Macromolecular diffractive imaging using imperfect crystals

Kartik Ayyer1, Oleksandr M. Yefanov1, Dominik Oberthür2, Shatabdi Roy-Chowdhury3,4, Lorenzo Galli1,2, Valerio Mariani1, Shibom Basu3,4, Jesse Cee4,3, Chelsie E. Conrad4,3,4, Raimund Fromme3,4, Alexander Schaffer3,4, Katerina Dörner1,3, Daniel J. James4,5, Christopher Kupitz3,6, Markus Metz2, Garrett Nelson4,5, Paulraj Lourdu Xavier1,2, Kenneth R. Beyerlein1, Marius Schmidt6, Iosifina Sarrou1, John C. H. Spence4,5, Uwe Weierstall4,5, Thomas A. White1, Jay–How Yang3,4, Yun Zhao4,5, Mengning Liang8, Andrew Aquila8, Mark S. Hunter8, Joseph S. Robinson8, Jason E. Koglin8, Sébastien Boutet8, Petra Fromme3,4, Anton Barty1 & Henry N. Chapman1,2,9

The three-dimensional structures of macromolecules and their complexes are mainly elucidated by X-ray protein crystallography. A major limitation of this method is access to high-quality crystals, which is necessary to ensure X-ray diffraction extends to sufficiently large scattering angles and hence yields information of sufficiently high resolution with which to solve the crystal structure. The observation that crystals with reduced unit-cell volumes and tighter macromolecular packing often produce higher-resolution Bragg peaks1,2 suggests that crystallographic resolution for some macromolecules may be limited not by their heterogeneity, but 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 that is equal to the incoherent sum of diffraction from rigid individual molecular complexes aligned along several discrete crystallographic orientations and that, consequently, contains more information than Bragg peaks alone3. Although such continuous diffraction patterns have long been observed—and are of interest as a source of information about the dynamics of proteins4— 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 phase4 the pattern directly. Using the molecular envelope conventionally determined at 4.5 Å as a constraint, we obtain a static image of the photosystem II dimer at a resolution of 3.5 Å. This result shows that continuous diffraction can be used to overcome what have long been supposed to be the resolution limits of macromolecular crystallography, using a method that exploits commonly encountered imperfect crystals and enables model-free phasing6,7.

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 , then the phase of the diffracted wave from this unit is changed by 2πs/d at a scattering angle θ 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 (r.m.s.) displacement σ along a coordinate, then Bragg intensities will accordingly diminish at a resolution of d = 2πσ corresponding to the well known Debye–Waller factor 4πσ2 exp(−4π2σ2), where q = 1/d = 2sin(θ/λ) for an X-ray wavelength λ (ref. 8). The disorder length σ is easily estimated from a so-called Wilson plot9 of the diffraction data. Less widely appreciated is the fact that although the Bragg peaks diminish exponentially with the square of scattering angle, an incoherent sum of the diffraction of individual structural units arises contrariwise8,10,11 (see Fig. 1 and Methods section ‘The effect of rigid body disorder on Bragg and continuous diffraction’) 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 crystals10–13, and have even been ascribed to the disorder in the arrangement of rigid structural units as large as single macromolecules10–13. Although the continuous diffraction was seen to follow the point-group symmetry of the crystal14, 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 in which 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 equation (5) 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 resolution16,17 (see Methods section ‘Constraint ratio’). Bragg intensities can be equivalently represented in real space, by Fourier transformation, as the three-dimensional (3D) pair-correlation function (or Patterson function8) of the crystal. This correlation function cannot always distinguish between object pairs existing only in one unit cell or straddling neighbours, but 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 algorithms18–22 of coherent diffractive imaging and aligned molecule diffraction16.
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 (XFEL) experiments yielded Bragg diffraction from crystals with volumes 1 μm$^3$ 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 r.m.s. displacement of molecules by 0.8 Å in each dimension (1.4-Å r.m.s. 3D displacement; see Methods section ‘The effect of rigid body disorder on Bragg and continuous diffraction’). Bragg diffraction from large PSII crystals (1 mm$^3$ in volume) extending to 1.9 Å was recently reported, achieved by dehydrating the crystal in the presence of small amphiphiles, leading to an increase in crystal contacts and a tighter crystal packing. 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 those shown in Fig. 2a, were recorded from room-temperature microcrystals of PSII in random orientations, as per the method of serial femtosecond crystallography (see Methods section ‘Crystallography’). The orientation of each individual crystal was determined from its diffraction pattern by indexing the Bragg peaks, allowing it to be inserted into a 3D array to build up the full intensity distribution 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 the lowest resolutions and become dominant as a function of 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-correlation function), which is obtained by a Fourier transform of the continuous diffraction intensities. We found it 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, and the second due to the incoherent addition of aligned single-molecule diffraction (see Methods section ‘Two data sources from one experiment’). Starting with a known search model for PSII, we first produced an electron density map of the PSII dimer by molecular replacement phasing to a resolution of 4.5 Å, 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 PSII dimer, shown in Fig. 3d. This mask forms the real-space constraint used to generate a 3D image of the electron density of the PSII dimer by iterative phase refinement of the continuous diffraction.

The iterative phasing of the continuous diffraction, covering a resolution range of 4.5–3.3 Å, was carried out using the difference-map...
algorithm\textsuperscript{16,19} (Methods section 'Iterative phasing'). 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 synthesize a 3.3-Å resolution structure. Averaging solutions obtained over multiple random starts produced a self-consistent electron density to a resolution of 3.5 Å, as determined by the Fourier shell correlation and the phase retrieval transfer function\textsuperscript{28} (Fig. 3e). Finally, pseudo-crystallographic refinement\textsuperscript{29} of this electron density was carried out, following a similar procedure as for single-particle cryo-electron microscopy data (Methods section 'Model refinement'). 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 $R_{\text{free}}$ factor at low resolution (Extended Data Fig. 1). Although the assembled continuous diffraction data are quite noisy, owing to the limited number (2,848) of diffraction patterns, the structure clearly benefits from their inclusion. Control experiments verified these benefits (see Methods section 'Control analyses and comparisons'). Further improvements could be expected with more measurements.

Our approach, demonstrated with serial diffraction data collected using 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 solely on the basis of their diffraction patterns. The improved structural refinement at a resolution beyond that of measurable Bragg peaks establishes the origin of the continuous diffraction and demonstrates its utility in obtaining substantial 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\textsuperscript{30}. We anticipate that the approach will be applicable to a range of macromolecular systems, and may allow \textit{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.

**Figure 2 | Molecular coherent diffraction.** a, An XFEL snapshot 'still' diffraction pattern of a PSII microcrystal 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 in c normal to the [100] axis. Speckles are clearly observed beyond the 4.5-Å extent of Bragg diffraction (indicated by the white circles in b and d) to the edge of the detector.
Figure 3 | An unbiased size estimate of the rigid structural unit. This estimate 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. b, c, The equivalent projections of the autocorrelation functions calculated from the 4.5-Å model of the PSII 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. d, A loose support generated by thresholding and dilating the 4.5-Å-resolution structure was used as the support constraint for iterative phasing. e, Plots of the Fourier shell correlation (FSC; blue) and phase retrieval transfer function (PRTF; red). A value of PRTF = 0.5 is reached at a resolution of 3.5 Å. The dashed line corresponds to the boundary between the phased Bragg and the continuous diffraction.

Figure 4 | Improvement in resolution and quality of electron density. This improvement is obtained by directly phasing the continuous transform using the method of coherent diffractive imaging. a, Electron density map of the PSII dimer after refinement using structure factors obtained only from Bragg peaks at 4.5-Å resolution, contoured at one standard deviation of the densities (1σ). b, Electron density map obtained by iterative phase retrieval on the continuous diffraction data using the support constraint (molecular envelope) of Fig. 3d at 3.5-Å resolution, contoured at 1σ. c, d, Electron density maps of two regions of the PSII dimer, using only the Bragg diffraction (models shaded green), and the Bragg and continuous diffraction (models shaded blue). c, Two antenna chlorophylls in the antenna protein PsbC with their His residues, contoured at 1.5σ. d, The haem group of PsbE/F, contoured at 1.25σ.
1. Heras, B. & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. Acta Crystallogr. D 61, 1173–1180 (2005).

2. Umemura, Y., Kawakami, K., Shen, J.-R. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature 473, 55–60 (2011).

3. Sayre, D. Some implications of a theorem due to Shannon. Acta Crystallogr. A 5, 843 (1952).

4. Wall, M. E., Adams, P. D., Fraser, J. S. & Sauter, N. K. Diffuse X-ray scattering to model protein motions. Structure 22, 182–184 (2014).

5. Fienup, J. R. Phase retrieval algorithms: a comparison. Appl. Opt. 21, 2758–2769 (1982).

6. Chapman, H. N. & Nugent, K. A. Coherent lensless X-ray imaging. Nature Photon. 4, S33–S39 (2010).

7. Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. Beyond crystallography: diffractive imaging using coherent x-ray light sources. Science 348, 530–535 (2015).

8. Guinier, A. X-Ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies Ch. 7 (Dover, 1994).

9. Rupu, B. Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology Ch. 7, 10 (Garland Science, 2009).

10. Doucet, J. & Benoit, J. P. Molecular dynamics studied by analysis of the X-ray diffuse scattering from lysozyme crystals. Nature 325, 643–646 (1987).

11. Caspar, D. L. D., Clargue, J., Salinque, D. M. & Clargue, M. Liquid-like movements in crystalline insulin. Nature 332, 659–662 (1988).

12. Faure, P. et al. Correlated intramolecular motions and diffuse X-ray scattering in lysozyme. Nature Struct. Biol. 1, 124–128 (1994).

13. Pérez, J., Faure, P. & Benoit, J. P. Molecular rigid-body displacements in a tetragonal lysozyme crystal confirmed by X-ray diffuse scattering. Acta Crystallogr. D 52, 722–729 (1996).

14. Wall, M. E., Ealick, S. E. & Gruner, S. M. Three-dimensional diffuse X-ray scattering from crystals of Staphylococcal nuclease. Proc. Natl Acad. Sci. USA 94, 6180–6184 (1997).

15. Welberry, T. R., Heerdegan, A. P., Goldstone, D. C. & Taylor, I. A. Diffuse scattering resulting from macromolecular frustration. Acta Crystallogr. B 67, 516–524 (2011).

16. Elser, V. & Milane, R. P. Reconstruction of an object from its symmetry-averaged diffraction pattern. Acta Crystallogr. A 64, 273–279 (2008).

17. Bates, R. H. T. Fourier phase problems are uniquely solvable in more than one dimension. I: underlying theory. Optik 61, 247–262 (1982).

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

19. Elser, V. Phase retrieval by iterated projections. J. Opt. Soc. Am. A 20, 40–55 (2003).

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

21. Robinson, I. K., Vartanyants, I. A., Williams, G. J., Pfeifer, M. A. & Pitney, J. A. Reconstruction of the shapes of gold nanocrystals using coherent X-ray diffraction. Phys. Rev. Lett. 87, 195505 (2001).

22. Chapman, H. N. et al. High-resolution ab initio three-dimensional X-ray diffraction microscopy. J. Opt. Soc. Am. A 23, 1179–1200 (2006).

23. Kupitz, C. et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. Nature 513, 261–265 (2014).

24. Chapman, H. N. et al. Femtosecond X-ray protein nanocrystallography. Nature 470, 73–77 (2011).

25. Boult, S. et al. High-resolution protein structure determination by serial femtosecond crystallography. Science 337, 362–364 (2012).

26. White, T. A. et al. CrystFEL: a software suite for snapshot serial crystallography. J. Appl. Cryst. 45, 335–341 (2012).

27. Yefanov, O. et al. Mapping the continuous reciprocal space intensity distribution of X-ray serial crystallography. Phil. Trans. R. Soc. B 369, 20130333 (2014).

28. Shapiro, D. et al. Biological imaging by soft X-ray diffraction microscopy. Proc. Natl Acad. Sci. USA 102, 15343–15346 (2005).

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

30. Spence, J. C. H. & Doak, R. B. Single molecule diffraction. Phys. Rev. Lett. 92, 10102 (2004).

Acknowledgements We acknowledge support of the Helmholtz Association through project-oriented funds and the Virtual Institute “Dynamic Pathways in Multidimensional Landscapes”; the DFG through the Gottfried Wilhelm Leibniz Program; the European Research Council under the European Union’s Seventh Framework Programme ERC Synergy Grant 609920 “AXIS” and Marie Curie FP7-PEOPLE-2011-ITN Grant 317079 “Nanoem”; the BMBF through Project 09E13GUL; the Graduate College “GRK 1355” at the University of Hamburg, the International Max Planck Research School UFAST, the BioXFEL Science Technology Center (award 1231306); and the US National Institutes of Health (NIH), National Institute of General Medical Sciences grants R01 GM095583, U54 GM094599, and R01 GM07463. Parts of the sample injector used at the Linac Coherent Light Source (LCLS) for this research was funded by the NIH, P41 GM103393, formerly P41RR021209. Use of the LCLS, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract number DE-AC02-76SF00515.

Author Contributions H.N.C. conceived the idea of molecular imaging using imperfect crystals, and developed it with K.A., O.M.Y., D.O., M.M., K.R.B. and L.G. K.A., O.M.Y., D.O., L.G. and V.M. carried out the diffusive imaging analysis. P.F. led the photosystem II project. Serial femtosecond crystallography measurements were led by P.F. and A.B. and performed with S.R.-C., J.C., D.O., M.A., K.S., D.J., G.N., U.W., Y.Z. and M.S. M.L., A.A., M.S.H., J.S.R., J.E.K. and S.Bo. set up and operated the CXI instrument. S.Ba., R.F., A.B. and M.M. performed on-site data evaluation at LCLS. S.Ba. and R.F. provided the initial search model and carried out model refinement on the atomic model obtained by molecular replacement from the Bragg peaks alone, and with the inclusion of continuous diffraction, have been deposited into the Protein Data Bank under accession codes 5E7C and 5E79, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.N.C. (henry.chapman@desy.de).

Author Information The coordinates of the atomic model obtained by molecular replacement from the Bragg peaks alone, and with the inclusion of continuous diffraction, have been deposited into the Protein Data Bank under accession codes 5E7C and 5E79, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.N.C. (henry.chapman@desy.de).
diffraction is equal to any particular snapshot in the limit of a large number of unit and, for uncorrelated motions of structural units in a crystal, the time-averaged considered as the incoherent sum of static snapshots over the time of the exposure, giving rise to a component of protein dynamics4,10–14. The time-averaged diffraction of a moving object can be computed by summing counts along circles of constant scattering from the water. The strength of the background depends linearly on 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 a strong linear correlation between these two signals. The incident Bragg-peak signal in Extended Data Fig. 3 for all of the 25,585 indexed patterns. The volume consisted of $501^3$ cubic voxels, each with a width of $1.364 \times 10^{-3} \text{Å}^{-1}$ ($q = 1/d$). 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 substantial 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 polarization 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 $5 \times 5$-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 $3 \times 3$-pixel distance of them, were replaced with the values from the $5 \times 5$-pixel median-filtered copy. The strength of the continuous diffraction signal for each frame was best estimated from pixels in the range of $0.1$–$0.2 \text{Å}^{-1}$ 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 signal intensities. The incident intensity for each snapshot diffraction pattern fluctuates, owing 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 all 25,585 patterns. 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 dynamics10–14. 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-averaged diffraction measurements cannot distinguish static disorder from dynamic. In the case of XFEL exposures, the positions of structural units such as entire macromolecules are certainly static. Following the approach of ref. 42, we first derive the diffraction of a translationally disordered crystal containing only one rigid object per unit cell (that is, PI symmetry). The density of this crystal in terms of the real-space coordinate $x$, $\rho_i(x)$, can be described by

$$\rho_i(x) = \rho_i(x) \otimes \sum_{n=1}^{N} \delta(x-a_n-a_i)$$

where $\rho_i(x)$ is the density of the asymmetric unit, repeated over $N$ positions in the crystal given by random displacement vectors $\Delta_n$, from ideal lattice positions $a_n$. In equation (1), $\delta(x)$ represents a convolution and $\Delta_n$ is the Dirac delta function. Without loss of generality, the mean displacement $\langle \Delta_n \rangle = 0$, and $\langle \Delta_n \rangle^2 = \sigma^2$.

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

$$P(x) = \rho_i(x) \otimes \sum_{n=1}^{N} \sum_{k=1}^{N} \delta(x-a_n-a_k) \otimes \sum_{k'=1}^{N} \delta(x+a_k-a_{k'})$$

The double sum in equation (2) can be understood for large crystals in the following manner. A constant lattice difference vectors $a_n-a_k$ 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 auto-correlation of the probability distribution of a single displacement. For a Gaussian with a standard deviation of $\sigma$, this is just another Gaussian with a standard deviation of $\sqrt{2}\sigma$. Thus

$$P(x) = \rho_i(x) \otimes \sum_{n=1}^{N} \sum_{k=1}^{N} \delta(x-a_n-a_k)$$

Keeping in mind that the Fourier transform ($F$) maps convolutions to multiplications, the intensity can be calculated from equations (2) and (3) as

$$I(q) = |F(q)|^2 \cdot \left| \sum_{n=1}^{N} \sum_{k=1}^{N} \delta(x-a_n-a_k) \right|$$

$$= \left| F(q) \right|^2 \cdot \left| N \langle \rho_i(x) \rangle + \exp(-x^2/4\sigma^2) \sum_{n=1}^{N} \sum_{k=1}^{N} \delta(x-a_n-a_k) \right|$$

$$= \left| F(q) \right|^2 \cdot \left| N \langle \rho_i(x) \rangle + \exp(-x^2/4\sigma^2) \sum_{n=1}^{N} \sum_{k=1}^{N} \delta(x-a_n-a_k) \right|$$

where $F(q)$ is the (complex-valued) Fourier transform of $\rho_i(x)$ at a resolution of $|q| = 1/d$. 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 $F(q)$, modulated by two terms. The second term is the reciprocal lattice, which gives the Bragg peaks, further modulated by the Debye–Waller factor, $\exp(-4\pi^2q^2\sigma^2$). The first term has no lattice associated with it, but is a monotonically increasing function with a value of 0 at $q = 0$ that asymptotes to a constant.
for \( q > 1/(2\pi\sigma) \). Thus, at high \( q \) values, the diffraction is proportional to the continuous diffraction pattern of the rigid unit, \( |F(x)|^2 \).

Generalizing further to the case of \( M \) asymmetric units in a unit cell, one can write the unit cell electron density as

\[
\rho_i(x) = \sum_{m=1}^{M} \sum_{n=1}^{N} \rho_i(R_{m}x + t_m) \otimes \delta(x - a_n - \Delta_{m,n})
\]

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

The Patterson function of \( \rho_i(x) \) 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

\[
P(x) = \sum_{m=1}^{M} \sum_{n=1}^{N} \rho_i(R_{m}x + t_m) \otimes \delta(x - a_n - \Delta_{m,n}) \otimes \delta(x + a_k + \Delta_{j,k})
\]

\[
= \sum_{m=1}^{M} \sum_{n=1}^{N} \rho_i(R_{m}x + t_m) \otimes \delta(R_{m}x - t_m) \otimes \sum_{n} \delta(x - a_n - \Delta_{m,n})
\]

\[
\otimes \sum_{j,k} \delta(x + a_k + \Delta_{j,k})
\]

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

\[
I(q) = \sum_{m=1}^{M} |F_m(x)|^2 \left[ N \left(1 - e^{-4\pi^2q^2\sigma^2}\right) + e^{-4\pi^2q^2\sigma^2} \sum_{n=1}^{N} e^{2\pi i (a_n - a_0) \cdot q} \right]
\]

\[
+ \sum_{m=1}^{M} \sum_{n=1}^{N} F_m(R_{m}x) F_n(x) e^{2\pi i (x - B_{m,n}) \cdot q} \left[ e^{-4\pi^2q^2\sigma^2} \sum_{n=1}^{N} e^{2\pi i (a_n - a_0) \cdot q} \right]
\]

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

\[
I(q) = N \sum_{m=1}^{M} |F_m(x)|^2 \left[ 1 - e^{-4\pi^2q^2\sigma^2}\right]
\]

\[
+ \sum_{m=1}^{M} \sum_{n=1}^{N} F_m(R_{m}x) e^{2\pi i q \cdot t_m} \left[ e^{-4\pi^2q^2\sigma^2} \sum_{n=1}^{N} e^{2\pi i (a_n - a_0) \cdot q} \right] (5)
\]

This equation has a similar form to equation (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, whereas 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, whereas the continuous diffraction depends only 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 \( e^{-4\pi^2q^2\sigma^2}\) 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 generalized to a number of non-identical rigid structural units in the unit cell.

Two data sources from one experiment. Structure determination could be considered in terms of a whole-pattern analysis, possibly based on equation (5). An alternative approach here was to treat the Bragg diffraction and continuous diffraction as two separate data sources that both encode structural information of the molecular system. The observable Bragg diffraction covers the range 0.03–0.22 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 0.22–0.35 Å−1 for this data set. Our strategy for using 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 equation (5), which predicts that the continuous diffraction is proportional to the incoherent sum of transforms of single rigid units aligned in various crystallographic orientations, we expect the statistics of the continuous diffraction intensities to follow that of molecular diffraction, which should be exponentially distributed. This can be seen in Extended Data Fig. 5, which shows a linear trend in the logarithmic histogram above a background level of around one 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 equation (5), this yields the incoherent sum of autocorrelation functions of the structural unit in various crystallographic orientations. To avoid the influence of the Bragg peaks on this calculation, this symmetrized autocorrelation function was generated from windowed regions of the full 3D pattern above a resolution of \( 1/q = 4.5 \) Å, following a procedure previously described. Using a Gaussian window function with a 15-pixel radius, 30 autocorrelation functions were generated in different locations in an orthogonal central section of the full 3D pattern. The functions produced from each subregion of reciprocal space were then averaged to obtain a single estimate of the projected symmetrized autocorrelation in the corresponding orthogonal direction (Fig. 3a). Although the windowing in reciprocal space reduces the spatial resolution of the determined autocorrelation map, it is still sufficient to identify the structural unit as an object of similar size to the PSII 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 by improvements obtained in structural refinement (Extended Data Fig. 1, discussed below). Another choice was also tested as described below. This result is not surprising because 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 EMT2Z was used to obtain an MTZ file from the integrated intensities produced by CrystFEL. Structure-factor amplitudes were generated using TRUNCATE. Molecular replacement phasing was performed using PHASER in PHENIX with the starting model being the PSII monomer from a published 1.9-Å-resolution structure (PDB, 3WU2). All ligands and cofactors 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 (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 PSII dimer (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 obtain the electron density of a single PSII dimer from the unit cell, phenix.cut_out_density was used to select the integrated density that 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 symmetrized 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 convoluting it with a Gaussian of width 4.4 Å (three 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.

Iterative phasing. The 3D continuous diffraction was phased using the difference-map algorithm. This algorithm iteratively refines the electron density image using an update rule that is 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 onto a constraint set C, \( P_C(x) \), is the nearest point to x that satisfies the constraint C. The difference map update rule is

\[
\mathbf{x} \rightarrow \mathbf{x} + \mathcal{P}_\Omega(2\mathcal{P}_S(x) - \mathbf{x}) - \mathcal{P}_S(x)
\]

where \( \mathcal{P}_S(x) \) is the support projection operator and \( \mathcal{P}_\Omega(x) \) 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 amplitudes by these amplitudes and phases. Second, the point-group symmetries were applied to the squared Fourier magnitudes. This point-group symmetrization has been discussed in ref. 16. In the absence of symmetry, the multiplicative factor applied to the Fourier amplitudes is \( \mathcal{P}_S(x) \), where \( \mathcal{P}_S(x) \) is the Fourier transform of the iterate and \( \mathcal{P}_S(x) \) the observed intensity. In the case of a symmetrized intensity, a similar factor is used, except that \( |\mathcal{P}_S(x)|^2 \) is replaced by the point-group symmetrized 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 is dependent on the lattice disorder length and detector pixel size. This scaling was determined empirically at an intermediate q range (corresponding to the resolution range 5.23–4.58 Å) in which both data types were non-zero and could be compared. The scale factor was chosen to minimize 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, probably owing 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, 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 Extended Data Fig. 3e, and the correlation between the continuous diffraction pattern and the symmetrized Fourier transform of the averaged iterate is shown in Extended Data Fig. 8a.

Model refinement. 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 diffraction 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. 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.

Control analyses and comparisons. Two control analyses were performed to test whether the phasing of the continuous diffraction was an artefact of the input data or arose through the continuous integer linear function was an average of a constant function multiplied by \( \frac{1}{3} \) for the orientations \( \Omega > 1 \), and so \( \Omega > 1 \).

\[ \Omega = \frac{V(A^k)}{2V(S)} \]

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 \( V(A^k) = 8V(S) \) and \( \Omega = 4 \).

The continuous diffraction produced by a disordered crystal is given by the incoherent sum of the transforms \( |F_i(m)|^2 \) for the orientations \( m \) as described in equation (5). In this case, the number of independent measurements is equal to the number of unknown coefficients required to describe the electron density of an object at a given resolution, is discussed in ref. 16. At a resolution of \( q = \frac{1}{4d} \), the object can be represented in real space by samples spaced by distances of \( d/2 \). The number of unknown coefficients to describe the object is equal to the volume that the object occupies divided by the voxel volume \( \frac{d^3}{8} \). The 3D 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, because the autocorrelation function is equivalent (through a Fourier transform) to the measured diffraction intensities. The autocorrelation function is Hermitian, because 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 \( V(A^k) \) divided by the same voxel volume \( \frac{d^3}{8} \). Thus

\[ \Omega = \frac{V(A^k)}{2V(S)} \]
crystal, $V_{\text{cell}}$. Thus, the Hermitian Patterson function contains half the number of independent coefficients as does the crystal unit cell itself, yielding $\Omega = 1/2$. This is Sayre’s result\(^2\), and is eight times lower than for the single (non-crystalline) cuboid object mentioned above. Unlike the case of incoherent addition of the terms into a 3D volume and the iterative phasing of the continuous diffraction is 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 PSII, there are four asymmetric units per unit cell. Therefore, the number of unknowns is reduced by a factor of four. The number of independent measurements is reduced owing to the symmetry of the Patterson function. In this case, the symmetry of the Patterson function is $Pmmm$, which results in a factor of four reduction compared with $P\overline{1}$, and leads once more to $\Omega = 1/2$.

Non-crystallographic symmetry does help for Bragg data, as is well known for the method of molecular replacement, because it 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 $\Omega$ for the PSII Bragg data, given the support $S$ of Fig. 3d. In this case, the number of independent measurements is proportional to the volume of the unit cell, divided by four (the number of identical copies) and divided by two. Compared with the number of unknown coefficients, which is proportional to $V(S)$, we calculate
\[
\Omega = \frac{V_{\text{cell}}}{8V(S)} = 0.86
\]
where the unit-cell volume is determined from the lattice parameters provided in Extended Data Table 1. This value suggests that the phase retrieval problem for this crystal is not solvable with only the Bragg data.

**Code availability.** The code used to perform the merging of diffraction patterns into a 3D volume and the iterative phasing of the continuous diffraction is available on request.

---

31. Liang, M. et al. The coherent X-ray imaging instrument at the Linac Coherent Light Source. J. Synchrotron Rad **22**, 514–519 (2015).
32. Emma, P. et al. First lasing and operation of an ångstrom-wavelength free-electron laser. Nature Photon. **4**, 641–647 (2010).
33. DePonte, D. P. et al. Gas dynamic virtual nozzle for generation of microscopic droplet streams. J. Phys. D **41**, 195505 (2008).
34. Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. Rev. Sci. Instrum. **83**, 035108 (2012).
35. Neutens, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. Nature **406**, 752–757 (2000).
36. Barty, A. et al. Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. Nature Photon. **6**, 35–40 (2012).
37. Hart, P. et al. The CSPAD megapixel x-ray camera at LCLS. In X-Ray Free-Electron Lasers: Beam Diagnostics, Beamline Instrumentation, and Application (eds Moeller, S. P. et al.) Proc. SPIE **8504**, 85040C (SPIE, 2012).
38. Barty, A. et al. Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. J. Appl. Cryst. **47**, 1118–1131 (2014).
39. White, T. A. et al. Crystallographic data processing for free-electron laser sources. Acta Crystallogr. **D69**, 1231–1240 (2013).
40. Yefanov, O. et al. Accurate determination of segmented X-ray detector geometry. Opt. Express **23**, 28459–28470 (2015).
41. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. Science **336**, 1030–1033 (2012).
42. Cowley, J. M. Diffraction Physics 2nd edn, 150–151 (Elsevier, 1981).
43. Wilson, A. J. C. The probability distribution of X-ray intensities. Acta Crystallogr. **A34**, 517–525 (1978).
44. French, G. S. & Wilson, K. S. On the treatment of negative intensity observations. Acta Crystallogr. **A34**, 517–525 (1978).
45. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Cryst. **40**, 658–674 (2007).
46. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. **D66**, 213–221 (2010).
47. Urzhumtsev, A., Afonine, P. V., Lunin, V. Y., Terwilliger, T. C. & Adams, P. D. Metrics for comparison of crystallographic maps. Acta Crystallogr. **D70**, 2593–2606 (2014).
Extended Data Figure 1 | Model refinement is improved at low resolution. A plot of the metric $R_{\text{free}}$ as a function of resolution shell $q$, showing a marked improvement of the model refined against the 3.5-Å diffractive image. The blue curve shows $R_{\text{free}}$ prior to the inclusion of the continuous diffraction and the red curve shows $R_{\text{free}}$ afterwards. Here $R_{\text{free}}$ is calculated using only Bragg intensities (which were excluded from the refinement) for a resolution below 4.5 Å.
Extended Data Figure 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 $q = 0.23\ \text{Å}^{-1}$, or a resolution of 4.3 Å.
**Extended Data Figure 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 0.22–0.34 Å$^{-1}$) and the integrated signal in all detected Bragg peaks, for all 25,585 indexed patterns. We chose the 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). b, The measurement from the liquid jet without crystals. c, d, Two representative patterns with speckle counts above the mean solvent background, but not above the threshold (c), and one of the 2,848 strongest patterns (d), as indicated by the positions of the yellow circles in a.
Extended Data Figure 4 | Patterson function of a disordered lattice.
A distorted lattice (left, black circles, with ideal positions in grey), with vectors (red arrows) connecting all lattice points that have difference vectors \( \mathbf{a}_n - \mathbf{a}_k = (1, 1) \). On the right, the arrows from the left panel 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 resultant distribution is the autocorrelation of the displacement distribution. This process can be repeated for all difference vectors, leading to equation (3).
Extended Data Figure 5 | Continuous diffraction exhibits Wilson statistics. Histogram of merged continuous intensities in a $q$ range of 0.22–0.25 Å$^{-1}$. Above a background level of around one photon per pixel per 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 Figure 6 | Central sections of the 3D full-pattern diffraction volume. a–c, Bragg intensities in planes normal to the three orthogonal reciprocal-space axes, $q_x$, $q_y$ and $q_z$, respectively, which were arbitrarily chosen to be parallel to the $c^*$, $b^*$ and $a^*$ axes of the PSII crystal. The 4.5-Å extent of the Bragg peaks is indicated by the red circle. d–f, Full-pattern diffraction intensities in central sections normal to the same three orthogonal axes as in a–c, respectively, obtained from that 2,848 strongest snapshot diffraction patterns. g–i, Continuous diffraction intensities calculated for a single PSII dimer using the model refined from the 4.5-Å Bragg data, for the same set of orthogonal planes as in a–c, respectively. The intensities were calculated from the incoherent sum of the squared modulus of the 3D molecular transform of a single (uncrystallized) 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 PSII dimer.
**Extended Data Figure 7 | Real-space orthogonal slices.**
a–c, Slices of the observed structure factor ($F_o$) electron density map, plotted as a grey scale, of a single PSII 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 Figure 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 (that is, constant on surfaces of constant q), iterative phasing using the support constraint of Fig. 3d fails, as indicated by this plot of the FSC 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 phase retrieval transfer function for the control could not be generated. b, Plot of the cross correlation (CC) between the measured diffraction and that calculated from the determined electron density of a dimer, symmetrized by the four crystallographic orientations.
Extended Data Figure 9 | Electron density maps of regions of the PSII dimer. a–f. Electron density maps based on maximum likelihood structure factors ($2mF_o-DF_c$) obtained using the Bragg diffraction (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, b, Non-haem iron coordinated by two His residues from D1 (chain A) and two from D2 (chain D), contoured at 1.5σ (a) and 4σ (b). Neighbouring Tyr and Lys residues are displayed as well. c, Part of an α-helix (chain T), showing that the side chains (for example, Arg and Phe) fit 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 1.25σ). 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 (maps contoured at 1.25σ). Using 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, Two chlorophylls and part of the transmembrane helix of chain C (maps contoured at 1.5σ). g, Matrix of Pearson correlation coefficients of electron density maps obtained from the model presented in ref. 2 (PDB, 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_1 2_1 2_1      | P2_1 2_1 2_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† (3.40–4.5)   | 30–3.5                     |
| **Wilson B (Å²)**      | 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†               | Difference Map        | MLHL†                      |
|-------------------------|-------------------|-----------------------|----------------------------|
| **No. reflections / phases** | 55,609            | 3.5 × 10⁷             | 769,253                    |
| **Resolution range (Å)** | 30.0–4.50 (4.58–4.50) | 30.0–3.20† (3.40–4.50) | A: 30.0–3.50 (3.54–3.50)   |
| **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                | 50,074                     |
| **No. voxels**          | 1.26 × 10⁸        | 0.012                 | 1.817                      |
| **Bond lengths (Å)**    | 0.005             | 0.005                 | 0.012                      |
| **Bond angles (°)**     | 1.400             | 1.400                 | 1.817                      |

*ML, maximum likelihood.
†MLHL, maximum likelihood combined with experimental phase probability distribution.
‡3.20-Å resolution in the centre edge of the 3D voxel array.