doi:10.1038/nature09750

Resolution shell (Å)	Number of unique reflections	Multiplicity	Complete-ness (%)	⟨Peak height/ noise⟩	Merged I/σ	R_{int}(I) (%)	R_{iso}(F) (%)
30.0	416	1369	100	889	16.8	44.1	37.6
16.2	382	1278	100	29.3	15.9	8.11	15.1
13.6	362	1195	100	19.2	16.2	8.77	12.3
12.1	366	1009	100	20.3	14.7	8.45	12.9
11.1	357	678	100	19.5	12.7	10.7	14.6
10.4	355	492	100	15.4	10.6	51.7	28.7
9.85	349	283	100	14.0	9.20	28.7	22.2
9.39	352	111	100	11.8	7.12	97.9	28.1
9.01	352	28.6	99.4	9.50	4.99	60.4	24.8
8.68	88	3.92	25.1	6.70	3.29	54.4	30.8
Total	3379	717	92.8	123	11.9	35.5	22.1

Table S1 | Statistics for the merged dataset of Photosystem I at 70 fs pulse duration. R_{int}(I) quantifies the consistency of the data by comparing merged datasets from odd- and even-numbered patterns. R_{iso}(F) compares the LCLS structure factors to a conventional single-crystal dataset collected at ALS beamline 8.2.2. The bottom centre edge of the bottom detector occurs at a resolution length of 10.0 Å. See Supplementary Discussion for details.

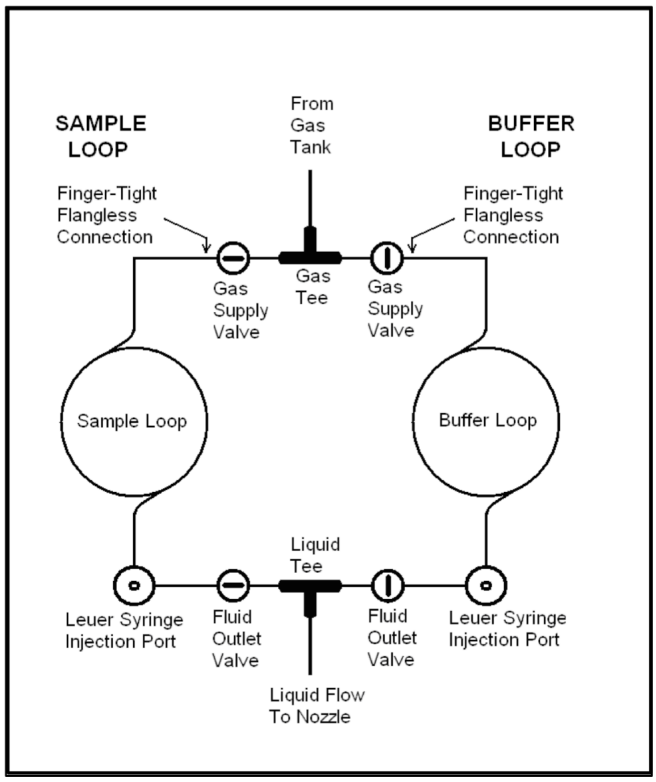

Figure S1| Schematic of the fluid lines. This scheme maintains constant flow through the nozzle when changing samples, to prevent freezing of the nozzle that occurs if the liquid stops flowing. The diluted Photosystem I nanocrystals (at 1 mg/ml) are contained in a "sample loop" of up to 10 ml volume. A "solvent loop" was integrated so that sample need not be consumed during X-ray alignment adjustments and to provide constant flow through the jet when changing the sample suspension.

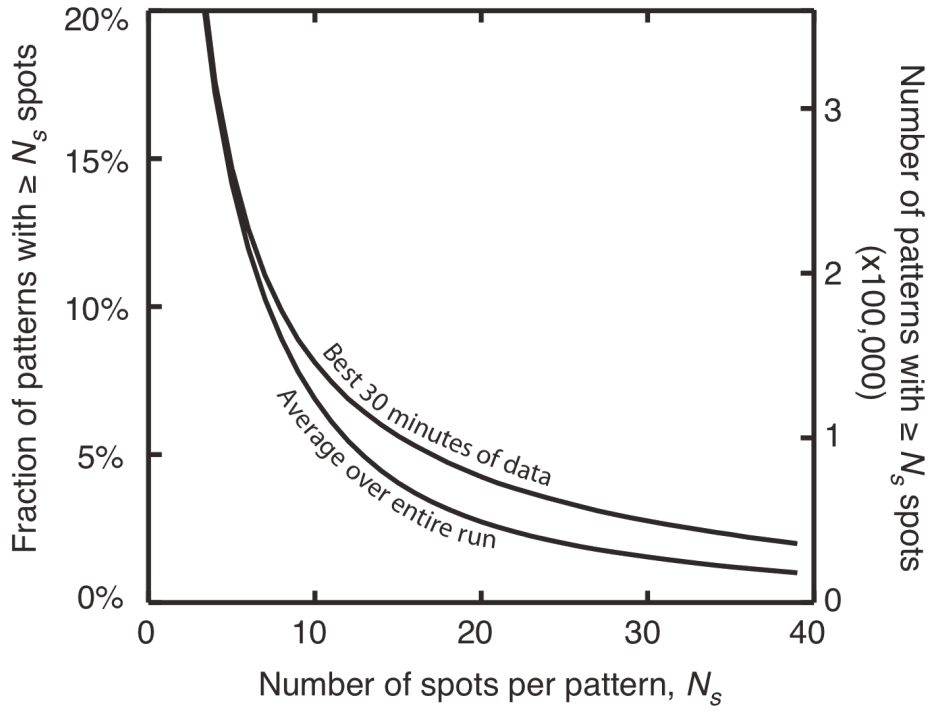

Figure S2 | Rate of hits. Complementary cumulative distribution of the number of single-pulse patterns with N_s or more measured Bragg reflections. From a run of 1.85 million pulses at 70 fs pulse duration, 6% of patterns (112,725) had 10 or more measured peaks, and 13% of those were indexed, for a total efficiency of 0.8%.

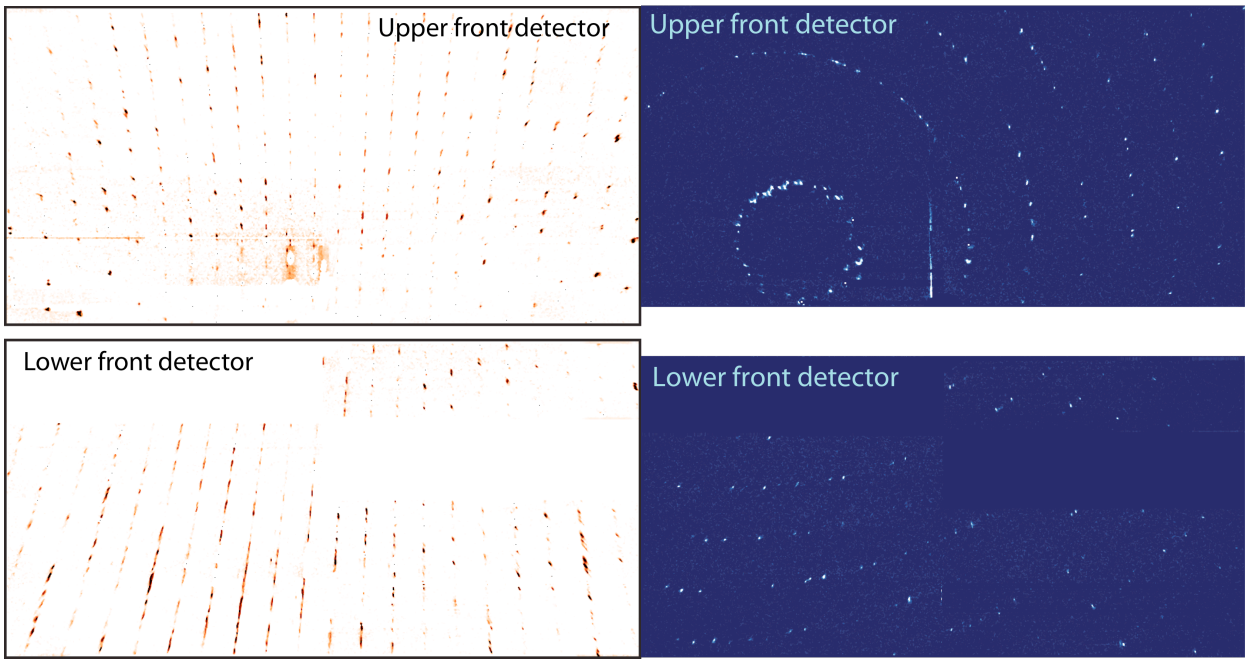

Figure S3 | Single-shot single crystal diffraction patterns of Photosystem I nanocrystals recorded with 70 fs pulses. (a, Left) A pattern from a crystal oriented with its *c* axis almost horizontal, perpendicular to the X-ray beam. (b, Right) Reciprocal lattice planes are seen as arcs, due to the intersection of the Ewald sphere with the lattice. Cuts through the shape transform, due to the finite crystal size, are evident when the intersection of the Ewald sphere with the reciprocal lattice planes is close to tangential, such as the almost continuous lines of intensity in the bottom left of (a) and the spots in the complete circle of reflections in the top left of (b).

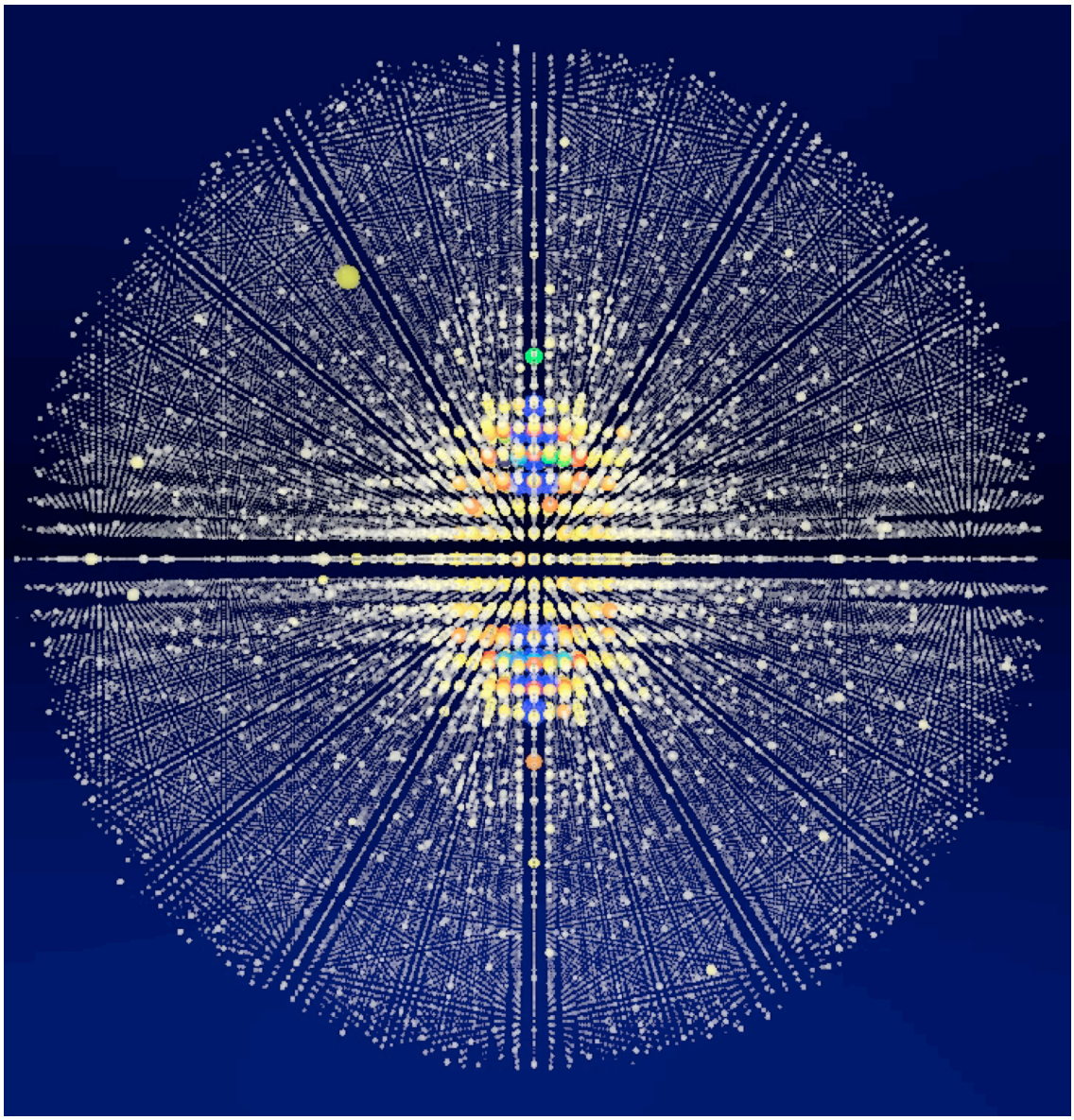

Figure S4 | 3D rendering of merged intensities. In this view the c axis is horizontal. Figure 3 (b) is the cut through the central section that is normal to the c axis.

Supplementary Methods

Diffraction Pattern Recording and Analysis

The crystals diffract very strongly so that some high intensity peaks are broadened due to saturation of the detectors, as can be seen for some peaks in Figs. 2 and 3. The detector pixels have a full well capacity of 282,000 electrons, corresponding to 571 X-ray photons of 1.8 keV (6.9 Å wavelength). When more than 571 photons arrive in a pixel on the front detectors, the generated electrons spill into neighbouring pixels. The integrated photon count in the peak is correct but the shape of the peak is lost in that region. The charge collected in the CCD wells is digitized into a 14-bit signal. The rear detectors were operated with a gain setting that produced 10.5 digital counts per X-ray photon, in order to match the well depth to the digitization range. In order to optimize the detection of weaker peaks on the front detectors, they were operated with a higher gain setting that produced 167 digital counts per X-ray photon, which saturated the analogue to digital converter (ADC) at 78 X-ray photons, and gave a readout noise less than a photon.

X-ray fluorescence from the water jet is significant and is a source of background on the diffraction patterns. For a 4-μm diameter jet we expect 210 oxygen *K*-shell fluorescence photons (525 eV) per pixel per pulse on the front detectors, whereas the smaller solid angle of the rear detectors will collect 2.9 fluorescence photons per pixel per shot. These photons are attenuated by an 8-μm thick polyimide foil placed between the water jet and the front CCDs. This foil has a transmission of 10^{-5} at 525 eV, whereas the 1.8 keV diffraction from the crystals is only attenuated by 0.5. A 4-mm diameter hole was punched in the polyimide filter for the direct FEL pulse to pass through. A shadow of the hole caused by the fluorescence photons can be faintly seen on the rear detectors, on the patterns in Fig. 2. This shadow is magnified by a factor of 14, as given by the source to foil and

source to detector distances. The fluorescence signal level in the unattenuated region is about 5 digital counts, in agreement with the abovementioned expectation.

Typical noise levels due to readout and diffuse photon background are lower than 1.3 photons/pixel. All patterns have been processed by removal of signal from known bad pixels, correction of residual common mode offsets, application of a pre-characterized gain (flat-field) correction, and subtracting background signal determined from a moving-window selection of 50 nearby no-hit frames to ensure similar jet and exposure conditions. For each pixel, the background is taken as the median of values at that location over the previous 50 frames. After background subtraction the locations of spots are determined by an intensity threshold and morphological analysis to exclude non-Bragg-peak events, producing a list of all peak locations in each image. Patterns with more than 10 Bragg peaks were selected for indexing. For the quantitative analyses of indexing (Figs. 2 and 3) and averaging intensities (Fig. 4) the saturated pixels were identified and the charge that leaked into neighbouring pixels was then integrated.

Supplementary Discussion of the quality of data collected by femtosecond protein nanocrystallography

Data Metrics

From the 1.85 million patterns collected from Photosystem I with X-ray pulses of 70 fs duration, 112,725 were identified as crystal hits with more than 10 Bragg spots. Of those, 15,445 were successfully indexed and gave rise to 2,424,394 observed reflections. Photosystem I crystals display the symmetry of space group $P6_3$. Due to the ambiguity of indexing this polar space group, the data obtained from the randomly oriented nanocrystals will appear as perfectly merohedrally twinned, displaying the apparent symmetry of spacegroup $P6_322$. The final merged dataset consists of 3,379 unique reflections after merging Friedel pairs. A 3D rendering of the final full dataset is shown in Fig. S4.

Metrics of the merged data quality are given in Table S1. The ⟨peak height/noise⟩ values are calculated from diffraction patterns as the ratios of the maximum photon counts per pixel in each predicted spot region divided by the noise estimated from the pixel variation in a ring around the spot. These height/noise values were then averaged over all spots that contributed to the merged data, in each resolution shell. In the highest-angle resolution shell, we found the ratio of the peak heights to the mean background noise level was 6.7, which compares well with the mean signal-to-noise ratio of 2 often used to define the outer resolution cut off, suggesting that high-resolution data extends beyond the edge of the detector. Multiplicity is the average number of measured spots per hkl index in each bin, totalling 2,424,394 observations divided by 3,379 indices, or 717. The merged I/σ was calculated for each hkl index using σ defined as the root-mean-square deviation of spot intensity values from their mutual mean (I), divided by the square root of the number of spots (N):

$$I_{hkl} = \left\langle I_{spot} \right\rangle \quad ; \quad \sigma_{hkl} = \frac{\sqrt{\sum_{N} \left(I_{spot} - I_{hkl} \right)^2}}{N}$$

Despite the high multiplicity, these I/σ values are lower than the ⟨peak height/noise⟩ because of the large error associated with unknown spot partialities. Estimates of these partialities may be obtained by a procedure analogous to traditional "postrefinement"[28], except that in this case the rocking curve of each nanocrystal is different, and dominated by the 3D transform of each crystal's shape. Algorithms fast enough to make postrefining this volume of data practical are still under development, but like traditional "postrefinement", such algorithms will require a rough initial estimate of the "full" spot intensities. Here, we simply sum over all partial reflections effectively integrating over the volume of the shape transform as well as integrating over the distributions of crystal shapes, orientations, and pulse fluence. Convergence of this multi-dimensional integral requires very high multiplicity[18], and this is consistent with the decrease in I/σ at around the 10Å shell, where the multiplicity is relatively low. This is mainly due to the geometry of the detector. The edge of the bottom detector intercepts a 40.5° scattering angle, and hence reflections in shells above 10.0 Å were recorded less frequently in the detector corners.

The convergence is also tested with an *R* factor comparison. Since the traditional R_{merge} statistic is not applicable here, we instead compared intensities merged from the even-numbered patterns to the odd-numbered, according to

$$R_{\mathrm{int}} = 2\frac{\sum_{\{\mathrm{hkl}\}}\left\|I_{\mathrm{even}}\right| - \left|I_{\mathrm{odd}}\right\|}{\sum_{\{\mathrm{hkl}\}}\left\|I_{\mathrm{even}}\right| + \left|I_{\mathrm{odd}}\right\|}$$

where I_{even} and I_{odd} are the merged intensities from the two interleaved datasets. Note that the factor of 2 makes the denominator an average. We find that the merging procedure converged to a relative intensity error of 36% overall, and below 10% in the middle resolution shells. Note that because R_{int} involves splitting the data into two subsets, it overestimates the error in the completely merged data set by a factor of $\sqrt{2}$.

3D Merging of Intensities

Although the Photosystem I structure is known, the original (1JB0) structure factors were too incomplete at low resolution for direct comparison with the LCLS data. We therefore compared the data with structure factors obtained from a single crystal (0.5 × 1 mm) of Photosystem I, cryogenically cooled to 100 K in a different solvent (2M sucrose in the same harvesting buffer used for LCLS data: 5mM MES pH 6.4, 0.02% β-dodecylmaltoside), and measured at the Advanced Light Source synchrotron beamline 8.2.2 with 12.4 keV photons. The metric R_{iso} shown in Table S1 quantifies the comparison between the single crystal and LCLS structure factors and is defined as

$$R_{\mathrm{iso}} = \frac{\sum_{\{\mathrm{hkl}\}} |F_{\mathrm{LCLS}} - F_{\mathrm{ALS}}|}{\sum_{\{\mathrm{hkl}\}} F_{\mathrm{LCLS}}}.$$

This metric was 22% overall and below 13% in the middle resolution shells, which is good agreement considering the differences in temperature and solvent composition. For example, a survey (not shown) of the 101 tetragonal lysozyme data sets in the PDB[35] found that the smallest R_{iso} across a temperature difference of 50 K or more was 20%. In addition, Photosystem I crystals contain 78% solvent, and the LCLS data were collected from nanocrystals in harvesting buffer, while the synchrotron data were collected from a large crystal with an additional 2M sucrose in its solvent channels. The large crystal was transferred in 10 steps from the harvesting buffer to the mixture of harvesting buffer with 2M sucrose over a time period of 2 hours prior to flash-cooling in liquid propane. Given the impact of bulk solvent on low-angle structure factors[36], it is quite remarkable that the R_{iso} between the LCLS and ALS data sets is as low as it is. Unfortunately, room-temperature data collection on single crystals of Photosystem I proved impractical due to severe radiation damage.

References

35. Berman, H. M., *et al*. The protein data bank, *Acta Cryst. D* **58**, 899-907 (2002).

36. Tronrud, D. E. TNT refinement package, *Methods Enzymol.* **277**, 306-319 (1997).