Structure of β-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy

Alberto Bartesaghi, Doreen Matthies, Soojay Banerjee, Alan Merk, and Sriram Subramaniam

Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Edited by Wah Chiu, Baylor College of Medicine, Houston, TX, and approved June 20, 2014 (received for review February 14, 2014)

We report the solution structure of Escherichia coli β-galactosidase (∼465 kDa), solved at ∼3.2-Å resolution by using single-particle cryo-electron microscopy (cryo-EM). Densities for most side chains, including those of residues in the active site, and a catalytic Mg\(^{2+}\) ion can be discerned in the map obtained by cryo-EM. The atomic model derived from our cryo-EM analysis closely matches the 1.7-Å crystal structure with a global rmsd of ∼0.66 Å. There are significant local differences throughout the protein, with clear evidence for conformational changes resulting from contact zones in the crystal lattice. Inspection of the map reveals that although densities for residues with positively charged and neutral side chains are well resolved, systematically weaker densities are observed for residues with negatively charged side chains. We show that the weaker densities for negatively charged residues arise from their greater sensitivity to radiation damage from electron irradiation as determined by comparison of density maps obtained by using electron doses ranging from 10 to 30 e/Å\(^2\). In summary, we establish that it is feasible to use cryo-EM to determine near-atomic resolution structures of protein complexes (∼500 kDa) with low symmetry, and that the residue-specific radiation damage that occurs with increasing electron dose can be monitored by using dose fractionation tools available with direct electron detector technology.

Significance

Atomic resolution models for proteins and protein complexes are usually obtained using X-ray crystallography or NMR spectroscopy, and in selected instances, by cryo-electron microscopy (cryo-EM) of ordered protein assemblies. The vast majority of high-resolution structures obtained using cryo-EM have been typically restricted to large, well-ordered entities such as helical or icosahedral assemblies or two-dimensional crystals. We show here that emerging methods in single-particle cryo-EM now allow structure determination at near-atomic resolution, even for much smaller protein complexes with low symmetry, by determining the structure of the 465-kDa enzyme β-galactosidase. In addition, by quantitative comparison of density maps obtained at different electron dosages, we demonstrate preferential sensitivity of residues such as Asp and Glu to damage upon irradiation with electrons.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. Freely available online through the PNAS open access option.

Data deposition: Density map derived by cryo-electron microscopy and the fitted atomic model have been deposited in the EMDataBank (EMDB), www.emdatabank.org (EMDB ID code EMD-5995), and the Protein Data Bank (PDB), www.pdb.org (PDB ID code 3J7H), respectively.

1To whom correspondence should be addressed. Email: s1@nih.gov.

This article contains supporting information online at www.pnas.org/suppl/supplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1402809111

PNAS | August 12, 2014 | vol. 111 | no. 32 | 11709–11714
Images were collected over a time period of 15 s, and the integrated exposure obtained after alignment of individual frames show well-dispersed particles and a variety of distinct views (Figs. S1 and S2). Processing of the individual frames recorded over the course of the exposure allows correction for beam-induced motion, which frequently leads to significant improvements in image quality, as in the example presented in Fig. S3 (see also Figs. S4 and S5). The information in the aligned images extends to resolutions close to ∼3 Å, as illustrated by the Thon rings visible in the Fourier transform of images and in the 1D power spectrum profile (Figs. S1 and S2). Automated particle selection followed by reference-free 2D classification (Fig. S6) was used to obtain class averages suitable for de novo determination of an initial map at ∼20-Å resolution by using EMAN2 (22). This map was used as a starting model for refinement of the structure by using FREALIGN (23). The final map, obtained from 11,726 particles, following 16 cycles of refinement (the first 8 by using binned, and the last 8 by using unbinned images) is shown in Fig. 1A, and has a resolution of ∼3.2 Å, as estimated by the gold standard (24) Fourier shell correlation (FSC) (Fig. 1B). Further, the resolution at which an FSC value of 0.5 is obtained between the experimental cryo-EM map and the map computed from the corresponding atomic model is also ∼3.2 Å (Fig. 1B), providing independent confirmation of map resolution. The FSC curve is different when the cryo-EM density map is compared with the map derived from the crystal structure (3.6 Å vs. 3.2 Å; Fig. 1B), providing a measure of the differences between the atomic models derived by cryo-EM and X-ray crystallography.

There are numerous structures for β-galactosidase determined by X-ray crystallography, the majority of which use constructs where the N-terminal 9 or 12 residues are replaced with a different sequence to facilitate crystallization (Table S1). Processing of the individual frames recorded over the course of the exposure allows correction for beam-induced motion, which frequently leads to significant improvements in image quality, as in the example presented in Fig. S3 (see also Figs. S4 and S5). The information in the aligned images extends to resolutions close to ∼3 Å, as illustrated by the Thon rings visible in the Fourier transform of images and in the 1D power spectrum profile (Figs. S1 and S2). Automated particle selection followed by reference-free 2D classification (Fig. S6) was used to obtain class averages suitable for de novo determination of an initial map at ∼20-Å resolution by using EMAN2 (22). This map was used as a starting model for refinement of the structure by using FREALIGN (23). The final map, obtained from 11,726 particles, following 16 cycles of refinement (the first 8 by using binned, and the last 8 by using unbinned images) is shown in Fig. 1A, and has a resolution of ∼3.2 Å, as estimated by the gold standard (24) Fourier shell correlation (FSC) (Fig. 1B). Further, the resolution at which an FSC value of 0.5 is obtained between the experimental cryo-EM map and the map computed from the corresponding atomic model is also ∼3.2 Å (Fig. 1B), providing independent confirmation of map resolution. The FSC curve is different when the cryo-EM density map is compared with the map derived from the crystal structure (3.6 Å vs. 3.2 Å; Fig. 1B), providing a measure of the differences between the atomic models derived by cryo-EM and X-ray crystallography.

Fig. 1. Cryo-EM density map of *E. coli* β-galactosidase at 3.2-Å resolution. (A) Surface representation of the density map derived by cryo-EM. The dimensions of the β-galactosidase tetramer are ∼180 Å x 140 Å x 87 Å. (B) Three different FSC plots to estimate resolution of the map shown in A. The curve in red shows the gold-standard FSC (24), with a value of 0.143 at 3.2-Å resolution, calculated between two independently refined halves of the dataset; the curve in blue shows the FSC (value of 0.5 at 3.6-Å resolution) calculated between the map in A and the map derived from the structure derived by X-ray crystallography PDB IDs code 1DP0 (21); and the curve in green shows the FSC (value of 0.5 at 3.2-Å resolution) calculated between the map in A and the map derived from the cryo-EM-derived atomic model.

Fig. 2. Atomic model for β-galactosidase tetramer derived by cryo-EM. (A) Cryo-EM map shown in surface representation with fitted atomic model of β-galactosidase. The four labeled protomers are colored magenta, black, green, and blue, respectively. The density corresponding to the 12 N-terminal residues of protomers 1 and 4 shown in mosaic12 are shown in the boxed region to highlight the absence of these residues in 96% of the 57 available X-ray structures. (B) Zoomed in region of the dimer interface highlighted in A showing a superposition of the cryo-EM-derived atomic model derived de novo from the experimental density map (N1 and N4 shown in magenta and blue, respectively); the cryo-EM-derived atomic model closely matches that derived in the X-ray models that report coordinates for this region (only 2 of the 57 reported structures).
atomic structural model of full-length β-galactosidase (Fig. 2 and Fig. S7).

Detailed comparison of the atomic model derived by cryo-EM with several X-ray structures shows small, but significant differences (Fig. 3). For example, the rmsd of all Cα atoms relative to the 1.7 Å resolution X-ray structure (PDB ID code 1DP0) is only ~0.66 Å (Fig. S8), but it is worth noting that individual Cα atoms differ by as much as 4.6 Å. As expected, the best agreement is in the more central regions, and the greatest differences are in the periphery of the tetramer (Fig. 3A). Interestingly, some of the largest deviations occur precisely at zones of crystal contacts (see also Figs. S9 and S10). Comparison of the crystal structure with the structure of the protein in solution in the absence of lattice contacts shows a measurable shift of the Cα trace and an alteration of the local side-chain conformations at these zones (Fig. 3B and C), although significant deviations are also observed elsewhere in the protein (Fig. 3D). These differences show that contact with neighboring molecules in the crystal results both in stabilization and local perturbation of the protein structure. Overall, this comparison confirms the potential of cryo-EM for de novo protein structure determination, especially in instances in which deletions, mutations, and crystal contacts can result in differences in conformation between the solution structure and the crystallized form.

Inspection of the cryo-EM density map reveals that, as expected from the estimated resolution, side-chain densities in β-strand and α-helical regions are well resolved (Fig. 4A and B) with densities comparable to those observed for the same regions in an X-ray crystallographic structure of β-galactosidase (Fig. S11). The structure of the active site, including the central His and Glu residues and the catalytic Mg2+ ion, can also be clearly discerned (Fig. 4C). The coordinating water molecules and Na+ ions could not be identified at our present resolution. Nevertheless, the quality of the map indicates that single-particle cryo-EM methods can now furnish sufficiently high resolution to provide mechanistic information on the structure of the active site under conditions where the enzyme is dispersed in solution and is not in the environment of a 3D crystal. Clearly resolved densities are observed not merely for bulky side chains such as tryptophan and tyrosine, but also for extended side chains such as arginine and lysine and smaller side chains such as Leu, Ile, and Val (Figs. 3–5). Interestingly, densities for positively charged and neutral residues are clearly resolved in virtually all instances (see selected Arg, Lys, His, and Gln residues shown in Fig. 5). However, densities for negatively charged residues, Glu and Asp, are typically weaker (Fig. 5B and Fig. S12). The residues chosen in Fig. 5B are from different regions of the protein and include those that are buried (e.g., Glu-979 and Asp-172) as well as those exposed to solvent (e.g., Glu-198 and Asp-130), indicating that these trends are independent of residue location. One way to quantitatively explore the trends in density values for different residue types in the map is to analyze the variation in map values as shown in Fig. 5C. These measurements add more depth to the observations already presented in the isosurface...
In Fig. S12, which shows that increasing the dose from 10 e⁻/Å² results to electron irradiation is also presented quantitatively are shown in Fig. 6). The preferential sensitivity of Asp and Glu residues compared with Asn and Gln residues, respectively. The trends we observe for preferential radiation damage showing increased loss of density (26%, 22%, and 10% for Glu-67, Glu-243, and Asp-96) are much more sensitive to radiation damage by comparison of maps obtained with progressively higher dose-fractionation capabilities of current direct detector technology (Fig. 6). This comparison reveals that there is a systematic loss of density for the negatively charged side chains in comparison with the similarly sized, but neutral side chains. (C) Plot of normalized averaged map values (as detailed in Fig. S12), shown for each residue type when only buried residues are considered (hatched bars) or when all residues including buried and exposed varieties are considered (filled bars). Buried residues are defined as those where <30% of their maximum possible surface area is exposed to solvent. Figure 5.

representations of the density map and show that Glu and Asp residues display, on average, ~30% and ~29% less density than the similarly sized neutral Gln and Asn residues, respectively. Further, we find that the pattern of preference for radiation damage is independent of the solvent exposure of the residues (Fig. 5C).

It has been proposed that a possible origin of the weaker densities for Glu and Asp residues is that the differences between X-ray and electron scattering factors result in a lower signal at medium resolutions from negatively charged, deprotonated carboxylate moieties (26). An alternative, more likely reason is that Glu and Asp side chains are more sensitive to damage that occurs with electron irradiation (27–30). Selective sensitivity of negatively charged residues such as Glu and Asp, and disulfide bonds (of which there are none in the E. coli β-galactosidase), has also been observed for structures determined by X-ray crystallography (31–33). To evaluate this possibility, we compared the density maps reconstructed from different fractions of the total exposure (10, 20, or 30 e⁻/Å²), taking advantage of the dose-fractionation capabilities of current direct detector technology (Fig. 6). This comparison reveals that there is a systematic loss of the side-chain density for Glu and Asp residues with increasing dose, and that this loss is not observed for positively charged or neutral residues in the same degree (selected examples are shown in Fig. 6). The preferential sensitivity of Asp and Glu residues to electron irradiation is also presented quantitatively in Fig. S12, which shows that increasing the dose from 10 e⁻/Å² to 45 e⁻/Å² results in an ~twofold increase in preferential damage of Asp and Glu residues compared with Asn and Gln residues, respectively. The trends we observe for preferential radiation damage of Asp and Glu residues can also be demonstrated by similar analysis of map value variations in a recently reported cryo-EM density map for Frh (10), an unrelated protein, at a resolution of 3.36 Å (Fig. S124).

In principle, there is no reason why the success in achieving near-atomic resolution structures for the proteasome at 3.3 Å (11), mammalian TRPV1 channel with regions at 3.4 Å (12), and now β-galactosidase at 3.2-Å resolution, cannot be extended to other proteins of similar, or even smaller size. One likely prerequisite for this approach to be applied successfully is that the protein sample is structurally homogenous even in the absence of stabilization provided by a crystal lattice. NMR spectroscopic studies of proteins suggest that most proteins dispersed in aqueous solution will not be uniformly ordered in all regions of the polypeptide, and are likely to be more ordered at their core and less ordered at the periphery, where there may be greater contact with the solvent. This type of variability does not necessarily preclude structure determination by cryo-EM, but simply results in a gradient in resolution from the center to the periphery of the molecule, as seen in the case of the density map of the TRPV1 channel, where the resolution declines from 3.4 Å at the center to lower values on the periphery, preventing the visualization of side-chain densities (12). Nevertheless, the presence of this type of gradient in resolution may be informative in its own and could identify regions of the protein that are more flexible than others. Additionally, the use of Fab fragments, interacting proteins, or small molecule ligands to stabilize conformations could help in some instances to stabilize the conformation to enable cryo-EM analyses (34, 35). Further, as methods for image processing and 3D classification continue to evolve (36–38), single-particle cryo-EM analyses can be expected to routinely yield not just a single averaged structure, but an ensemble of conformations present in solution (39), thus complementing higher resolution structural studies of specific, stabilized conformations by X-ray crystallography.

The methods used to prepare cryo-EM specimens and obtain raw electron micrographs are also likely to have a significant impact on the quality of the final map(s) obtained. Dispersion of protein complexes in a layer of ice as thin as possible, to
minimize background contribution from the vitrified buffer, and the use of electron optical conditions that allow recording of images at the highest possible resolution, are prerequisites for obtaining high-resolution maps. In addition, other factors unique to the protein may be especially relevant. Some proteins tend to accumulate at the air-water interface, with potential for partial denaturation. In turn, this nonrandom distribution can lead to aberrant projection images, preferential orientations relative to the vitreous ice layer, and, ultimately, to lower resolution. Some protein complexes that are relatively stable in bulk solution may also be more prone to dissociation when it is subject to interfacial forces during the blotting process that precedes plunge freezing.

Beam-induced specimen movement has long been recognized as an important factor in degradation of image quality (29). The development of direct electron detectors that allow dose fractionation now provides a way to begin to resolve this problem (40, 41). By collecting dose fractionated images, the total exposure can be divided into a large number of images, each at a very small dose, enabling correction for drift resulting from beam-induced movement and also from thermal drifts of the specimen stage. The emerging consensus from a number of recent studies (10, 11, 15, 40), which is confirmed in our present analysis, is that, on average, the largest movements tend to occur early in the exposure (Fig. S3); these early frames can be either discarded or down-weighted during image processing. Similarly, later frames collected during the exposure are expected to display greater radiation damage effects and can also be discarded to preserve high-resolution features, as shown in our analysis. Our finding that the use of just the first 10 e⁻/Å² produces density maps with the least radiation damage is consistent with estimates of the optimal dose to obtain density maps at 3-Å resolution by using cryo-EM (42). In addition to reducing the effects of beam-induced motion and mitigating the effects of radiation damage, we show that dose fractionation can also be used to improve the accuracy of contrast transfer function (CTF) estimation, which is an important resolution-limiting factor in single-particle cryo-EM (ref. 43; Fig. S5). Finally, the images we used in our analysis were recorded by using a direct electron detector at the end of an energy filter, demonstrating that the use of the filter does not limit structure determination to resolutions of ~3 Å, and may even be beneficial because of the increased contrast of energy-filtered images.

The advances in cryo-EM that allow determination of structures of small protein complexes and membrane proteins at near-atomic resolution mark a critical shift in structural biology. It is also encouraging that we were able to determine the structure at ~3.2-Å resolution from as few as ~12,000 projection molecular images, bringing it closer than ever before to the theoretical limit identified by Henderson (44) for the number of “perfect” images needed to obtain atomic resolution from cryo-EM. Structure determination via X-ray diffraction requires that proteins be coaxed into conditions that enable formation of well-ordered 3D crystals; however, in some cases, the process of crystallization can significantly alter the native solution conformation of the protein. NMR spectroscopy provides an alternative route to determine solution structures, but in practice, it is not feasible for molecules larger than ~50 kDa. With cryo-EM, structures of protein complexes as large as ~90 MDa and as small as ~300 kDa have now been determined at near-atomic resolution. There is every reason to hope that the structures of much smaller proteins will also be determined by cryo-EM methods. Because specimens for cryo-EM are plunge frozen rapidly from the aqueous phase, the cryo-EM density reflects the averaged, most probable protein conformation in solution and, in combination with image classification, can provide structures of a range of conformations that are sampled by the protein. Even when a number of crystal structures are already available, the determination of an atomic resolution model that captures the conformation and active site geometry in solution can be important for detailed quantitative analysis of the reaction mechanism. The prospect that the determination of protein structures to atomic resolution will no longer be limited by size or by the need for crystallization represents a significant and exciting horizon in structural biology. The stage is now set for the application of these methods to analyze structures of a variety of biologically and medically relevant multiprotein complexes and membrane protein assemblies, which have historically represented the most challenging frontier in structural biology.

**Materials and Methods**

**Sample Preparation.** β-galactosidase (Enzyme Commission 3.2.1.23; catalog no. G5635; Sigma-Aldrich) was subjected to gel filtration on a Superdex-200 size-exclusion chromatography column connected to an AKTA FPLC apparatus (GE Healthcare Bio-Sciences) with an elution buffer comprised of 25 mM Tris at pH 8, 50 mM NaCl, 2 mM MgCl₂, and 0.5 mM tri(2-carboxyethyl)phosphine. Solution samples were deposited on 200 mesh Quantifoil R2/2 grids (Quantifoil Micro Tools) and plunge frozen by using a Leica EM GP instrument (Leica Microsystems).

**Data Acquisition.** The grids were imaged by using a Titan Krios transmission electron microscope (FEI Company) aligned for parallel illumination and operated at 300 kV, with the specimen maintained at liquid nitrogen temperature. Individual images were recorded by using a K2 Summit camera (Gatan) operated in superresolution counting mode with a physical pixel size of 1.275 Å and a superresolution pixel size of 0.6375 Å. The detector was placed at the end of a GIF Quantum energy filter (Gatan), operated in zero-energy-loss mode with a slit width of 20 eV. The dose rate used was ~5 e⁻/Å²·s (equivalent to ~3 e⁻/Å²·s at the specimen plane) to ensure operation in the linear range of the detector (11). The total exposure time was 15.2 s, and intermediate frames were recorded every 0.4 s giving an accumulated dose of ~45 e⁻/Å² and a total of 38 frames per image.

**Image Processing.** From a dataset of 834 micrographs, with defocus values ranging from ~1.0 to ~2.5 μm, a subset of 509 micrographs displaying the highest resolution Thon ring profiles, least astigmatism, and clearly visible, well-separated particles were then selected for further analysis. To compensate for drift and beam-induced motion during the electron exposure, individual frames of each movie were aligned by cross-correlation using the cumulative average of previously aligned frames as a reference to align the remaining frames (for more details, see SI Materials and Methods). The increased signal-to-noise ratio of the cumulative average of frames, which was used as a reference to align the raw frames, results in significant improvements in the accuracy of motion correction as evidenced by better defined Thon rings in the drift-corrected images (Fig. S3). In terms of the cross-correlation between power spectra extracted from the corrected averages and the theoretical CTF profiles, our results compare favorably with those produced by the technique presented in ref. 11, which uses redundant measurements between raw frames to improve the consistency of motion correction (Fig. S4).

**Accuracy of CTF determination** is one of the main factors that impact resolution in single-particle cryo-EM, especially in the near-atomic resolution regime. Taking advantage of the greater ability of direct electron detectors to produce dose fractionated movies, we show that in addition to providing opportunities for motion correction and mitigation of radiation damage effects, movies also provide a way to improve the accuracy of CTF determination and, therefore, the resolution of single-particle reconstructions. The way we accomplish this improved accuracy is by estimating the parameters of the contrast transfer function of each micrograph by using radially averaged power spectra obtained by periodogram averaging with tiles extracted from all frames of each movie. Compared with the standard approach of using tiles extracted from the average of frames (after drift correction), our method does not require frames to be prealigned and results in better defined 1D CTF profiles, allowing more accurate determination of defocus (Fig. S5 and for more details, see SI Materials and Methods). The improvement in the resolvability of zero crossings obtained by using this approach allowed us to set the upper resolution limit for CTF estimation at 3 Å while still providing reliable CTF fits.

**Model Refinement and Analysis.** Particles (24,750) were picked automatically from the best 509 micrographs by detecting the local maxima of
correlation of each image with a Gaussian disk of 100 Å in radius and used to build and refine the entire atomic model in the environment of UCSF Chimera (46) and averaged for each residue type. To distinguish between exposed and buried residues, the surface area of each residue was calculated by using SURF (47, 48). The maximum surface area for each residue type was calculated by using the GLY-XXX-GLY motif, analyzed for each residue (XXX) one at a time for each of the 20 aa.

To study the effects of radiation damage on the final map, a subset of particles that was used in the reconstruction was reextracted from the averages of the first 8, 16, and 24 frames, and the reconstruction using each set was obtained by using identical alignment parameters (i.e., Euler angles and displacements for each particle were unchanged). Figures were produced by UCSF Chimera (46) or PyMOL (The PyMOL Molecular Graphics System, Version 1.3., Schrödinger, LLC). The final map and refined atomic model have been deposited with the Electron Microscopy Data Bank (EMDB ID code EMD-3599) and Protein Data Bank (PDB ID code 3JTH), respectively.

ACKNOWLEDGMENTS. We thank Mrs. Lesley A. Earl and Mario J. Borgnia for helpful comments on the manuscript, Dr. Erin Tran for assistance with specimen preparation, and Robert Mueller for technical assistance with electron microscopy. The complete set of electron micrographs used to obtain the density map presented here is available through the Electron Microscopy Pilot Image Archive (EMPIAR). This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (Nih). This research was supported in part by funds from the Center for Cancer Research, National Cancer Institute, NIH.

Supporting Information

Bartesaghi et al. 10.1073/pnas.1402809111

SI Materials and Methods

Motion Correction. To compensate for drift and beam-induced motion during the electron exposure, individual frames of each movie were aligned by cross-correlation by using the cumulative average of previously aligned frames as a reference to align the remaining frames. Adopting principles similar to those used in iterative image alignment in single-particle cryo-EM (1), we start by aligning two consecutive frames to each other (typically the middle frames of each movie) and then use their aligned average as a reference to align the next frame. The average at each step is updated to include all of the previously aligned frames, and this process is repeated until all frames have been aligned. The entire process is then iterated until incremental changes in the estimated shifts of each frame are negligible; usually 5–8 iterations are sufficient to achieve convergence. All cross-correlation operations were implemented by using IMOD’s tiltcorr command using the cumulative option (2). For the comparison of our frame alignment strategy with the results produced by motioncorr (3) (Fig. S4), unbinned superresolution images were used as input to the program by using default parameter settings (increasing the offset for the number of frames used for comparison with values up to 10, using larger values for the B-factor parameter or binning of the data, did not improve the results). The corrected sum produced by motioncorr was then used as input to ctffind3 (4) with the range of frequencies used to fit the contrast transfer function (CTF) set from 50 to 3.5 Å. Identical settings were used to estimate the CTF of the uncorrected averages and the corrected averages by using our cumulative cross-correlation strategy.

CTF Estimation. Parameters of the contrast transfer function for each micrograph were obtained by using radially averaged power spectra obtained by periodogram averaging with tiles extracted from individual frames of each movie. The power spectra from individual movie frames were obtained by using tomos (5) and then radially averaged to estimate a single defocus value by using tomoctffind (5). Although the lower signal-to-noise ratio of individual frames is known to deteriorate the signal in the power spectra (6), the subsequent radial averaging operation results in a significant increase in the cross-correlation values against the 1D theoretical CTF profile (Fig. S5). This improvement is attributed to the combination of the radial averaging operation (assuming astigmatism is negligible), with the fact that the power spectrum obtained from the individual frames has no influence of either local or global drift during the exposure (due to the shift invariant property of the power spectrum) and does not suffer from dampening of signal at high resolution due to interpolation and averaging operations needed to produce drift-corrected images.

Model Refinement. Side-chain densities were visualized clearly throughout the map, allowing us to build and refine the entire atomic model in the environment of the program COOT (7). More specifically, rigid body fitting of an X-ray structure [Protein Data Bank (PDB) ID code 1DP0] to the map was first carried out by using Chimera (8). The transformed structure was used as a starting point to the refinement in COOT (version 0.7.1), which was possible after translating the map to a new origin (0,0,0), using EMAN. The structure was refined into the density map in a semimanual way by using planar peptide and Ramachandran restraints, no secondary structure restraints, with a weight matrix of 60% during multiple rounds of processing for the entire 1,023-aa peptide chain. Side-chain rotamers were tested for their consistency with the density. The N-terminal domain, and multiple segments of poor agreement with the initial structure, were constructed de novo. Noncrystallographic D2 symmetry was applied at the end.

Fig. S1. Cryo-EM of *Escherichia coli* β-galactosidase. (A) Average of 38 aligned movie frames acquired over a 15.2-s exposure window (45 e⁻/Å² total accumulated dose) with images recorded at 300 kV and −1.7 μm defocus in superresolution mode by using a physical pixel size of 1.275 Å. (Scale bar: 50 nm.) (B) Fast Fourier transform (FFT) of image in A showing the extent of Thon rings present in the data extending to ∼2/3 of the edge of the transform, which corresponds to 2.55 Å. (C) One-dimensional power spectrum profile obtained by radially averaging the Fourier transform of the image in A; peaks are clearly visible out to 3.2-Å resolution.
Fig. S2. Cryo-EM imaging of E. coli β-galactosidase at 300 kV in superresolution mode by using a physical pixel size of 1.275 Å. Averages of 38 aligned movie frames acquired over a 15.2-s exposure window (45 e−/Å² total accumulated dose) using −1 μm (A) and −2 μm (C) defocus and corresponding power spectra (obtained by periodogram averaging using tiles from all frames) showing the resolution to which Thon rings are visible (B and D). The edge of the transform corresponds to 2.55 Å, and only the nonredundant half of the FFT is shown. (Scale bars: 50 nm.)
Fig. S3. Improvement in image quality obtained by aligning individual movie frames. (A and B) Comparison between averages of unaligned (A) and aligned (B) movie frames showing correction of the blurring effect caused by beam-induced motion (extreme case). (Scale bars: 25 nm.) (C) Trajectory of movement for each of the 38 frames during the 15.2-s exposure used in A. Displacements are measured in multiples of the superresolution pixel (1 pixel = 0.6375 Å). (D) Comparison between the FFTs of the averages of unaligned (Left) and aligned (Right) movie frames showing recovery of uniform Thon rings after correction for image movement during the exposure.

Fig. S4. Improved accuracy of frame alignment by using a cumulative cross-correlation approach. Distribution of cross-correlation values for the dataset of 509 images reported by ctffind3 (1) between theoretical CTF curves and power spectra obtained from uncorrected frame averages (cyan), corrected-averages using motioncorr (2) (blue), and corrected averages using our cumulative cross-correlation strategy (green). The plots show that the use of cumulative cross-correlation increases the accuracy of frame alignment beyond that achieved by motioncorr (2).

Fig. S5. Use of intermediate frames to improve accuracy of CTF determination. (A) Radially averaged power spectra used for CTF determination obtained from the aligned average of frames (blue) and from the raw individual frames without alignment (green). Cross-correlation values reported by tomoctffind (1) against the theoretical CTF curve at $-1.7 \mu m$ defocus improve from 0.64 to 0.86 when using the individual frames. (B) Distribution of cross-correlation values reported by tomoctffind for the set of 509 images using frame averages corrected by cumulative cross-correlation (blue), and for power spectra obtained from individual movie frames (green), showing a substantial improvement in the quality of the CTF fits and increase in accuracy of defocus estimation when using the individual movie frames.

Fig. S6. Two-dimensional reference-free classification and ab initio generation of an initial model. (A) Representative 2D class averages of cryo-EM β-galactosidase images obtained with the program e2refine2d.py in the EMAN2 suite shows characteristic views of β-galactosidase. (B) Alternating class averages (odd-numbered columns) and corresponding reprojections of the ab initio 3D model obtained with e2initialmodel.py (even-numbered columns) show a good match between computed model projections and reference-free class averages.
Fig. S7. Ramachandran plot of the atomic model of the E. coli β-galactosidase structure derived from cryo-EM data, indicating 4,004 residues (>98%) in preferred and 72 residues (<2%) in allowed regions.

Fig. S8. Histogram showing the distribution of rmsd values of individual Cα atoms between corresponding residues in the cryo-EM and X-ray (PDB ID code 1DP0) structures. Values range between 0.04 and 4.64 Å, with an average of 0.66 Å (yellow dashed line).
Fig. S9. Residues involved in crystal contacts. (A) Visualization of one of the crystal contacts (region ** in Fig. 3) highlighting the intermolecular interaction (indicated by the arrow) of two β-gal tetramers (shown in gray and orange) in the 3D crystal lattice (PDB ID code 1DP0). The presence of this crystal contact correlates with the difference in local backbone conformation between the crystal and cryo-EM structures. (B) Detailed view of the interaction between a peripheral loop from one tetramer (shown in gray) with three distinct regions on the neighboring tetramer in the crystal lattice.
Fig. S10. Comparison between cryo-EM and X-ray structures of E. coli β-galactosidase. (A) rmsd values of all Cα atoms between corresponding residues in five representative X-ray models [PDB ID codes 1JYX, 1BGL, 1DP0, 3DYP, and 1HN1] relative to the cryo-EM–derived atomic model are plotted according to the amino acid sequence. These five structures were chosen based on uniqueness of space group (1BGL is monoclinic and the rest are orthorhombic), presence of a ligand (1JYX is complexed with IPTG), best and worst resolutions of the determined structures (1DPO at 1.70 Å and 1HN1 at 3.0 Å), and an active-site mutant structure (3DYP). (B) X-ray structure color-coded for variation in rmsd values. (C) X-ray structure color-coded to display regions of the protein in close contact with neighboring molecules in the crystals used to obtain the five X-ray structures. Atoms that are within 4 Å of neighboring tetramers in the crystal lattice are marked as red spheres, indicating areas of protein-protein contact in the crystal lattice.
Fig. S11. Electron density map of β-galactosidase derived by X-ray crystallography at 3.0-Å resolution. To compare the cryo-EM density map at 3.2-Å resolution with a map obtained at comparable resolution using X-ray crystallography, 2F_o-F_c maps obtained from the X-ray structure at 3.0-Å resolution (PDB ID code 1HN1) of the same regions of the protein shown in Figs. 4 and 5 are presented. Residues 849–856 in a β-strand (A), residues 395–405 in an α-helix (B), and residues 413–420 in the active site along with density for a Mg^{2+} ion partly coordinated by His-418 and Glu-416 (C) (compare A–C with corresponding images shown in Fig. 4 A–C for the cryo-EM density map). (D) Densities for the same set of Arg residues as shown in Fig. 5 for the cryo-EM density map are shown for the 2F_o-F_c map corresponding to the 1HN1 X-ray structure.
Fig. S12. Proportional variation in side-chain density measured for each residue type on different cryo-EM maps. (A) Plots of normalized average map values for each type of residue for the ∼3.2-Å map of β-galactosidase (β-gal, Upper, black) and the 3.36-Å map of F420-reducing [NiFe] hydrogenase (1) (Frh, Lower, green). Density for each type of residue is measured by averaging density map intensities sampled at the position of each nonhydrogen atom for all residues of that type. Values are normalized with respect to the mean density over all residues, resulting in numbers greater than 1 for values above average and numbers smaller than 1 for values below average. SD values are shown as intervals on all plots. Loss of density as a result of radiation damage is seen dominantly in negatively charged residues, Asp and Glu, independent of their exposure to solvent. All Asp (marked by asterisk) and Glu (marked by diamond) residues show on average 30% and 29% less density than the corresponding Asn (asterisk) and Gln (diamond) residues in the case of β-gal, and 35% and 29% in the case of Frh. Most of the cysteine residues in Frh are bonded to FeS clusters, suggesting a possible explanation for the increased averaged map values for this type of residue in Frh compared with β-gal. (B) Plots of normalized average map values per residue on three β-gal maps using a total dose of 10 (blue), 20 (cyan), or 30 e⁻/Å² (purple). Proportional loss of density at Glu and Asp with respect to Gln or Asn increases with higher total dose, with 18% less density measured for Glu and Asp compared with Gln or Asn in the map calculated at ∼10 e⁻/Å², 23% at ∼20 e⁻/Å², and 24% at ∼30 e⁻/Å². This tendency is also shown in Fig. 6 on selected side chains with map density visualized as iso-surfaces.

<table>
<thead>
<tr>
<th>PDB ID code</th>
<th>Resolution, Å</th>
<th>N-terminal residues of construct used for protein expression</th>
<th>N-terminal residues in PDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>4DUV</td>
<td>2.3</td>
<td>MGGSHHHHHHGMASMTGGQGMGRDLYDDDIDMDPYVLQRRRDWNPGV</td>
<td>MIDPVLQRRRDWNPGV</td>
</tr>
<tr>
<td>3VD3</td>
<td>2.8</td>
<td>MGGSSHHHHHGMASMTGGQGMGRDLYDDDIDMDPYVLQRRRDWNPGV</td>
<td>RDWNPGV</td>
</tr>
<tr>
<td>3SEP</td>
<td>2.1</td>
<td>MGGSSHHHHHGMASMTGGQGMGRDLYDDDIDMDPYVLQRRRDWNPGV</td>
<td>VVLQRRRDWNPGV</td>
</tr>
<tr>
<td>3MVO</td>
<td>2.2</td>
<td>MGGSHHHHHHGMASMTGGQGMGRDLYDDDIDMDPYVLQRRRDWNPGV</td>
<td>VVLQRRRDWNPGV</td>
</tr>
<tr>
<td>3MUY</td>
<td>2.5</td>
<td>GSHMLEDPVVLQRRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>3IAP</td>
<td>2.0</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>3I3E</td>
<td>2.0</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>3DYP</td>
<td>1.8</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>3EIF</td>
<td>3.0</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>1PX3</td>
<td>1.6</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>1JYX</td>
<td>1.5</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>1HN1</td>
<td>3.0</td>
<td>GSHMLEDPVVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>1F4H</td>
<td>2.8</td>
<td>ITDSLAVVLQRRRDWNPGV</td>
<td>ITDSLAVVLQRRRDWNPGV</td>
</tr>
<tr>
<td>1DP0</td>
<td>1.7</td>
<td>GSHMLEDPVLQRRDWNPGV</td>
<td>RRDNPGV</td>
</tr>
<tr>
<td>1BGL</td>
<td>2.5</td>
<td>TMITDSLAVVLQRRDWNPGV</td>
<td>ITDSLAVVLQRRRDWNPGV</td>
</tr>
</tbody>
</table>

Conserved residues are underlined.