High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography

Sébastien Boutet,1* Lukas Lomb,2,3 Garth J. Williams,1 Thomas R. M. Barends,2,3 Andrew Aquilà,4 R. Bruce Doak,5 Uwe Weierstall,6 Daniel P. DePonte,4 Jan Steinbrener,2,3 Robert L. Shoeman,2,3 Marc Messerschmidt,1 Anton Barty,7 Thomas A. White,4 Stephan Kassemeyer,2,8 Richard A. Kirian,5 M. Marvin Seibert,1 Paul A. Montanez,1 Chris Kenney,6 Ryan Herbst,6 Philip Hart,6 Jack Pines,6 Gunther Haller,7 Sol M. Gruner,7,8 Hugh T. Philipp,7 Mark W. Tate,7 Marianne Hromalik,9 Lucas J. Koerner,10 Niels van Bakel,11 John Morse,12 Wilfred Ghonsalves,1 David Arlund,1 Michael J. Bogan,14 Carl Coleman,4 Raimund Fromme,15 Christina Y. Hampton,14 Mark S. Hunter,15 Linda Johansson,13 Gergely Katona,13 Christopher Kupitz,15 Mengning Liang,4 Andrew V. Martin,9 Karol Nass,16 Lars Redecke,17 Francesco Stellato,4 Nicusor Timneanu,18 Dingjie Wang,5 Nadia A. Zatsepin,5 Donald Schafer,1 James Defever,2 Richard Neutze,9 Petra Fromme,16 John C. H. Spence,5 Henry N. Chapman,4,16 Ilme Schlichting2,3

1Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. 2Max-Planck-Institut für Medizinische Forschung, Jahnstrasse 29, 69120 Heidelberg, Germany. 3Max Planck Advanced Study Group, Center for Free-Electron Laser Science, Notkestrasse 85, 22607 Hamburg, Germany. 4Center for Free-Electron Laser Science, DESY, Notkestrasse 85, 22607 Hamburg, Germany. 5Department of Physics, Arizona State University, Tempe, AZ 85287, USA. 6Particle Physics and Astrophysics, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. 7Department of Physics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA. 8Wilson Laboratory, Cornell University, CHESS, Ithaca, NY 14853, USA. 9Electrical and Computer Engineering, SUNY Oswego, Oswego, NY 13126, USA. 10The Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA. 11Nikhef, National Institute for Subatomic Physics, Science Park 105, 1098 XG, Amsterdam, Netherlands. 12European Synchrotron Radiation Facility, 38043 Grenoble Cedex, France. 13Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden. 14PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA. 15Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85287–1604, USA. 16University of Hamburg, Luruper Chaussee 149, 22761 Hamburg, Germany. 17Joint Laboratory for Structural Biology of Infection and Inflammation, Institute of Biochemistry and Molecular Biology, University of Hamburg, and Institute of Biochemistry, University of Lübeck, at DESY, Hamburg, Germany. 18Laboratory of Molecular Biophysics, Department of Cell and Molecular Biology, Uppsala University, Husargatan 3 (Box 596), SE-751 24 Uppsala, Sweden. *To whom correspondence should be addressed. E-mail: sboutet@slac.stanford.edu

Structure determination of proteins and other macromolecules has historically required the growth of high-quality crystals sufficient large to diffract x-rays efficiently while withstanding radiation damage. We applied serial femtosecond crystallography (SFX) using an x-ray free-electron laser (XFEL) to obtain high resolution structural information from microcrystals (<1×1×3 μm³) of the well-characterized model protein lysozyme. The agreement with synchrotron data demonstrates the immediate relevance of SFX for analyzing the structure of the large group of difficult-to-crystallize molecules.

Elucidating macromolecular structures by x-ray crystallography is an important step in the quest to understand the chemical mechanisms underlying biological function. Although facilitated greatly by synchrotron x-ray sources, the method is limited by crystal quality and radiation damage (1). Crystal size and radiation damage are inherently linked, as reducing radiation damage requires lowering the incident flux. This in turn calls for large crystals that yield sufficient diffraction intensities while reducing the dose to individual molecules in the crystal. Unfortunately, growing well-ordered large crystals can be difficult in many cases, particularly for large macromolecular assemblies and membrane proteins. In contrast, micron-sized crystals are frequently observed. Although diffraction data of small crystals can be collected using micro-focus synchrotron beamlines, this remains a challenging approach due to the rapid damage suffered by these small crystals (I).

Serial femtosecond crystallography (SFX) using x-ray free-electron laser (XFEL) radiation is an emerging method for 3D structure determination using crystals ranging from a few micrometers to a few hundred nanometers in size and potentially even smaller. This method relies upon x-ray pulses that are both sufficiently intense to produce high quality diffraction while of short enough duration to terminate before the onset of significant radiation damage (2–4). X-ray pulses of only 70 femtoseconds duration terminate before any chemical damage processes have time to occur, leaving primarily ionization and X-ray induced thermal motion as the main sources of radiation damage (2–4). SFX therefore promises to break the correlation between sample size, damage and resolution in structural biology. In SFX, a liquid microjet is used to introduce fully hydrated randomly oriented crystals into the single-pulse XFEL beam (5–8), as illustrated in Fig. 1. A recent low-resolution proof-of-principle demonstration of SFX performed at the Linac Coherent Light Source (LCLS) (9) using crystals of photosystem I ranging in size from 200 nm to 2 μm produced interpretable electron density maps (6). Other demonstration experiments using crystals grown in vivo (7) as well as in the lipidic sponge phase for membrane proteins (8) were recently published. However, in all these cases, the x-ray energy of 1.8 keV (6.9 Å) limited the resolution of the collected data to approximately 8 Å. Data collection to a resolution better than 2 Å became possible with the recent commissioning of the LCLS Coherent X-ray Imaging (CXI) instrument (10). The CXI instrument provides hard x-rays pulses suitable for high-resolution crystallography and is equipped with Cornell-SLAC Pixel Array Detectors (CSPADs) consisting of 64 tiles of 192 × 185 pixels each, arranged as shown in Fig. 1 and figs. S1 and S2. The CSPAD supports the 120 Hz readout rate required to measure each x-ray pulse from LCLS (11).

Here we describe SFX experiments performed at CXI analyzing the structure of hen egg white lysozyme (HEWL) as a model system using microcrystals of approximately 1×1×3 μm³ (4, 11). HEWL is an extremely well-characterized protein that crystallizes easily. It was the first enzyme to have its structure determined by x-ray diffraction (12), and has since been thoroughly characterized to very high resolution (13). Lysozyme has served as a model system for many investigations, including...
Table 1. SFX and synchrotron data and refinement statistics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>40 fs* pulses</th>
<th>5 fs* pulses</th>
<th>SLS RT data 3 ****</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>1.32 Å</td>
<td>1.32 Å</td>
<td>0.9997 Å</td>
</tr>
<tr>
<td>X-ray focus [μm²]</td>
<td>~10</td>
<td>~10</td>
<td>100 x 100</td>
</tr>
<tr>
<td>Pulse energy/fluence at sample</td>
<td>600 μJ/4x10⁷ph/pulse</td>
<td>53 μJ/3.5x10⁵ph/pulse</td>
<td>N.A./2.5 x10⁷ph/s</td>
</tr>
<tr>
<td>Dose [MGy]</td>
<td>33.0 per crystal</td>
<td>2.9 per crystal</td>
<td>0.024 total</td>
</tr>
<tr>
<td>Dose rate [Gy/s]</td>
<td>8.3 x 10⁶</td>
<td>5.8 x 10⁶</td>
<td>9.6 x 10⁶</td>
</tr>
<tr>
<td>Space group</td>
<td>P4₁2₁2</td>
<td>P4₁2₁2</td>
<td>P4₁2₁2</td>
</tr>
<tr>
<td>Unit cell length [Å], α=β=γ=90°</td>
<td>a=b=79, c=38</td>
<td>a=b=79, c=38</td>
<td>a=b=79, c=38</td>
</tr>
<tr>
<td>Oscillation range/exposure time</td>
<td>Still exp. / 40 fs*</td>
<td>Still exp. / 5 fs*</td>
<td>1.0°, 0.25 s</td>
</tr>
<tr>
<td># collected diffraction images</td>
<td>1471615</td>
<td>1997712</td>
<td>100</td>
</tr>
<tr>
<td># of hits/indexed images</td>
<td>66442/12247</td>
<td>40115/1057</td>
<td>n.a./100</td>
</tr>
<tr>
<td>Number of reflections</td>
<td>n.a.</td>
<td>n.a.</td>
<td>70960</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>9921</td>
<td>9743</td>
<td>9297</td>
</tr>
<tr>
<td>Resolution limits [Å]</td>
<td>35.3-1.9</td>
<td>35.3-1.9</td>
<td>35.4-1.9</td>
</tr>
<tr>
<td>Completeness**</td>
<td>98.3% (96.6%)</td>
<td>98.2% (91.2%)</td>
<td>92.6% (95.1%)</td>
</tr>
<tr>
<td>I/σ(I)**</td>
<td>7.4 (2.8)</td>
<td>7.3 (3.1)</td>
<td>18.24 (5.3)</td>
</tr>
<tr>
<td>Rmerge***</td>
<td>0.158</td>
<td>0.159</td>
<td>n.a.</td>
</tr>
<tr>
<td>Wilson R-factor******</td>
<td>28.3 Å²</td>
<td>28.5 Å²</td>
<td>19.4 Å²</td>
</tr>
<tr>
<td>R-factor/R-free******</td>
<td>0.1960.0229</td>
<td>0.1890.227</td>
<td>0.1660.200</td>
</tr>
<tr>
<td>Rmsd bonds, Rmsd angles******</td>
<td>0.006 Å, 1.00°</td>
<td>0.006 Å, 1.03°</td>
<td>0.007 Å, 1.05°</td>
</tr>
<tr>
<td>PDB code</td>
<td>4ET8</td>
<td>4ET9</td>
<td>4ETC</td>
</tr>
</tbody>
</table>

* Electron bunch length ** Highest resolution shell: 2.0-1.9 Å

*** Rmerge as defined in (19) \( R_m = \left( \sum |F_{obs}^2 - |F_{calc}^2| / \sum |F_{obs}^2 + |F_{calc}^2| \right) \)

**** Statistics from XDS (19) ***** Calculated with TRUNCATE (20)

****** Calculated with PHENIX (27)

Even though the underlying radiation damage processes differ due to the different time scales of the experiments using an XFEL and a synchrotron/rotating anode (femtoseconds vs. seconds/hours), no features related to radiation damage are observed in difference maps calculated between the SFX and the low-dose synchrotron data (Fig. 2B). In addition to local structural changes, metrics like I/I₀ and the Wilson B-factor are most often used to characterize global radiation damage in protein crystallography (16). I/I₀ is not applicable to the SFX data. However, the Wilson-B factors of both SFX data sets show values typical for room temperature data sets and do not differ significantly from those obtained from synchrotron and rotating anode data sets collected with different doses, using similarly grown larger crystals kept at room temperature and fully immersed in solution (17) (Table 1 and table S1). The R-factors calculated between all collected data sets do not show a dose dependent increase (fig. S4). However, higher R-factors are observed for the SFX data, indicating a systematic difference. This is not caused by non-convergence of the Monte Carlo integration since scaling the 40 fs and 5 fs data together does not affect the scaling behavior. Besides non-isomorphism or radiation damage, possible explanations for this difference could include suboptimal treatment of weak reflections, the difficulties associated with processing still diffraction images and other SFX-specific steps in the method. SFX is an emerging technique, and data processing algorithms, detectors and data collection methods are under continuous development.

A simple consideration shows the attainable velocities of atoms in the sample depend on the deposited X-ray energy versus the inertia of those atoms: \( v = \sqrt{3k_BT/m} \), where m is the mass of a carbon atom, for example, T is temperature and k_B is Boltzmann’s constant. For an impulse absorption of energy at the doses of our LCLS measurements we predict average velocities less than 10 Å / ps, which gives negligible displacement during the FEL pulses. On the timescale of femtoseconds, radiation damage is primarily caused by impulsive rearrangement of atoms and electron density, rather than the relatively slow processes of chemical bond breaking typical in conventional crystallography using much longer exposures at much lower dose rates (the dose rate in this experiment was approximately 0.75 MGy per femtosecond).

Neither the SFX electron density maps nor the Wilson B-factors suggest obvious signs of significant radiation damage. Very short pulses (5 fs electron bunch) are not expected to produce observable damage, according to simulations (3). Furthermore, it has been reported that the...
actual x-ray pulses are shorter than the electron bunches for XFELs, making the pulse duration possibly shorter than the relevant Auger decays (17). The agreement between the SXF results using 40 fs pulses and 5 fs pulses suggests similar damage characteristics for the two pulse durations based on the available data. Our results demonstrate that under the exposure conditions used, SXF yields high quality data suitable for structural determination. SXF reduces the requirements on crystal size and therefore the method is of immediate relevance for the large group of difficult-to-crystallize molecules, establishing SXF as a very valuable high-resolution complement to existing macromolecular crystallography techniques.

References and Notes
11. Materials and methods are available as supporting material on Science Online.
23. The cheetah software package is available via the url: http://www.desy.de/~barty/cheetah/

Acknowledgments: Portions of this research were carried out at the Linac Coherent Light Source, a National User Facility operated by Stanford University on behalf of the U.S. Department of Energy (DOE), Office of Basic Energy Sciences (OBER) and at the Swiss Light Source, beamline X10SA, Paul Scherrer Institute, Villigen, Switzerland. The CXI instrument was funded by the LCLS Ultrafast Science Instruments (LUSI) project funded by the U.S. Department of Energy, Office of Basic Energy Sciences. We acknowledge support from the Max Planck Society, the Hamburg Ministry of Science and Research and the Joachim Herz Stiftung, as part of the Hamburg Initiative for Excellence in Research (LEXI); the Hamburg School for Structure and Dynamics in Infection; the U.S. NSF (awards 0417142 and MCB-1021557); the NIH (award 1R01GM095583); the German Federal Ministry for Education and Research (grants 1OKX0806 and 1OKX0807); AMOS program within the Chemical Sciences, Geosciences, and Biosciences Division of the OBER, Office of Science, U.S. DOE; the Swedish Research Council; the Swedish Foundation for International Cooperation in Research and Higher Education. We thank A. Meinhart and E. Hofmann for collecting the synchrotron data set, M. Gebhart for help preparing the crystals, and M. Hayes and the technical staff of SLAC and the LCLS for their great support in carrying out these experiments. Special thanks to G. M. Stewart, T. Anderson, and SLAC Infomedia for generating Fig. 1. The structure factors and coordinates have been deposited with the Protein Data Bank (accession codes 4ET8, 4ET9, 4ETA, 4ETB, 4ETC, 4ETD, 4ETE). The Arizona Board of Regents, acting for and on behalf of Arizona State University and in conjunction with R.B.D., U.W., D.P.D., and J.C.H.S., has filed U.S. and international patent applications on the nozzle technology applied herein.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1217737/DC1
Materials and Methods
Figs. S1 to S7
Table S1
References (22–20)
12 December 2011; accepted 21 May 2012
Published online 31 May 2012
10.1126/science.1217737
Fig. 1. Experimental geometry for serial femtosecond crystallography at the Coherent X-ray Imaging instrument. Single pulse diffraction patterns from single crystals flowing in a liquid jet are recorded on a CSPAD at the 120 Hz repetition rate of LCLS. Each pulse was focused at the interaction point using 9.4 keV x-rays. The sample-to-detector distance (z) was 93 mm.

Fig. 2. (A) Final, refined 2m\(F_{\text{obs}}\)-DF\(_{\text{calc}}\) (1.5\(\sigma\)) electron density map (18) of lysozyme at 1.9 Å resolution calculated from 40 fs pulse data. (B) \(F_{\text{obs}}\)[40 fs]-\(F_{\text{obs}}\)[synchrotron] difference Fourier map, contoured at +3\(\sigma\) (green) and -3\(\sigma\) (red). No interpretable features are apparent. The synchrotron dataset was collected with a radiation dose of 24 kGy.