Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED

M Jason de la Cruz1, Johan Hattne1, Dan Shi1, Paul Seidler2, Jose Rodriguez2, Francis E Reyes1, Michael R Sawaya2, Duilio Cascio2, Simon C Weiss3, Sun Kyung Kim2, Cynthia S Hinck3, Andrew P Hinck3, Guillermo Calero3, David Eisenberg2 & Tamir Gonen1

Traditionally, crystallographic analysis of macromolecules has depended on large, well-ordered crystals, which often require significant effort to obtain. Even sizable crystals sometimes suffer from pathologies that render them inappropriate for high-resolution structure determination. Here we show that fragmentation of large, imperfect crystals into microcrystals or nanocrystals can provide a simple path for high-resolution structure determination by the cryoEM method MicroED and potentially by serial femtosecond crystallography.

Large and perfect crystals are desirable for traditional structure determination because they yield a strong signal over background. In reality, crystals of biological material are rarely perfect. In the mosaic model, a real, imperfect crystal is composed of several small but well-ordered blocks1. These mosaic blocks have a finite size, are misaligned with respect to each other, and may be composed of unit cells with different dimensions2. Depending on the nature and degree of disorder between the mosaic blocks, an imperfect crystal may exhibit a plethora of pathologies, which may hamper subsequent data reduction, limit the resolution of the final model, and even prevent structure determination altogether (Supplementary Figs. 1–9).

Small crystals are usually not affected by such defects and therefore may yield superior data quality where diffraction is not limited by the number of diffracting unit cells in the crystal3. For these reasons, methods such as serial femtosecond crystallography (SFX)4 at an X-ray free-electron laser (XFEL) and the electron cryomicroscopy (cryoEM) method microelectron diffraction (MicroED)5 are actively being developed. These methods can yield structures to resolutions better than 1 Å from crystals that are significantly smaller than those which are required for standard crystallography, ~10,000× smaller in volume for XFEL and ~1,000,000× smaller for MicroED. In fact, in both these techniques large crystals can prove problematic. This is because large crystals can clog up the nozzle of liquid jet-based SFX sample delivery systems6,7, while in MicroED the large electron-scattering cross-section implies that absorption extinguishes diffraction when the sample is too thick8.

Here we show that sonication, vigorous pipetting, or vortexing can be used to break large imperfect crystals into small, single-crystal fragments that are suitable for data collection and atomic structure determination by MicroED (Fig. 1). Delicate samples may benefit from gentler fragmentation by vortexing, while harsher methods such as pipetting and sonication are required to break more robust crystals.

Fragmentation was tested on eight proteins with molecular weights ranging from 0.7 to 34.6 kDa and solvent contents between 30% and 60%: lysozyme, TGF-βm–TßRII, xylanase, thaumatin, trypsin, proteinase K, thermolysin, and a segment of the tau protein (Fig. 2, left column). Six of these proteins are standard samples that readily form crystals that are large—too big for MicroED and SFX experiments using liquid injectors—and without crystal growth optimization sometimes exhibit pathologies. Two of the samples (the TGF-βm–TßRII complex and tau peptide) represent challenging, new cases. As with the standard samples, we could obtain large and imperfect crystals for TGF-βm–TßRII complex and tau without growth optimization. These large crystals were broken apart by one of the three fragmentation approaches (see Online Methods) before preparation on cryoEM grids. Micrometer- or nanometer-sized crystal fragments appeared evenly distributed on the grid; and even when grids were densely populated, single-crystal data sets could be collected by using the selected area aperture.

In many cases, the untreated crystals were mosaic, yielded diffraction patterns with multiple lattices, or were otherwise unsuitable for standard crystallographic experiments (Fig. 1 and Supplementary Figs. 1–9). In the case of the amyloid-forming peptide of tau, what appeared to be large crystals were in fact crystal bundles that produced low-resolution powder-like diffraction. These problems have traditionally been overcome by modifying the crystallization conditions to optimize crystal growth and quality—a process that can be tedious and labor intensive, particularly when crystal pathologies do not become apparent until data processing. Breaking of large crystals by physical means produced fragments that appeared crystallographically

1Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, Virginia, USA. 2Howard Hughes Medical Institute, UCLA-DOE Institute, Departments of Biological Chemistry, Chemistry & Biochemistry, and Molecular Biology Institute, UCLA, Los Angeles, California, USA. 3Department of Structural Biology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Correspondence should be addressed to T.G. (gonen@janelia.hhmi.org).

RECEIVED 2 AUGUST 2016; ACCEPTED 28 DECEMBER 2016; PUBLISHED ONLINE 13 FEBRUARY 2017; DOI:10.1038/NMETH.4178
homogeneous and yielded diffraction data at atomic resolution and void of the above artifacts (Supplementary Table 1). MicroED has already allowed rapid structure determination from several peptide fragments that could not be solved using micro-focus synchrotron sources in spite of many months of crystal optimization (Supplementary Figs. 2–4). For each sample, data

Figure 1 | Crystals before and after fragmentation and their X-ray and MicroED diffraction patterns. (a) Top row, light micrographs of imperfect crystals before fragmentation. Scale bars are 500 µm, except for tau peptide and TGF-βm–TßRII, where the scale bar is 50 µm. Bottom row, electron micrographs of fragmented crystals. Scale bars are 5 µm. (b) An example of X-ray diffraction from a thaumatin crystal exhibiting multiple lattices. (c) MicroED from a fragmented thaumatin crystal does not suffer from the same pathologies. (d) An example of X-ray diffraction from trypsin crystals yielding only powder-like diffraction. (e) MicroED analysis of fragmented trypsin crystals does not suffer from the same pathologies. The inset shows a close up of the spot indicated by the blue circle. X-ray diffraction patterns were collected on a CuKα home over a 0.37° rotation range; MicroED patterns were recorded as detailed in the Online Methods. Further examples are shown in Supplementary Figures 1–9.
were collected from two to ten crystal fragments, and data sets were merged for completeness and multiplicity (Supplementary Table 1). Although we have previously published two structures where a single nanocrystal was sufficient for structure determination, multicrystal merging is generally preferred. This partly because the grid on which the crystals are mounted limits the accessible rotation range during data collection and partly because small, weakly scattering samples require a more intense beam to yield statistically accurate measurements of the Bragg reflections and may be irreversibly damaged before a complete data set can be obtained. Recording narrow wedges with fewer exposures for each crystal involves a trade-off between limiting the damage to the sample and maintaining sufficient beam intensity for the signal from high-resolution reflections to be accurately integrated. Because the resolutions obtained by MicroED are comparable to or better than those obtained by X-ray diffraction before sonication, it appears that fragmentation only broke apart the large crystals into small crystal domains and did not otherwise damage the lattice order.

The macromolecular structures in this study were phased using molecular replacement (MR) or by direct methods and refined using electron-scattering factors. In all cases, the refined model fits the calculated density well (Fig. 2, middle column) with an overall real-space correlation coefficient ranging from 0.72 to 0.91 (RSCC; Supplementary Table 1). The resulting simulated-annealing (SA) composite omit maps (Fig. 2, right column) match their respective models very well, indicating that the data are indeed high quality and unbiased. Further, the SA omit maps for lysozyme reveal depressions or holes in the aromatic rings of amino acid side chains, while for proteinase K toroidal densities were observed even for proline residues at the 1.6-Å resolution cutoff. A well-coordinated calcium ion is clearly visible in the omit map for trypsin, as is one of the iodides in the xylanase structure. Individual atoms are visible in the density for the tau peptide and calcium ions are visible in the OMIT maps for xylanase and trypsin, lysozyme, and for proline residues in the proteinase K structure. Iodide is shown above the mean, eight residues in loop regions of the tau peptide model indicate a –sheet generated by crystal fragments. Left column, ribbon representations of the corresponding macromolecules. The tau peptide model indicates a β-sheet generated by crystal packing. Middle column, 2mF –DF peaks for hydrogen atoms; for one residue, modeled hydrogens allowed us to unambiguously place the correct sidechain rotamer. All structures were determined between 1.1–2.9-Å resolution. We note that the R-factors from model refinement are generally higher in MicroED than are typical values obtained in X-ray crystallography at similar resolutions. Others and we have made such observations in previously published studies. We believe that the main reason for this is the lack of adequate electron-scattering factors in the crystallographic refinement software. X-rays are scattered by the electron cloud of an atom, while electrons are scattered by the atomic Coulomb potential, which arises from the nucleus as well as its surrounding electrons. Current electron-scattering-factor tables do not properly account for this difference, and this may contribute to the residual between observed and calculated structure factor amplitudes leading to higher than normal crystallographic and free R-factors. We note that the gap between R_work and R_free is small, indicating no concerns of overfitting (Supplementary Fig. 10).

MicroED is inherently well suited to studying crystal fragments. Electrons interact strongly with matter so that large crystals are not required to detect high-resolution Bragg reflections. MicroED has been used to determine the structures of the enzymes lysozyme, proteinase K, catalase, and Ca^2+-ATPase. It was also recently used to solve the 1.4-Å-resolution structure of TGF-βRII one of the disulfide bonds is shown. Right column, SA composite omit maps contoured at 1 σ above the mean, except for the maps from lysozyme and proteinase K, which are contoured at 1.5 σ above the mean. Depressions or holes can be observed in the density of side chains of aromatic residues for tau peptide and lysozyme, and for proline residues in the proteinase K structure. Iodide and calcium ions are visible in the omit maps for xylanase and trypsin, respectively. All figures were generated using PyMOL.
the toxic nonamyloid-β component (NAC) core of α-synuclein, where diffraction data were collected from crystals smaller than the wavelength of visible light. We previously found that crystals thinner than ~400 nm are suitable for MicroED and routinely yield atomic-resolution information. The highest resolution structures reported so far by using MicroED are from four prion protein fragments determined at 1-Å resolution solved by direct phasing methods. The resolution of macromolecules whose crystals we investigated, we believe this approach will be broadly applicable whenever large crystals are available. Fragmentation widens the scope of diffraction methods such as MicroED and SFX to include samples that do not exclusively form tiny crystals but instead form large, imperfect crystals.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

Acknowledgments

The Goeken laboratory is supported by funds from the Howard Hughes Medical Institute. The Eisenberg laboratory is supported by HHMI, NIH grant 1R01-AG029430 (D.E.), and DOE grant DE-FC02-02ER63421 (D.E.). The Calero laboratory is supported by NIH grants RO1GM120292 and RO1DK102495 (G.C.) and BioXFEL-STC123106 (G.C.). The Hinck laboratory is supported by NIH grants R01-GM58670 and R01-CA172886 (A.P.H.).

Author Contributions

D.S. and T.G. designed the experiment; M.J.d.L.C. and F.E.R. prepared lysozyme, xylanase, thaumatin, trypsin, proteinase K, and thermolysin samples; J.R. prepared tau peptide samples; S.C.W., S.K.K., C.S.H., A.P.H., and G.C. prepared TGF-β1–TβRII samples; M.J.d.L.C., D.S., J.R., and S.C.W. collected data; M.J.d.L.C., J.H., P.S., M.R.S., D.C., S.C.W., and G.C. analyzed data and refined and determined models; J.H., J.R., and M.R.S. prepared figures; M.J.d.L.C., J.H., and T.G. wrote the manuscript with contributions from all authors.

Competing financial interests

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Darwin, C.G. Philosophical Magazine Series 6 43, 800–829 (1922).
2. Nave, C. Acta Crystallogr. D Biol. Crystallogr. 54, 848–853 (1998).
3. Cusack, S., et al. Nat. Struct. Biol. 5 (Suppl.), 634–637 (1998).
4. Schlichting, I. IUCrJ 2, 246–255 (2015).
5. Shi, D., Nannenga, B.L., Iadanza, M.G. & Gonen, T. eLife 2, e01345 (2013).
6. Weierstall, U. Phil. Trans. R. Soc. Lond. B 369, 20130337 (2014).
7. Sierra, R.G. et al. Acta Crystallogr. D Biol. Crystallogr. 68, 1584–1587 (2012).
8. Henderson, R. Q. Rev. Biophys. 28, 171–193 (1995).
9. Nannenga, B.L., Shi, D., Leslie, A.G.W. & Gonen, T. Nat. Methods 11, 927–930 (2014).
10. Nannenga, B.L., Shi, D., Hatte, J., Reyes, F.E. & Gonen, T. eLife 3, e03600 (2014).
11. Collieux, C. et al. in International tables for crystallography, Vol. C (eds. Prince, E. & Wellberry, T.R.) Ch. 4.3 (International Union of Crystallography, 2006).
12. Yonekura, K., Kato, K., Ogasawara, M., Tomita, M. & Toyoshima, C. Proc. Natl. Acad. Sci. USA 112, 3368–3373 (2015).
13. Hatte, J., Shi, D., de la Cruz, M.J., Reyes, F.E. & Gonen, T. J. Appl. Crystallogr. 49, 1029–1034 (2016).
14. Rodriguez, J.A. et al. Nature 525, 486–490 (2015).
15. Sawaya, M.R. et al. Proc. Natl. Acad. Sci. USA 113, 11232–11236 (2016).
16. Von Laue, M. in Nobel Lectures Physics 1901–1921, 347–355 (Elsevier, 1967).
17. Stevenson, H.P. et al. Acta. Crystallogr. D Struct. Biol. 72, 603–615 (2016).
18. Bublitz, M. et al. IUCrJ 2, 409–420 (2015).
19. The PyMOL Molecular Graphics System v. 1.8 (Schrödinger, LLC, 2015).
ONLINE METHODS

Protein crystallization. Large crystals (>500 µm along the longest edge) were grown via hanging-drop vapor diffusion at room temperature using previously established protocols. All enzymes were purchased from Sigma-Aldrich (St. Louis, Missouri) unless otherwise noted.

Tau peptide (VQIVYK) was dissolved in distilled water, and crystals were prepared by mixing tau peptide with arachidonic acid and mecloxylyn sulfosalicylate. Crystals formed with a 1:2 drop peptide to precipitant ratio from 65–70% ethylene glycol in Tris pH 8.5.

Lysozyme (G. gallus) crystals were prepared by equilibrating 50 mg/ml lysozyme in 50 mM sodium acetate over 1.7 M sodium chloride, 50 mM sodium acetate pH 4.7.

TGF-β3–TßRII was expressed and purified as described before (ref. 20 and unpublished data). Crystals were grown from 0.5 µl of protein (20 mg/ml), 0.25 µl mother liquor, and 0.2 µl seed stock in 100 mM HEPES–NaOH pH 7.5 and 45% MPD.

Xylanase from T. reesei (Hampton Research, Aliso Viejo, California) was dialyzed against 10 mM bicine pH 9.0, 1 mM magnesium sulfate, and 1 mM DTT; and it was combined with a precipitant solution containing 0.3 M sodium iodide, 1.2–1.3 M ammonium sulfate, and 100 mM bicine pH 9.0 (ref. 21).

Thaumatin (T. danielli) crystals were grown from 2 µl of protein (25 mg/ml in water) and Hampton Research Index Reagent 26 (1.1 M ammonium tartarate pH 7.0).

Trypsin (B. taurus) was dissolved (60 mg/ml) in 10 mg/ml benzamidine, 3 mM calcium chloride and equilibrated against 4% (w/v) PEG4000, 0.2 M lithium sulfate, 0.1 M MES pH 6.5, and 15% ethylene glycol.

Proteinase K (E. album) crystals were grown by combining 2 µl of protein solution (50 mg/ml) with 2 µl of precipitant solution (1.0–1.3 M ammonium sulfate, 0.1 M Tris pH 8.0)13.

Thermolysin from B. thermoproteolyticus (Hampton Research, Aliso Viejo, California) crystals were prepared by equilibrating a 160 mg/ml solution of thermolysin (45% dimethyl sulfoxide, 50 mM Tris pH 7.5, and 2.5 M cesium chloride) over 0.5 ml of water22.

Crystal fragmentation. Drops containing the crystals were placed in separate microfuge tubes and suspended in crystal mother liquor. A sonicating water bath with electronic control (Elmasonic P30H, Singen, Germany) was set at its lowest power (30% at 37 kHz) for gentle agitation of the crystals in the tube. With the tube sealed, its tip was briefly submerged in the activated water bath for 0.5 s. Alternatively, crystals were fragmented by vigorously pipetting a crystal suspension in mother liquor (trypsin and thaumatin) or vortexed with 0.5 mm disruption glass beads in a 1.5 ml reaction tube for 2 s (TGF-β3–TßRII). A detailed protocol for crystal fragmentation is available as a Supplementary Protocol and is published online23.

MicroED data collection. The solution containing fragmented crystals was then applied to transmission electron microscope (TEM) grids with carbon film support, and it was plunged frozen in liquid ethane. Frozen grids were mounted in a Gatan Model 626 cryospecimen holder and examined using an FEI Tecnai F20 field-emission TEM operated at an accelerating voltage of 200 kV, which corresponds to a de Broglie wavelength of 0.025 Å, and the grids were screened for crystals in over-focused diffraction mode. Where a single still shot revealed strong diffraction, data were collected as continuous rotation-tilt series8. Individual frames were recorded on a TVIPS TemCam-F416 as 4 s exposures while the stage was rotating at 0.09°/s, except for tau peptide, which was rotated at 0.29°/s during 2 s exposures. For the macromolecular samples, the absolute tilt angle was generally <35° (<65° for tau peptide). Data sets collected from each crystal spanned between 30° and 135°, corresponding to a total dose no greater than 1.2–5.5 e-/Å² at the given rotation rate. The selected area aperture of the TEM was used to limit the area from which data were collected, making it possible to select a single crystal fragment. The virtual detector distance varied between 0.73 mm and 3.6 m for the different data sets, corresponding to a maximum resolution between 0.95 Å and 2.0 Å in the detector corners at an acceleration voltage of 200 kV. The movie frames recorded in rolling-shutter mode on the TVIPS TemCam-F416 CMOS camera were converted to SMV format, while preserving as much as possible of the metadata necessary for subsequent processing24. Detailed protocols for data collection were recently published25.

Structure determination. Sweeps were corrected to account for negative pixel values13, then they were indexed and integrated in MOSFLM26 using its graphical interface iMOSFLM27 or in XDS28. To further probe the scattering power of the crystallites, diffraction patterns were initially processed without imposing any resolution cutoffs other than those entailed by the geometry of the experiment. Except for the tau peptide and TGF-β3–TßRII, the high-resolution limits due to factors other than the area of the detector were instead determined case by case to give stable processing during refinement of the atomic model. Multicrystal scaling and merging were performed in AIMELESS29. Structures were phased by molecular replacement using MOLREP30 from the PDB entries 3j6k, 1ktz, 2dfb, 4ek0, 2ptn, 5frk, and 2ttl for lysozyme, TGF-β3–TßRII, xylanase, thaumatin, trypsin, proteinase K, and thermolysin, respectively, or by direct methods using SHELEX31. A free R set comprising approximately 5% of the unique reflections was copied from the deposited data for each search model. The deposited data for the xylanase and trypsin models do not define a free set; instead, a new set was chosen using freerflag32.

Maximum likelihood structure refinement was carried out in phenix.refine33 using electron-scattering factors11. Explicit water molecules were automatically modeled by phenix.refine and subsequently manually curated in Coot34. SA composite omit maps were computed from the refined models using CNS35 with electron-scattering factors36. For these calculations, charged species were manually removed from the phasing models because they are not included in CNS’s electron-scattering library. The SA protocol was defined to exclude ~5% of the structure at a starting temperature of 2,500 K. Protocols for data analysis in MicroED were recently published25.

Statistics. The number of values used to calculate the statistics in Supplementary Table 1 are given as “# total reflections” and “# unique reflections.” No other statistical tests are used in the manuscript.

Data availability statement. Atomic coordinates and structure factors were deposited to the Protein Data Bank (PDB; 5k7n, 626 cryospecimen holder and examined using an FEI Tecnai F20 field-emission TEM operated at an accelerating voltage of 200 kV, which corresponds to a de Broglie wavelength of 0.025 Å, and the
5k7o, 5ty4, 5k7p, 5k7q, 5k7r, 5k7s, and 5k7t) and the Electron Microscopy Data Bank (EMDB; 8216, 8217, 8472, 8218, 8219, 8220, 8221, and 8222), and the raw data were uploaded to the Structural Biology Data Grid\textsuperscript{37} (10.15785/SBGRID/284, 10.15785/SBGRID/285, 10.15785/SBGRID/368, 10.15785/SBGRID/286, 10.15785/SBGRID/287, 10.15785/SBGRID/288, 10.15785/SBGRID/289, and 10.15785/SBGRID/290).

20. Zúñiga, J.E. \textit{et al.} \textit{J. Mol. Biol.} \textbf{354}, 1052–1068 (2005).
21. Watanabe, N., Akiba, T., Kanai, R. & Harata, K. \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{62}, 784–792 (2006).
22. English, A.C., Done, S.H., Caves, L.S.D., Groom, C.R. & Hubbard, R.E. \textit{Proteins} \textbf{37}, 628–640 (1999).
23. de la Cruz, M.J. \textit{et al.} Micro- and nanocrystal preparation for MicroED and XFEL serial crystallography by fragmentation of imperfect crystals. \textit{Protocol Exchange} \texttt{http://dx.doi.org/10.1038/protex.2017.010} (2017).
24. Hattne, J. \textit{et al.} \textit{Acta Crystallogr. A Found. Adv.} \textbf{71}, 353–360 (2015).
25. Shi, D. \textit{et al.} \textit{Nat. Protoc.} \textbf{11}, 895–904 (2016).
26. Leslie, A.G.W. & Powell, H.R. (eds.) \textit{Evolving Methods for Macromolecular Crystallography} (Springer, 2007).
27. Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R. & Leslie, A.G.W. \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{67}, 271–281 (2011).
28. Kabsch, W. \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{66}, 125–132 (2010).
29. Evans, P.R. & Murshudov, G.N. \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{69}, 1204–1214 (2013).
30. Vagin, A. & Teplyakov, A. \textit{J. Appl. Crystallogr.} \textbf{30}, 1022–1025 (1997).
31. Sheldrick, G.M. \textit{Acta Crystallogr. A Found. Adv.} \textbf{71}, 3–8 (2015).
32. Winn, M.D. \textit{et al.} \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{67}, 235–242 (2011).
33. Afonine, P.V. \textit{et al.} \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{68}, 352–367 (2012).
34. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. \textit{Acta Crystallogr. D Biol. Crystallogr.} \textbf{66}, 486–501 (2010).
35. Brunger, A.T. \textit{Nat. Protoc.} \textbf{2}, 2728–2733 (2007).
36. Gonen, T. \textit{et al.} \textit{Nature} \textbf{438}, 633–638 (2005).
37. Meyer, P.A. \textit{et al.} \textit{Nat. Commun.} \textbf{7}, 10882 (2016).