. 3e, and the correlation between the continuous diffraction pattern and the symmetrised Fourier transform of the averaged iterate is shown in Extended Data Fig. 8a.

5. Model Refinement

In order to interpret the 3.5 Å resolution electron density volume image, and to compare with the lower-resolution refinement of the Bragg data, we refined a structural model to best match the diffractive image. Although it would make sense to refine a molecular model to the electron density map in real space, refinement was done in reciprocal space following a protocol recently established for interpreting cryo-electron microscopy maps using conventional crystallographic software[29]. The 3D electron density image (and hence the complex-valued Fourier amplitudes) was taken to be a crystal of P1 symmetry consisting of one PSII dimer in a lattice much larger than that of the real crystal. The phase probabilities, written in the form of Hendrickson–Lattman (HL) coefficients, were used for reciprocal space refinement which was performed against an MLHL target (maximum likelihood with experimental phase probability distribution) using PHENIX. Refinement statistics are shown in Extended Data Table 1.

6. Control analyses and comparisons

Two control analyses were performed to test whether the phasing of the continuous diffraction was an artefact of the input model. In one test the continuous diffraction was

replaced by its orientational average, constant for a given q magnitude. In this case, any improvement in the quality of the reconstruction would be purely an artefact of the support constraint. It was found that different random starts led to uncorrelated reconstructions, as shown by the FSC plot in Extended Data Fig. 8b. In the second test the rigid structural unit was assumed to be the photosystem-II-monomer instead of the PSII dimer. Not only is this physically unexpected, due to the stronger binding that forms the PSII dimer as compared with crystal contacts, but the symmetrised molecular transform of a monomer exhibits a larger speckle size and smaller extent of the autocorrelation than observed (data not shown).

A comparison of the electron density maps of the molecular structure are given in Fig. 4 and Extended Data Fig. 9, where improvements can be seen by including the continuous diffraction. In Extended Data Fig. 9, as well as showing the electron density obtained by iterative phasing of the continuous diffraction data we show the model density obtained by pseudo-crystal refinement. The correlation of the electron density maps[47] is given in Extended Data Fig. 9g, including a comparison with a previous high-resolution structure of Umena *et al.*[2] truncated to 3.5 Å resolution.

7. Constraint Ratio

The possibility to phase the diffraction pattern *ab initio*, such as by iterative methods, requires that the information content in the measurement exceeds that required to synthesise the electron density image. Such is not the case when measuring at Bragg peaks alone. Elser & Millane[16] discuss the feasibility of phase retrieval in terms of a quantity Ω, which represents the ratio of independent measurements to the number of unknown coefficients required to describe the electron density of an object at a given resolution. At a resolution $q = 1/d$ the object can be represented in real-space by samples spaced by distances $d/2$. The number of unknown coefficients to describe the object is equal to the volume that the object occupies divided by the voxel volume $d^3/8$. The three-dimensional region occupied by the object is known as the support of the object, S, with a volume given by $V(S)$. The number of independent Fourier space measurements is likewise determined by the volume of the support of the autocorrelation function of the object, since the autocorrelation function is equivalent (through a Fourier transform) to the measured diffraction intensities. The autocorrelation function is Hermitian, since it is the Fourier transform of the real-valued intensities, and thus the number of independent coefficients is equal to half the volume of the support of the autocorrelation function $V(A_S)$ divided by the same voxel volume $d^3/8$. Thus,

$$\Omega = \frac{V(A_S)}{2V(S)}. \quad (11)$$

In the absence of additional information, phase retrieval is feasible when $\Omega > 1$. As an example, consider an object that is shaped as a cuboid. The autocorrelation function is twice the length in each dimension, giving $V(A_S) = 8\ V(S)$ and $\Omega = 4$.

The continuous diffraction produced by a disordered crystal is given by the incoherent sum of the transforms $|F_A(\mathbf{R}_m\mathbf{q})|^2$ for the orientations m, as described in Eqn. (9). In this case the

number of independent measurements is equal to the volume of the union of the autocorrelation supports, $V(\cup_m \mathbf{R}_m A_S)$ divided by the number of symmetry operations of the crystal point group. In the case of photosystem II there are four molecular orientations. The support of the photosystem II dimer S is shown in Fig. 3d and one slice through the union of the autocorrelation supports is $\cup_m \mathbf{R}_m A_S$ given in Fig. 3b. The volume of this symmetrised support is 16.6 times larger than V_S, giving $\Omega = 2.1$ when considering the four orientations and Hermitian symmetry. This is lower than the case of only one orientation of the dimer where we find that the volume of the autocorrelation of the single support S of Fig. 3d is 10.9 times larger than V_S, so we would have $\Omega = 5.4$ in that case. The effect of the symmetry is therefore not just a matter of counting symmetry operations but it depends on the shape of the autocorrelation function compared with the shape of the symmetrised autocorrelation. We see that even though the symmetry reduces the ratio Ω it is still greater than unity, indicating the feasibility of phase retrieval from the continuous diffraction data alone.

Compared with the continuous diffraction pattern of a single object, Bragg data contains fewer independent measurements, reducing the ratio Ω. In a crystal one certainly knows that the object is contained in a volume equal to that of the unit cell, V_{cell}. However, the autocorrelation function (known in this case as the Patterson function) is periodic with a unit cell volume that is equal to that of the crystal, V_{cell}. Thus, the Hermitian Patterson function contains half the number of independent coefficients as the crystal unit cell itself, giving $\Omega = 1/2$. This is Sayre's result[2], eight times lower than for the single (non-crystalline) cuboid object mentioned above. Unlike the case of incoherent addition of the continuous diffraction of rigid units in the various crystallographic orientations, however, the constraint ratio for Bragg data does not depend on the space group. Although the unit cell volume becomes larger to accommodate more molecules in a crystal of higher symmetry, this does not lead to a change in the information content even though it gives a higher density of Bragg peaks. For the space group $P2_12_12_1$ of photosystem II, there are four asymmetric units per unit cell. Thus, the number of unknowns is reduced by a factor of 4. The number of independent measurements is reduced due to the symmetry of the Patterson function. In this case the symmetry of the Patterson function is P*mmm*, giving a factor of 4 reduction compared with P-1, leading once more to $\Omega = 1/2$.

Non-crystallographic symmetry does help for Bragg data, as is well known for the method of molecular replacement, since this reduces the number of unknown coefficients needed to describe the object without necessarily reducing the number of independent measurements. A better knowledge of the support of the object also helps in this case, allowing the application of solvent flattening (equivalent to applying the support constraint in iterative phasing). We can calculate the constraint ratio Ω for the photosystem II Bragg data, given the support S of Fig. 3d. In this case the number of independent measurements is proportional to the volume of unit cell, divided by four (the number of identical copies) and divided by two. Compared with the number of unknown coefficients, proportional to $V(S)$, we calculate

$$\Omega = \frac{V_{cell}}{8V(S)} = 0.86,$$

where the unit cell volume is determined from the lattice parameters provided in Extended Data Table 1. This suggests that the phase retrieval problem for this crystal is not solvable with just the Bragg data.
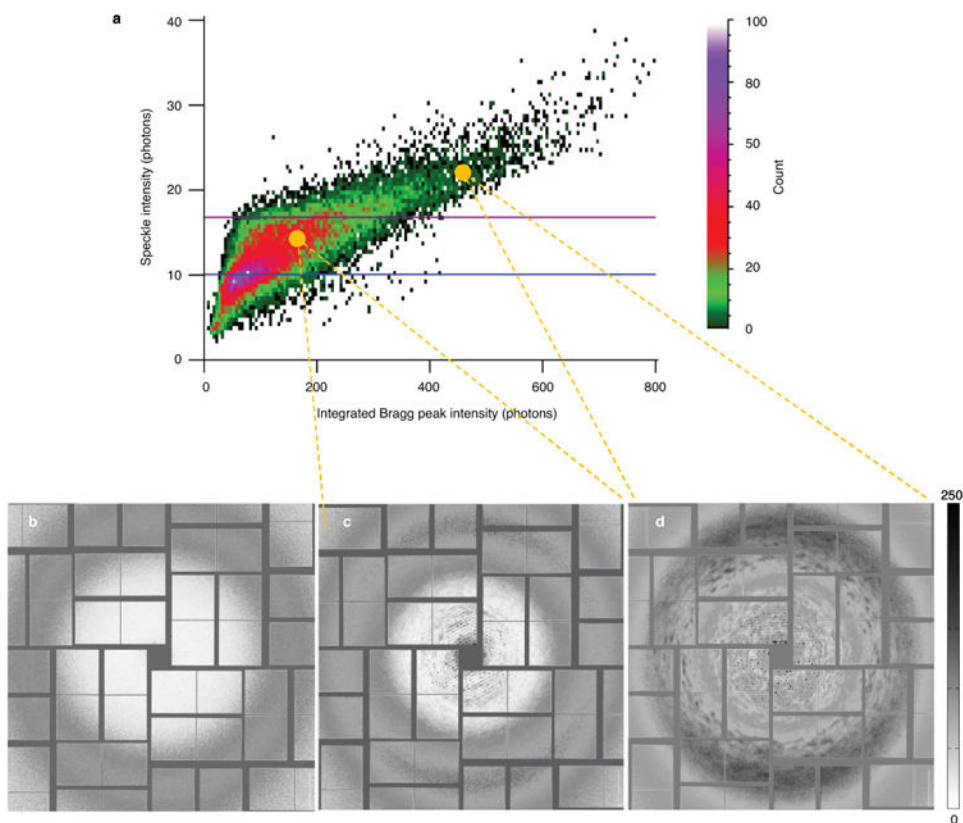
Extended Data
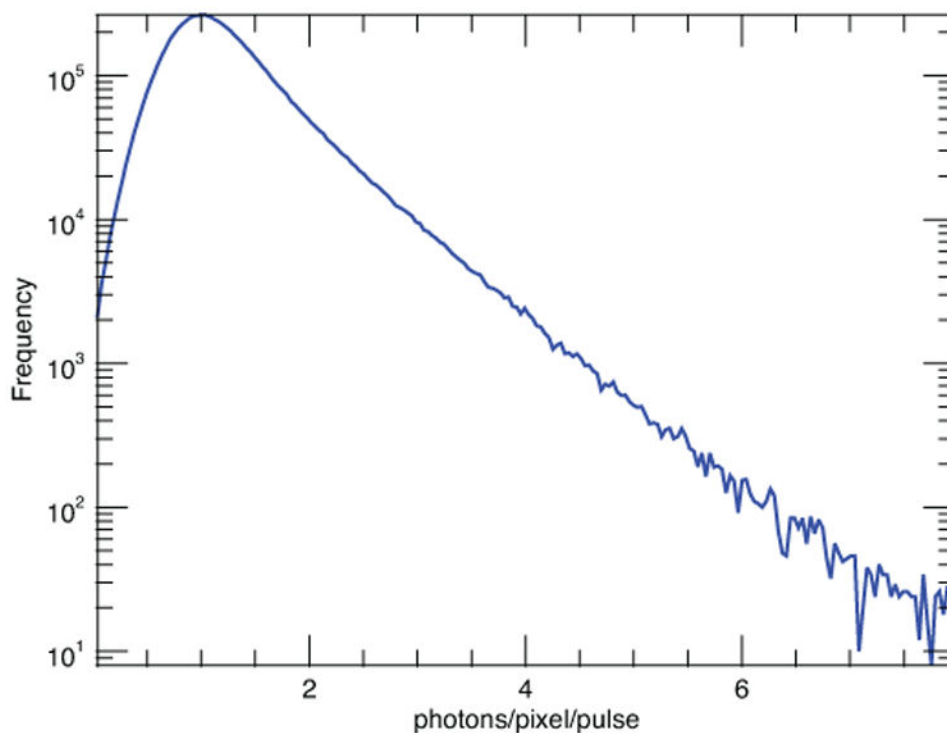
Extended Data Fig. 1. Continuous diffraction exhibits Wilson statistics
Histogram of merged continuous intensities in a q range of 0.22 Å^{-1} to 0.25 Å^{-1}. One can see that, above a background level of around 1 photon / pixel / pulse, the logarithm of the histogram follows a linear trend with negative slope characteristic of the exponential distribution predicted by Wilson statistics[43].
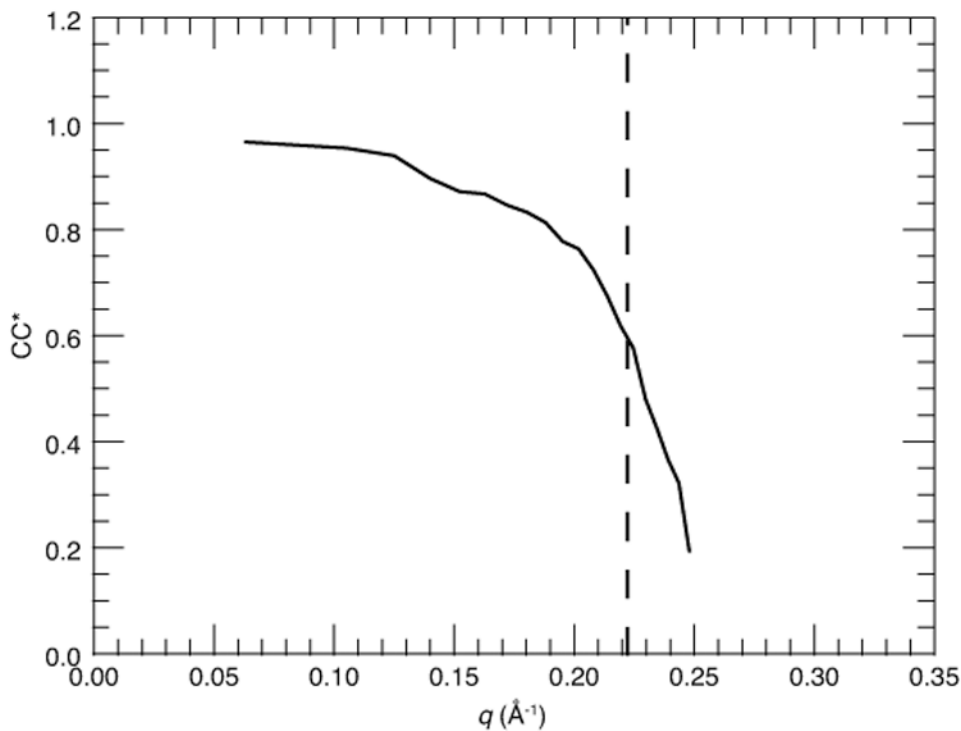
Extended Data Fig. 2. Data quality and resolution of Bragg diffraction
Plot of the reduced Pearson correlation coefficient, CC* (ref. 41) as an estimate of the consistency of the integrated Bragg intensities determined from 25,585 indexed patterns. A value of CC* = 0.5 is reached at $q = 0.23$ Å^{-1}, or a resolution of 4.3 Å.

Extended Data Fig. 3. Strongest continuous diffraction occurs with strongest Bragg diffraction
a, Two-dimensional histogram of patterns, sorted by the integrated counts in the continuous component of the diffraction pattern (in a q range of 0.22 Å^{-1} to 0.34 Å^{-1}) and the integrated signal in all detected Bragg peaks, for all 25,585 indexed patterns. We chose 2,848 patterns with the strongest continuous diffraction signal above 17 X-ray counts (purple line) in the q region to generate the 3D continuous pattern shown in Fig. 2c. The featureless background due to scattering from the solvent contributes 10 X-ray counts (blue line); the measurement from the liquid jet without crystals is shown in **b**. Two representative patterns with speckle counts above the mean solvent background but not above the threshold (**c**), and one of the strongest 2,848 patterns (**d**).
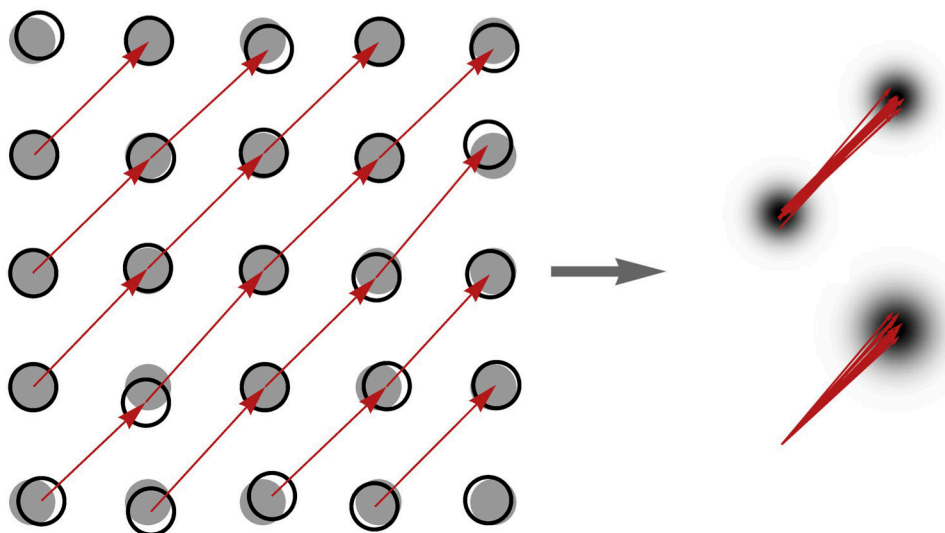
Extended Data Fig. 4. Patterson function of a disordered lattice
A distorted lattice (left, with ideal positions in gray), with vectors connecting all lattice points having difference vectors $\mathbf{a}_n - \mathbf{a}_K = (1,1)$. On the right, the arrows from the left are translated in two ways. Upper right: the heads and tails are both displaced from their ideal positions. Bottom right: the tails are lined up, resulting in the distribution of head positions forming a broader Gaussian. In the limit of a large crystal, the resulting distribution is just the autocorrelation of the displacement distribution. This process can be repeated for all difference vectors, leading to Eqn. (3).
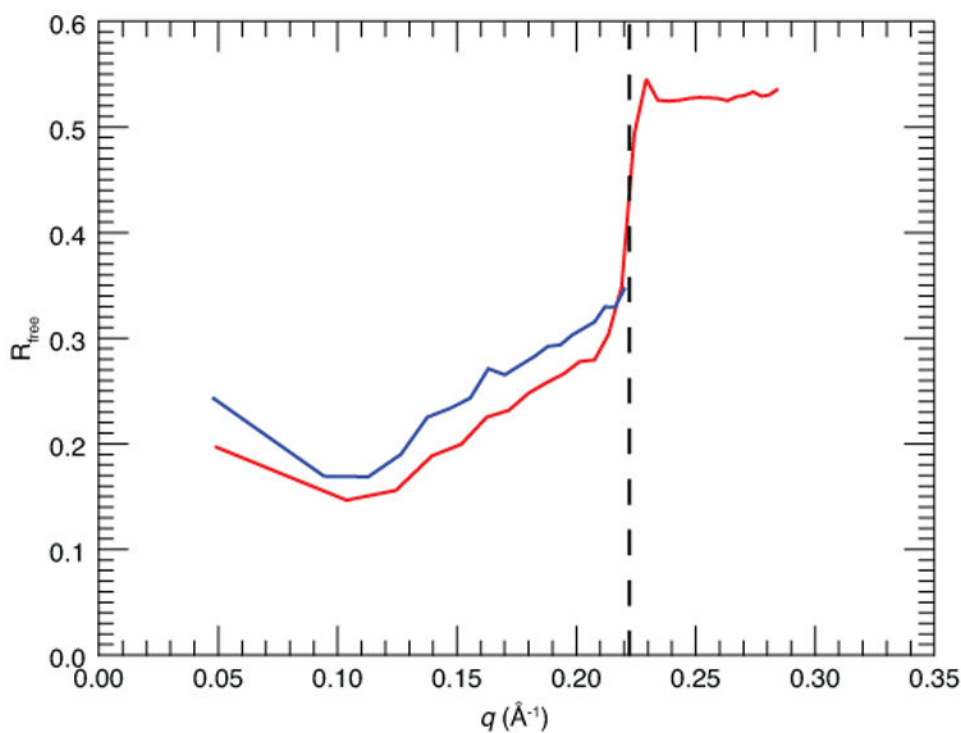
Extended Data Fig. 5. Model refinement is improved at low resolution

A plot of the metric *R*-free, as a function of resolution shell *q*, shows a marked improvement of the model refined against the 3.5 Å diffractive image. Here *R*-free is calculated only using Bragg intensities (which were excluded from the refinement) for resolution below 4.5 Å.
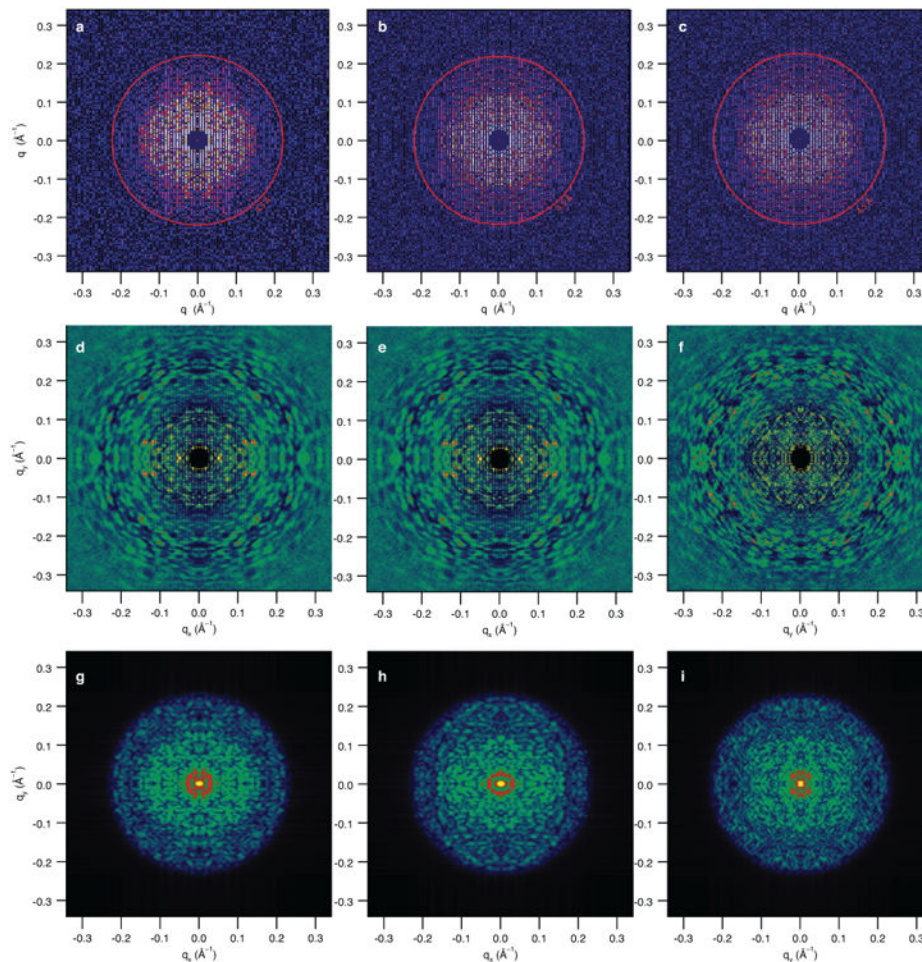

Extended Data Fig. 6. Central sections of the 3D full-pattern diffraction volume
a–c, Bragg intensities in planes normal to the three orthogonal reciprocal-space axes, q_z, q_y, and q_x, respectively, which were arbitrarily chosen to be parallel to the c*, b*, and a* axes of the PS II crystal. **d–f**, full-pattern diffraction intensities in central sections normal to the same three orthogonal axes, obtained from 2,848 strongest snapshot diffraction patterns. **g–i**, continuous diffraction intensities calculated for a single PSII dimer, from the model refined from the 4.5 Å Bragg data, for the same set of orthogonal planes as **a–c** and **d–f**. The intensities were calculated from the incoherent sum of the squared modulus of the 3D molecular transform of a single (uncrystallised) PSII dimer in each of the four orientations of the 222 point group. All panels are plotted on the same scale, with the experimental data (**d–f**) extending to $q = 0.33$ Å^{-1} at the centre edge. The agreement between each of **d–f** with the corresponding panel **g–i** is further evidence that the rigid structural unit is the photosystem II dimer.
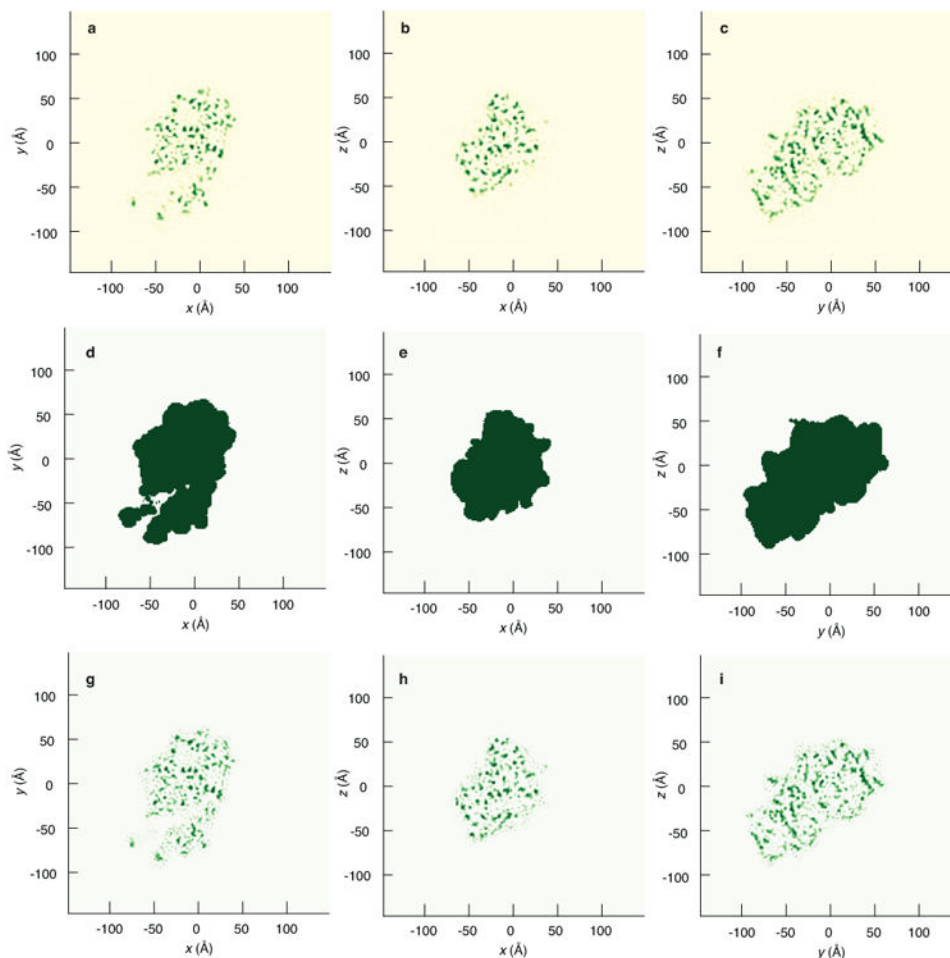
Extended Data Fig. 7. Real space orthogonal slices

a–c, slices of the F_o electron density map, plotted as a grey-scale, of a single photosystem II dimer obtained from the 4.5 Å Bragg intensities following model refinement of that data. Each slice is one pixel thick (1.5 Å) and is normal to the z, y, and x real-space axes, respectively (conjugate to q_z, q_y, and q_x). **d–f**, slices through the 3D real-space support constraint used to for iterative phasing. The support was generated by blurring the 4.5 Å resolution electron density map by 2.2 Å and then thresholding to achieve a binary mask. **g–i**, slices through the 3.5 Å resolution image obtained by iterative phasing of the continuous diffraction, using the support constraint illustrated in **d–f** and Fig. 3d.
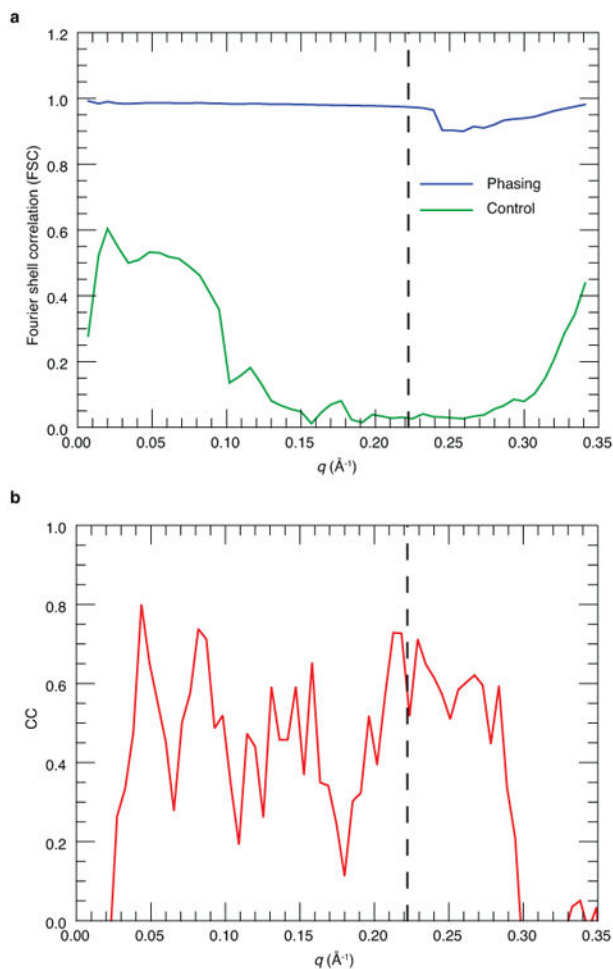
Extended Data Fig. 8. The continuous pattern is consistent with diffraction from a rigid object in crystallographic orientations

a, When the continuous diffraction intensities are substituted for intensities averaged over all orientations (i.e. constant on surfaces of constant q), iterative phasing using the support constraint of Fig. 3d fails, as indicated by this plot of the Fourier shell coefficient as a function of resolution for solutions (green) obtained from two independent phasing trials. The blue curve is the same FSC shown in Fig. 3e for comparison. The iterations never converged, so a PRTF for the control could not be generated. **b,** Plot of the correlation between the measured diffraction and that calculated from the determined electron density of a dimer, symmetrised by the four crystallographic orientations.
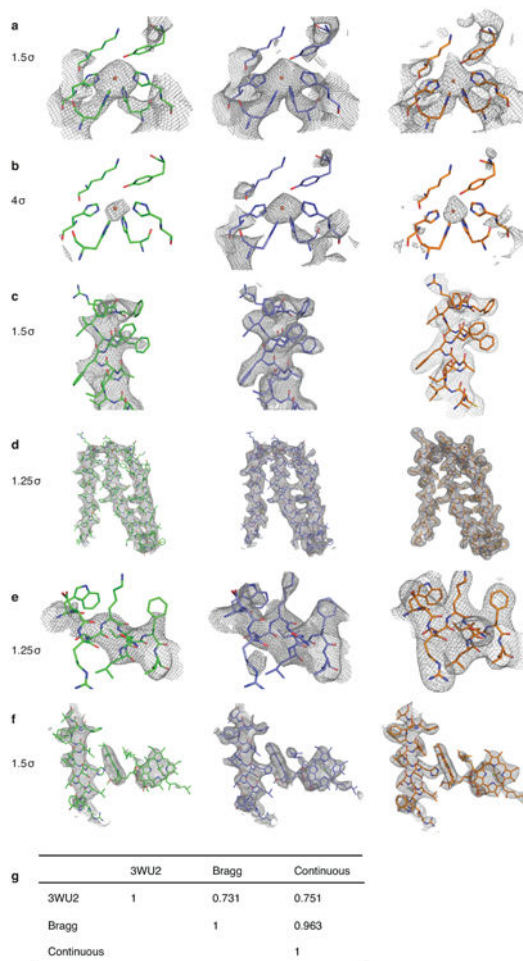
Extended Data Fig. 9. Electron density maps (2mFo-DFc) of regions of the PSII dimer
(**a**-**f**) Electron density maps obtained utilising the Bragg diffraction (with the model shaded green, left), the Bragg and continuous diffraction (model shaded blue, middle), and computed from pseudo-crystal refinement (model shaded brown, right). The maps are rendered at various density levels relative to the standard deviation of the overall density. **a** and **b**, Non-heme iron of chain A coordinated by four Histidin from chain A contoured at **a**, 1.5σ and **b**, 4σ. Neighboring Tyr and Lys residues are displayed as well. **c**, Part of an α–helix (chain T), showing that the side chains (e.g. Arg and Phe) are fitting better in the electron density when applying our new method (maps contoured at 1.5σ). **d**, helices from chains Y and Z (maps contoured at 1.5σ). The density map shows more details and better agreement to the model when applying the analysis using the Bragg and continuous diffraction. **e**, Detailed view of a section of chain Z, map contoured at 1.25σ. Utilising only the Bragg diffraction, no electron density is visible at this level around the side chains of Trp, Lys and Arg. Again, the model fits better into the map, when using the continuous diffraction. **f**, The primary electron donor chlorophyll dimer (P680) and part of the transmembrane helix of chain C, contoured at 1.5σ. **g**, Matrix of Pearson correlation coefficients of electron density maps obtained from the Umena model[2] (PDB code 3WU2) the model refined from Bragg diffraction, and that obtained from the continuous diffraction.

Extended Data Table 1
Data collection, phasing, and refinement

	Bragg diffraction	Continuous diffraction	Pseudo-crystal refinement
Point Group	222	222	1
Space Group	$P2_12_12_1$		P1
a, b, c (Å)	133.25, 226.26, 307.09	735.3, 735.3, 735.3	250.8 250.8 250.8
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution range (Å)	30–4.5 (4.62–4.5)	30–3.20[c]	30–3.5
Wilson B (Å^2)	191.6		
Indexed / Oriented patterns	25,585	2,848	
Completeness (%)	99.9		
I/sig(I)	5.01 (1.49)		
R-split (%)	11.48 (46.16)		
CC*	0.99 (0.94)		
Maximum dose (MGy)	275		
Refinement and phasing			
Method	ML[a]	Difference Map	MLHL[b]
No. reflections / phases	55,609	3.5×10^7	769,253
Resolution range (Å)	30.0–4.50 (4.58–4.50)	30.0–3.20[c]	A: 30.0–3.50 (3.54–3.50) B: 30.0–4.50 (4.55 – 4.50)
Rwork/Rfree (%)	24.8 / 27.2 (34.9 / 36.2)		A: 31.7 / 32.4 (53.2 / 54.2) B: 21.7 / 23.0 (39.6 / 40.6)
No. atoms	50,074		50,074
No. voxels		1.26×10^8	
Bond lengths (Å)	0.005		0.012
Bond angles (°)	1.400		1.817

[a]ML: Maximum likelihood

[b]MLHL: Maximum likelihood combined with experimental phase probability distribution.

[c]3.20 Å resolution in the centre edge of the 3D voxel array.

Acknowledgments

References

1. Heras B, Martin JL. Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Cryst D. 2005; 61:1173–1180. [PubMed: 16131749]

2. Umena Y, Kawakami K, Shen JR, Kamiya N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature. 2011; 473:55–60. [PubMed: 21499260]

3. Sayre D. Some implications of a theorem due to Shannon. Acta Cryst. 1952; 5:843.

4. Wall ME, Adams PD, Fraser JS, Sauter NK. Diffuse X-ray scattering to model protein motions. Structure. 2014; 22:182–184. [PubMed: 24507780]

5. Fienup JR. Phase retrieval algorithms: a comparison. Appl Opt. 1982; 21:2758–2769. [PubMed: 20396114]

6. Chapman HN, Nugent KA. Coherent lensless X-ray imaging. Nature Photon. 2010; 4:833–839.

7. Miao J, Ishikawa T, Robinson IK, Murnane MM. Beyond crystallography: Diffractive imaging using coherent x-ray light sources. Science. 2015; 348:530–535. [PubMed: 25931551]

8. Guinier, A. X-Ray Diffraction: In Crystals, Imperfect Crystals, and Amorphous Bodies. Dover; 1994.

9. Rupp, B. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. 1st. Garland Science; New York: 2009.

10. Doucet J, Benoit JP. Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. Nature. 1987; 325:643–646. [PubMed: 3808065]

11. Caspar DLD, Clarage J, Salunke DM, Clarage M. Liquid-like movements in crystalline insulin. Nature. 1988; 332:659–662. [PubMed: 3282173]

12. Faure P, et al. Correlated intramolecular motions and diffuse x-ray scattering in lysozyme. Nature Struct Biol. 1994; 1:124–128. [PubMed: 7656016]

13. Pérez J, Faure P, Benoit JP. Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by X-ray diffuse scattering. Acta Cryst D. 1996; 52:722–729. [PubMed: 15299635]

14. Wall ME, Ealick SE, Gruner SM. Three-dimensional diffuse X-ray scattering from crystals of Staphylococcal nuclease. Proc Nat Acad Sci. 1997; 94:6180–6184. [PubMed: 9177191]

15. Welberry TR, Heerdegen AP, Goldstone DC, Taylor IA. Diffuse scattering resulting from macromolecular frustration. Acta Cryst B. 2011; 67:516–524. [PubMed: 22101541]

16. Elser V, Millane RP. Reconstruction of an object from its symmetry-averaged diffraction pattern. Acta Cryst A. 2008; 64:273–279. [PubMed: 18285621]

17. Bates RHT. Fourier phase problems are uniquely solvable in more than one dimension: 1. Underlying theory. Optik. 1982; 61:247–262.

18. Fienup JR. Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. J Opt Soc Am A. 1987; 4:118–123.

19. Elser V. Phase retrieval by iterated projections. J Opt Soc Am A. 2003; 20:40–55.

20. Miao J, Charalambous P, Kirz J, Sayre D. Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens. Nature. 1999; 400:342–344.

21. Robinson IK, Vartanyants IA, Williams GJ, Pfeifer MA, Pitney JA. Reconstruction of the shapes of gold nanocrystals using coherent X-ray diffraction. Phys Rev Lett. 2001; 87:195505. [PubMed: 11690423]

22. Chapman HN, et al. High-resolution *ab initio* three-dimensional X-ray diffraction microscopy. J Opt Soc Am A. 2006; 23:1179–1200.

23. Kupitz C, et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. Nature. 2014; 513:261–265. [PubMed: 25043005]

24. Chapman HN, et al. Femtosecond X-ray protein nanocrystallography. Nature. 2011; 470:73–77. [PubMed: 21293373]

25. Boutet S, et al. High-resolution protein structure determination by serial femtosecond crystallography. Science. 2012; 337:362–364. [PubMed: 22653729]

26. White TA, et al. CrystFEL: a software suite for snapshot serial crystallography. J Appl Cryst. 2012; 45:335–341.

27. Yefanov O, et al. Mapping the continuous reciprocal space intensity distribution of x-ray serial crystallography. Phil Trans Roy Soc B. 2014; 369:1647.

28. Shapiro D, et al. Biological imaging by soft X-ray diffraction microscopy. Proc Nat Acad Sci. 2005; 102:15343–15346. [PubMed: 16219701]

29. Fischer N, et al. Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. Nature. 2015; 520:567–570. [PubMed: 25707802]

30. Spence JCH, Doak RB. Single molecule diffraction. Phys Rev Lett. 2004; 92:198102. [PubMed: 15169448]

31. Liang M, et al. The Coherent X-ray Imaging instrument at the Linac Coherent Light Source. J Synchr Rad. 2015; 22:514–519.

32. Emma P, et al. First lasing and operation of an ångstrom-wavelength free-electron laser. Nature Photon. 2010; 4:641–647.

33. DePonte DP, et al. Gas dynamic virtual nozzle for generation of microscopic droplet streams. J Phys D. 2008; 41:195505.

34. Weierstall U, Spence JCH, Doak RB. Injector for scattering measurements on fully solvated biospecies. Rev Sci Instrum. 2012; 83:035108. [PubMed: 22462961]

35. Neutze R, Wouts R, van der Spoel D, Weckert E, Hajdu J. Potential for biomolecular imaging with femtosecond x-ray pulses. Nature. 2000; 406:753–757.

36. Barty A, et al. Self-terminating diffraction gates femtosecond x-ray nanocrystallography measurements. Nature Photon. 2012; 6:35–40.

37. Hart, P., et al. The CSPAD megapixel x-ray camera at LCLS. in X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Applications. In: Stock, SR., editor. Proc SPIE. Vol. 8504. 2012. p. 85040C

38. Barty A, et al. Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. J Appl Cryst. 2014; 47:1118–1131. [PubMed: 24904246]

39. White TA, et al. Crystallographic data processing for free-electron laser sources. Acta Cryst D. 2013; 69:1231–1240. [PubMed: 23793149]

40. Yefanov, O., et al. Opt Express. 2015. Accurate determination of segmented X-ray detector geometry. submitted

41. Karplus PA, Diederichs K. Linking crystallographic model and data quality. Science. 2012; 336:1030–1033. [PubMed: 22628654]

42. Cowley, JM. Diffraction Physics. Elsevier; New York: 1981. p. 150-151.

43. Wilson AJC. The probability distribution of X-ray intensities. Acta Cryst. 1949; 2:318–321.

44. French GS, Wilson KS. On the treatment of negative intensity observations. Acta Cryst A. 1978; 34:517–525.

45. McCoy AJ, et al. Phaser crystallographic software. J Appl Cryst. 2007; 40:658–674. [PubMed: 19461840]

46. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D. 2010; 66:213–221. [PubMed: 20124702]

47. Urzhumtsev A, Afonine PV, Lunin VY, Terwilliger TC, Adams PD. Metrics for comparison of crystallographic maps. Acta Cryst D. 2014; 70:2593–2606. [PubMed: 25286844]
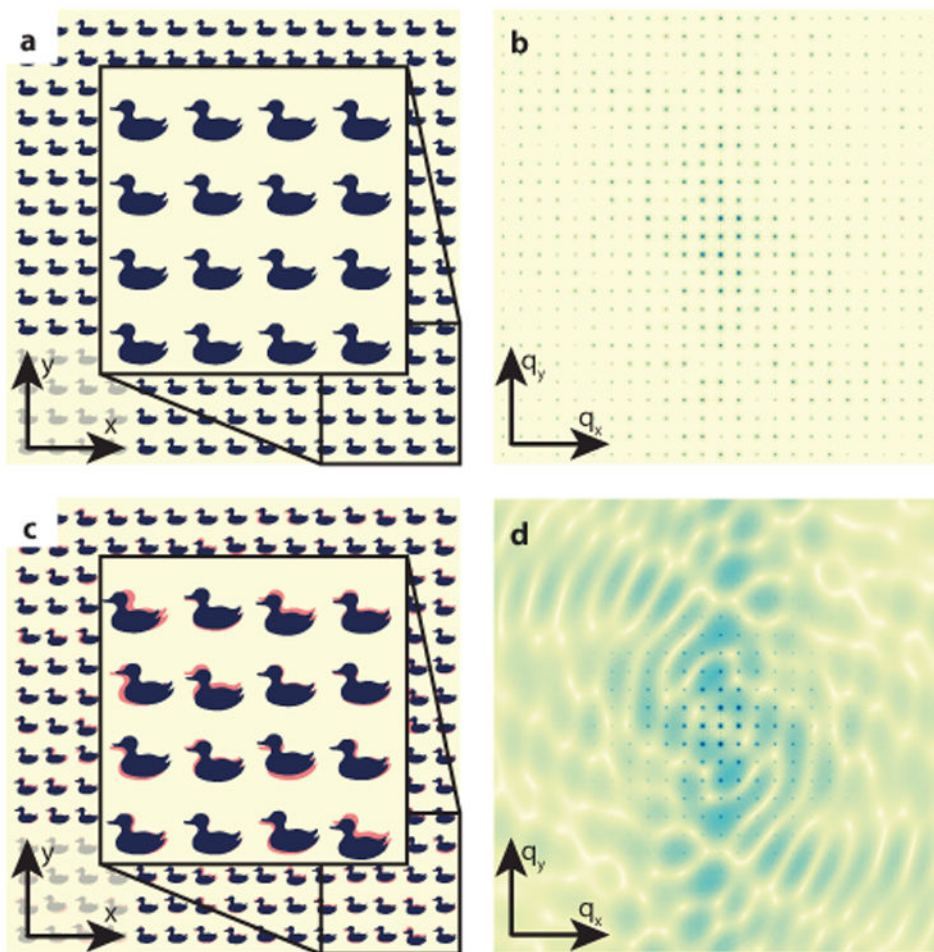
Figure 1. Lattice disorder reveals the continuous molecular transform

Diffraction from an arrangement of objects in a perfect lattice **a** results in regularly-spaced Bragg peaks through constructive interference **b**. Translational disorder of the objects at a length scale σ, **c**, disrupts Bragg interference beyond a reciprocal lattice resolution length $d = 1/q = 2\pi\,\sigma$, **d**. The loss of correlation gives rise to the incoherent sum of the Fraunhofer diffraction patterns of individual objects, which increases with q in balance with the diminishing Bragg intensities. Here, σ was chosen to be 4% of the lattice spacing, leading to reduction of Bragg peaks at the 4th order.
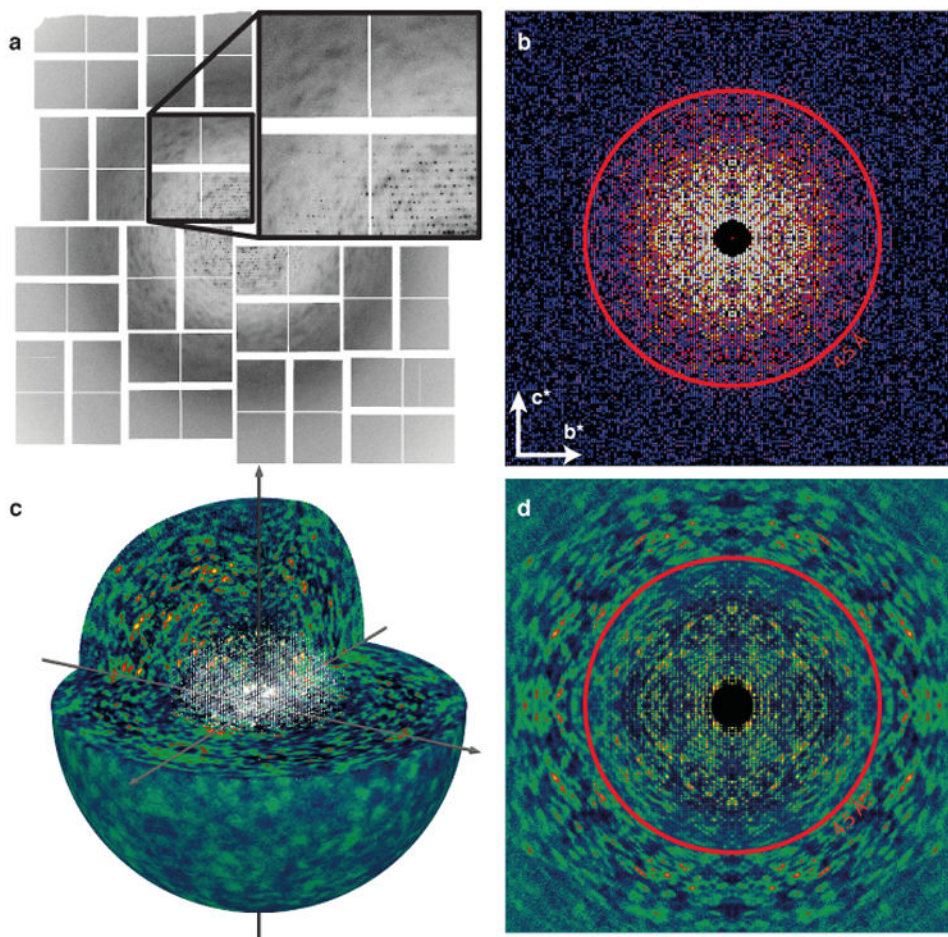
Figure 2. Molecular coherent diffraction
An X-ray FEL snapshot "still" diffraction pattern of a photosystem II microcrystal, **a**, shows a weak speckle structure beyond the extent of Bragg peaks which is enhanced in this figure by limiting the displayed pixel values. **b**, Structure factors obtained from Bragg-peak counts from 25,585 still patterns, displayed as a precession-style pattern of the [001] zone axis. **c**, A rendering of the entire 3D diffraction volume assembled from the 2,848 strongest patterns. **d**, A central section of the diffraction volume normal to the [100] axis. Speckles are clearly observed beyond the extent of Bragg diffraction to the edge of the detector.
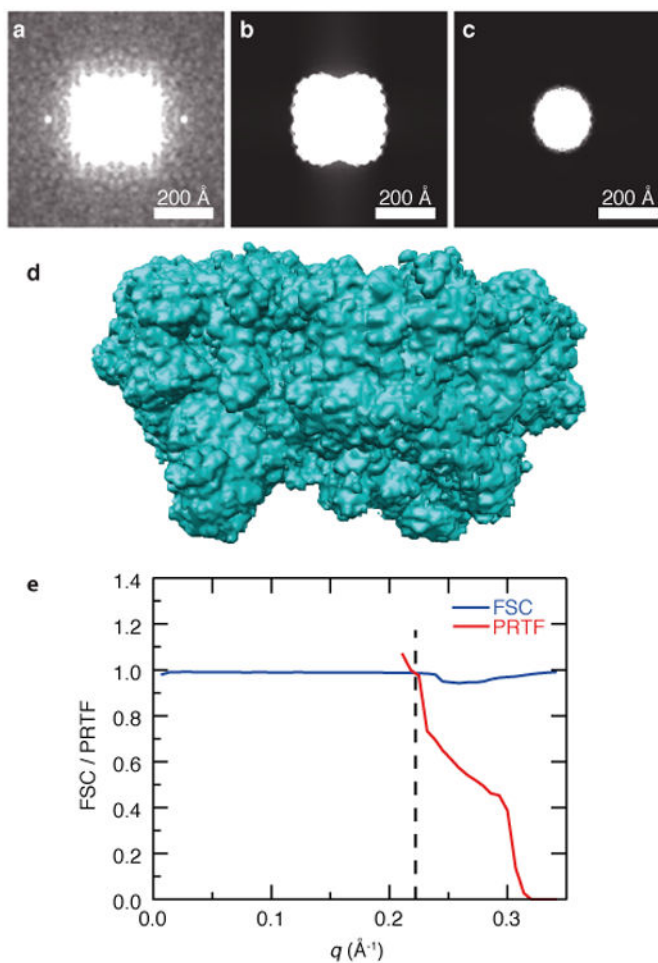
Figure 3.

An unbiased size estimate of the rigid structural unit is obtained by a Fourier transform of the continuous diffraction intensities, yielding the autocorrelation function (3D pair-distribution function. **a**, Projection of the experimentally-determined 3D autocorrelation along the crystal c axis. The equivalent projections through the autocorrelation functions calculated from the 4.5 Å model of photosystem II dimer, **b**, and the monomer, **c**, after applying the point group symmetries of the crystal. The extent of the rigid structural unit matches the size and shape of the PSII dimer. A loose support, **d**, generated by thresholding and dilating the 4.5 Å resolution structure was used as the support constraint for iterative phasing. The Fourier shell correlation (FSC) and phase retrieval transfer function (PRTF), **e**, indicates a resolution of 3.5 Å.
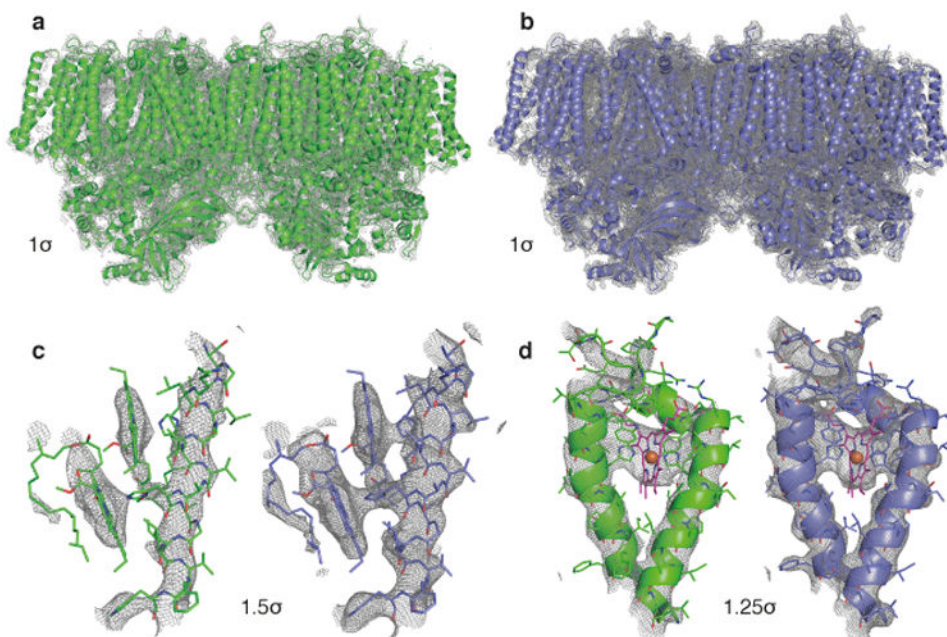
Figure 4.
Improvement in electron density resolution and quality is obtained by directly phasing the continuous transform by the method of coherent diffractive imaging. **a**, Electron density map of the PS II dimer after refinement using structure factors obtained only from Bragg peaks at 4.5 Å resolution. **b**, Electron density obtained by iterative phase retrieval on the continuous diffraction data using the support constraint (molecular envelope) of Fig. 3d at 3.5 Å resolution. **c** and **d**, electron density maps (2Fo-DFc) of two regions of the PSII dimer, utilising only the Bragg diffraction (with the model shaded green) and the Bragg and continuous diffraction (model shaded blue). **c**, Two antenna Chlorophylls in the antenna protein PsbC with their His ligands. **d**, The heme group of PsbE/F.