Achieving better-than-3-Å resolution by single-particle cryo-EM at 200 keV

Mark A Herzik Jr[1,2], Mengyu Wu[1,2] & Gabriel C Lander[1]

Nearly all single-particle cryo-EM structures resolved to better than 4-Å resolution have been determined using 300-keV transmission electron microscopes (TEMs). We demonstrate that it is possible to obtain reconstructions of macromolecular complexes of different sizes to better than 3-Å resolution using a 200-keV TEM. These structures are of sufficient quality to unambiguously assign amino acid rotameric conformations and identify ordered water molecules.

Recent technical breakthroughs in the field of three-dimensional (3D) single-particle cryo-electron microscopy (cryo-EM) have allowed for the visualization of biological macromolecular assemblies in near-native states at unprecedented resolutions[1,2]. To date, the vast majority of high-resolution cryo-EM reconstructions have been obtained using TEMs operating at 300 keV equipped with a direct electron detector (DED). TEMs operating at 300 keV offer minimized inelastic scattering and specimen charging compared to those at lower voltages, the benefits of which confer an advantage in imaging thicker specimens[3]. However, the high cost of purchasing and operating high-end TEMs may be unfeasible for many institutions.

Here we expand upon prior studies using 200-keV TEMs[4,5], and demonstrate that a Thermo Fisher Talos Arctica (200 keV) paired with a Gatan K2 Summit DED can produce cryo-EM reconstructions of *Thermoplasma acidophilum* 20S proteasome and rabbit muscle aldolase at ~3.1-Å and ~2.6-Å resolution, respectively. The resulting maps are of sufficient quality to identify ordered water molecules and unambiguously assign amino acid rotameric conformations. Our work demonstrates that, through use of a 200-keV TEM equipped with a DED, macromolecules of varying sizes and internal symmetries can be reconstructed at resolutions previously ascribed only to 300-keV microscopes.

All data were acquired using a base-model Talos Arctica TEM operating at 200 keV. Careful alignment of the TEM was performed immediately before each data collection to maximize parallel illumination[6] (Online Methods, **Supplementary Protocol** and **Supplementary Fig. 1**) and ensure Thon rings were visible beyond ~3-Å resolution in the power spectrum of images collected over amorphous carbon. Coma-free alignment was performed according to Glaeser *et al.*[6] immediately before data collection, as implemented in Leginon[7]. Data were collected using Leginon, and image pre-processing was performed using the Appion pipeline[8]. Images were acquired at a nominal magnification of 45,000× (calibrated pixel size of 0.91-Å at detector level) with the Gatan K2 Summit DED operating in super-resolution mode. To minimize the effects of electron-beam-induced motions, samples were prepared on gold grids[9], and exposures were acquired at the center of holes (1.2-μm diameter) using a beam (~2-μm diameter) large enough to cover the hole entirely and contact the surrounding gold substrate.

Mechanical and beam-induced motions were corrected, and dose weighting was performed using MotionCor2 (ref. 10). Whole-image contrast transfer function (CTF) estimation was performed using CTFFind4 (ref. 11), and micrographs yielding CTF estimates below an Appion confidence value of 0.9 were discarded. Per-particle CTF estimates calculated using gCTF[12] were used for all subsequent processing steps. Reference-free 2D classification and subsequent 3D classification and 3D auto-refinement were performed using RELION 2.0 (ref. 13). All reported resolutions are based on the gold-standard Fourier shell correlation (FSC) criterion[14] (Online Methods).

To assess the resolving capabilities of a Talos Arctica TEM, we sought to determine the structure of the *Thermoplasma acidophilum* 20S proteasome, a symmetric ~700-kDa protein complex with high thermal stability, previously determined to ~2.8-Å resolution using cryo-EM[15]. We collected 629 movies of frozen-hydrated 20S (**Fig. 1**) with a defocus range of −0.8 μm to −1.8 μm using stage movements to navigate to each desired exposure target. A final stack of 106,581 particles yielded a reconstruction estimated as ~3.1-Å resolution (**Fig. 1a**, **Supplementary Fig. 2** and **Online Methods**). Local resolution estimation of this EM density indicates that most of the map resolved to better than ~3.2-Å resolution, with the core of the molecule resolved to ~2.8-Å resolution (**Fig. 1a** and **Supplementary Fig. 2**). Indeed, an overlay of the ~2.8-Å 20S proteasome reconstruction previously solved using a 300-keV instrument (EMDB EMD-6287) reveals that the EM densities are of comparable quality, with both maps possessing clearly resolvable side-chain densities (**Supplementary Fig. 3**).

As an alternative to stage position navigation, we collected images using stage position to navigate to the center of four holes before image shifting the beam to each exposure target (**Supplementary Fig. 4**). This data collection strategy has the

benefit of being able to collect more micrographs per hour, on average, as compared to using stage position for navigation (**Supplementary Table 1**). However, substantial image shifting can introduce coma, which manifests as phase error in the structure factors. To determine the extent to which beam tilting affects the attainable resolution, we collected 394 videos of frozen-hydrated 20S using image shift to navigate to acquisition targets. Using a data processing and refinement strategy similar to that employed for the stage position data set, we obtained a final stack of 96,254 particles that yielded a reconstruction of ~3.3-Å resolution (**Fig. 1c** and **Supplementary Fig. 5**). An estimation of the local resolution of this EM density indicates that most of the map resolves to ~3.2-Å resolution or better, with parts of the core resolved to better than 3 Å (**Fig. 1c**).

Comparison of the two 20S reconstructions generated using our 200-keV TEM reveals that maps of comparable quality can be obtained, with a resolution loss less than ~0.2 Å, with image shift versus stage position for exposure target navigation. Visual inspection of the 20S EM densities indicated only modest differences in side-chain resolvability (**Fig. 1b,d**), suggesting that the extent of the coma introduced through our image shift navigation strategy minimally disrupts optimal image formation (**Supplementary Note 1**).

The best-resolved regions of the 20S reconstruction from the stage movement data set feature density that can be attributed to well-ordered water molecules (**Fig. 1**). Notably, the location of many of these waters correlates with those previously observed for the *T. acidophilum* 20S proteasome determined by cryo-EM (EMDB EMD-6287) and X-ray crystallography (PDB 1YAR) (**Supplementary Fig. 3**). Further examination of the density reveals that some side chains exhibit alternative conformations that lie within expected rotameric positions (**Fig. 1h**). To the best of our knowledge, these are the first observations of water molecules and/or alternative side-chain conformations in a reconstruction obtained using a 200-keV TEM.

To date, all but three (EMDB EMD-6337, EMD-8355, and EMD-8356) structures determined to resolutions better than 5-Å using a 200-keV TEM possess high internal symmetry (EMDB EMD-2791 and EMD-9565) and/or comprise a molecular weight of ≥700 kDa (EMDB EMD-2792, EMD-6396, EMD-6397, EMD-6458, EMD-6584, EMD-8332, EMD-8333, EMD-8334, EMD-8402, and EMD-9572). In fact, only two structures <200 kDa have been reported to better than 3-Å resolution (EMDB EMD-8191 and EMD-3488), both determined using a 300-keV TEM. To determine whether near-atomic resolution was feasible for complexes in this size range using a 200-keV TEM, we sought to determine the structure of aldolase, a ~150-kDa homotetrameric enzyme (Online Methods).

We collected 784 recordings of frozen-hydrated aldolase (**Supplementary Fig. 6**) collected using acquisition settings similar to those used for the 20S proteasome. A final stack of 83,910 particles yielded a ~2.6-Å resolution reconstruction (Online Methods, **Fig. 2a,b** and **Supplementary Fig. 6**). Local resolution estimates reveals that most of the map was resolved to better than 2.8 Å, with the best-ordered regions of the map resolved to ~2.5 Å (**Fig. 2c**). The final reconstruction has clear side-chain density for most of the molecule and well-resolved backbone density (**Fig. 2d,e**). The best-refined model agrees well with previously published structures determined by
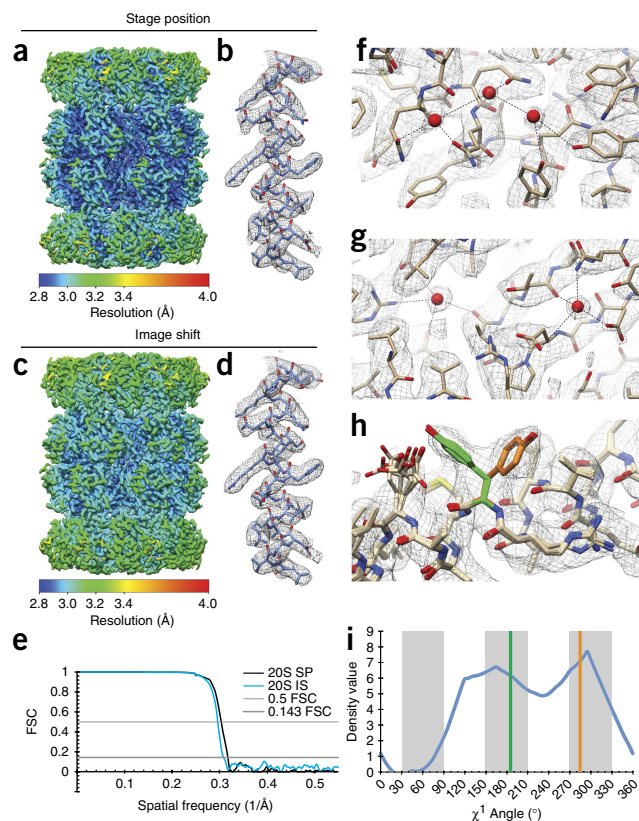

Figure 1 | Cryo-EM reconstructions of the 20S proteasome at ~3.1-Å resolution. (**a**–**d**) Local resolution estimates of the final *T. acidophilum* 20S proteasome reconstructions determined using either stage position (**a**) or image shift (**c**) navigation to exposure targets, with α-helices from the β-subunits shown in stick representation from each reconstruction (**b**,**d**), with corresponding EM density (gray mesh, zoned 2 Å within atoms). (**e**) Gold-standard FSC curves indicate a final resolution of ~3.1 Å for the stage position data set (SP; black) and ~3.3 Å for the image shift data set (IS; blue). (**f**,**g**) Ordered water molecules (red spheres) in the ~3-Å 20S EM density (gray mesh). Putative hydrogen bonds to the water molecules are shown as black dotted lines. (**h**) Tyr58 of the β-subunit adopts two alternative rotameric positions. The top 10 models refined against the density distribute into these rotameric positions (orange (60%) and green (40%)). (**i**) EMRinger[24] analysis of the EM density corresponding to Tyr58 (blue line) rotameric positions (thick gray bars). Vertical lines indicate refined χ^1 angles for Tyr58 (orange indicates conformation 1; green indicates conformation 2).

X-ray diffraction, with an overall r.m.s. deviation of 0.45 Å (PDB 6ALD). Further examination of the density revealed that the best-resolved regions of the map possess densities attributed to ordered water molecules (**Fig. 2f**). Moreover, the positions of these water molecules are conserved between those identified in PDB 6ALD (**Supplementary Fig. 7**).

TEMs operating at 300 keV have been shown to be optimal for imaging thicker biological specimens owing to reduced inelastic scattering, which limits image blurring associated with charging effects and radiolysis[3,16]. The deleterious effects of beam-induced motion and radiation damage can be accounted for computationally, through fractionation into movie frames using a DED[10,17–21]. For these reasons, 300-keV TEMs coupled with DEDs remain the flagship instruments in the field of high-resolution single-particle cryo-EM, accounting for nearly 98% of all reconstructions reported to 4-Å resolution or better. Our results demonstrate that
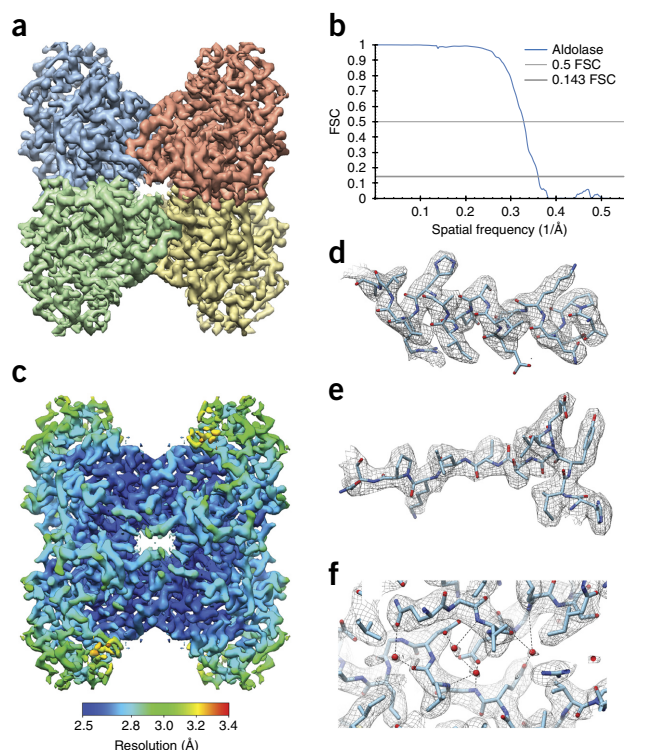
Figure 2 | Structure of aldolase at ~2.6-Å resolution. (**a**) The ~2.6-Å aldolase EM density (D2 symmetric) segmented on the basis of protomer organization. (**b**) Gold-standard FSC curve indicates a final resolution of ~2.6 Å at 0.143 FSC. (**c**) The local resolution estimate of the final aldolase reconstruction. Most of the molecule is resolved to better than 2.8 Å, with the core of the molecule resolved to ~2.5 Å. (**d**,**e**) Representative α-helix (**d**) and loop (**e**) within the aldolase EM density (gray mesh, zoned 2 Å within atoms) showing the unambiguous assignment of side-chain conformations. (**f**) The placement of putative ordered water molecules (red spheres).

cryo-EM reconstructions resolved to ~3-Å resolution and better are achievable using a 200-keV TEM coupled with a K2 Summit DED for samples of different sizes and internal symmetries. The maps generated herein are of sufficient quality for *de novo* model building as well as the identification of ordered waters and alternative side-chain conformations.

We sought to assess the resolution limit of our 200-keV instrument using the *T. acidophilum* 20S proteasome, which has served as a metric for characterizing the potential of TEMs and imaging accessories in numerous studies[2,4,15,18]. Notably, both of our 20S reconstructions using a 200-keV TEM showed clear density attributed to ordered water molecules. The presence of alternative amino acid rotamers in the density map further attests to the quality of the reconstruction.

In contrast to the 20S proteasome, which benefits from size, high internal symmetry and high thermostability, rabbit muscle aldolase represents a more challenging target for cryo-EM. Using illumination conditions similar to those used for determining the structures of 20S proteasome, we obtained a final aldolase reconstruction with a resolution of ~2.6 Å. While the smaller size (and thus fewer voxels) of aldolase used in the calculation of FSC may partly influence the higher reported resolution for aldolase relative to the 20S, it is also likely that the smaller size of aldolase contributes to a more isotropic resolution that is not negatively influenced by flexible peripheral domains. However, differences in ice thickness, overall

sample stability, and minor differences in microscope alignment cannot be ruled out as contributors.

Typically, micrometer underfocus values (e.g. ≥2 µm underfocus) have been prescribed for imaging targets <200 kDa[22]. More recently, strategies implementing phase plates combined with a small amount of defocus to boost image contrast have been demonstrated to be effective for resolving small protein complexes[23]. However, despite the small size of the aldolase particles used in this study, micrographs of sufficient contrast were obtained using underfocus values as low as −0.6 µm (**Supplementary Table 1**). We believe this is largely attributable to the thin ice of the aldolase specimen coupled with high particle density to boost particle contrast in the collected micrographs and facilitate alignment of movie frames. Our results suggest that it is feasible to achieve near-atomic-resolution reconstructions of a sub-200-kDa specimen using a conventional defocus approach (**Supplementary Note 2**).

We regard this work as a demonstration of the capabilities of a 200-keV TEM rather than a comparison of the benefits of operating at 200 versus 300-keV. Performing such a comparison would require great care in ensuring consistency in as many parameters as possible, including specimen quality, imaging conditions (for example, magnification, defocus range and pixel size) and data processing. Ideally, such experiments would be performed on a single grid using a microscope that is capable of operating at 200 or 300 keV.

We selected these specimens for their robust structural integrity and limited conformational heterogeneity to probe the resolution capabilities of a 200-keV TEM. Certainly, many macromolecular complexes will not be as well-behaved as those presented in this study, but our findings serve as a proof of principle. Given that 200-keV TEMs have substantially lower costs than 300-keV instruments, we hope that this work expands the field of high-resolution cryo-EM to a broader range of research institutes.

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.

AUTHOR CONTRIBUTIONS

M.A.H. and M.W. performed all cryo-EM experiments and analyses. M.A.H., M.W. and G.C.L. contributed to the experimental design and manuscript preparation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

BRIEF COMMUNICATIONS

1. Merk, A. *et al. Cell* **165**, 1698–1707 (2016).
2. Danev, R., Tegunov, D. & Baumeister, W. *eLife* **6**, e23006 (2017).
3. Henderson, R. *Q. Rev. Biophys.* **28**, 171–193 (1995).
4. Campbell, M.G. *et al. J. Struct. Biol.* **188**, 183–187 (2014).
5. Li, X. *et al. J. Mol. Biol.* **429**, 79–87 (2017).
6. Glaeser, R.M., Typke, D., Tiemeijer, P.C., Pulokas, J. & Cheng, A. *J. Struct. Biol.* **174**, 1–10 (2011).
7. Suloway, C. *et al. J. Struct. Biol.* **151**, 41–60 (2005).
8. Lander, G.C. *et al. J. Struct. Biol.* **166**, 95–102 (2009).
9. Russo, C.J. & Passmore, L.A. *Science* **346**, 1377–1380 (2014).
10. Zheng, S.Q. *et al. Nat. Methods* **14**, 331–332 (2017).
11. Rohou, A. & Grigorieff, N. *J. Struct. Biol.* **192**, 216–221 (2015).
12. Zhang, K. *J. Struct. Biol.* **193**, 1–12 (2016).
13. Kimanius, D., Forsberg, B.O., Scheres, S.H. & Lindahl, E. *eLife* **5**, e18722 (2016).
14. Henderson, R. *et al. Structure* **20**, 205–214 (2012).
15. Campbell, M.G., Veesler, D., Cheng, A., Potter, C.S. & Carragher, B. *eLife* **4**, e06380 (2015).
16. Egerton, R.F. *Ultramicroscopy* **145**, 85–93 (2014).
17. Campbell, M.G. *et al. Structure* **20**, 1823–1828 (2012).
18. Li, X. *et al. Nat. Methods* **10**, 584–590 (2013).
19. Rubinstein, J.L. & Brubaker, M.A. *J. Struct. Biol.* **192**, 188–195 (2015).
20. Grant, T. & Grigorieff, N. *eLife* **4**, e06980 (2015).
21. Scheres, S.H. *eLife* **3**, e03665 (2014).
22. Cheng, Y., Grigorieff, N., Penczek, P.A. & Walz, T. *Cell* **161**, 438–449 (2015).
23. Khoshouei, M., Radjainia, M., Baumeister, W. & Danev, R. *Nat. Commun.* **8**, 16099 (2017).
24. Barad, B.A. *et al. Nat. Methods* **12**, 943–946 (2015).

ONLINE METHODS

A step-by-step description of our procedures is available as a **Supplementary Protocol** and at *Protocol Exchange*[25].

Cryo-EM sample handling and grid preparation. Pure aldolase isolated from rabbit muscle was purchased as a lyophilized powder (Sigma-Aldrich) and solubilized in 20 mM HEPES (pH 7.5), 50 mM NaCl at ~3 mg/ml. Aldolase was further purified by size-exclusion chromatography using a Sepharose 6 10/300 (GE Healthcare) column equilibrated in solubilization buffer. Peak fractions were pooled and concentrated to 1.6 mg/ml immediately before cryo-EM grid preparation.

Archaeal 20S proteasome (*T. acidophilum*) was kindly donated by Z. Yu and Y. Cheng and used without further modification.

For cryo-EM, 3 μL purified aldolase (1.6 mg/ml) or 20S proteasome (0.5 mg/ml) were dispensed on freshly plasma cleaned UltrAuFoil R1.2/1.3 300-mesh grids (Electron Microscopy Services) and manually blotted[26] using a custom-built manual plunger in a cold room (≥95% relative humidity, 4 °C). Sample was blotted for ~7 s with Whatman No. 1 filter paper immediately before plunge freezing in liquid ethane cooled by liquid nitrogen. In order to provide enough signal for proper CTF estimation, we strived to achieve a particle concentration that maximized the number of particles contained within the holes without resulting in overlapping particles and/or aggregation (**Supplementary Note 3**).

Cryo-EM data acquisition, **image processing, and refinement.** Microscope alignments were performed on a cross-grating calibration grid. Condenser alignments and stigmation were performed as described for a two-condenser-lens system[27]. Proper eucentric height of the specimen was determined using Leginon before setting focus. Parallel illumination of the beam was maximized in diffraction mode by first adjusting the defocus to bring the objective aperture into focus in the front focal plane of the diffraction lens followed by adjustments of beam intensity to minimize the spread of gold powder diffraction[28] (**Supplementary Fig. 1**). The resulting beam-intensity value was saved in Leginon and remained unchanged throughout data collection. The objective aperture was centered, and any objective lens astigmatism was minimized. Exposure rate measurements on the K2 Summit DED were then collected to determine whether changes to spot size were necessary to achieve the desired exposure rate. In the event that changes in exposure rate were required, as was the case for increasing per frame cumulative dose during aldolase imaging, adjustments to spot size were made and alignments were repeated, including maximization of parallel illumination. Coma-free alignment was then performed as described previously[6]. Daily adjustments, if necessary, were made during data collection to maintain lens stigmation and ensure the beam was coma-free.

All cryo-EM data were acquired using the Leginon automated data-acquisition program[7]. All image pre-processing (frame alignment, CTF estimation, particle picking) were performed in real-time using the Appion image-processing pipeline[8] during data collection.

Images of frozen hydrated aldolase or 20S proteasome were collected on a Thermo Fisher Talos Arctica transmission electron microscope (TEM) operating at 200 keV. Movies were collected using a Gatan K2 Summit direct electron detector operating in super-resolution mode (super-resolution pixel size 0.455 Å/pixel) at a nominal magnification of 45,000× corresponding to a physical pixel size of 0.91 Å/pixel.

***Thermoplasma acidophilum* 20S proteasome.** 629 movies (68 frames/movie) of 20S proteasome were collected using stage position to navigate to the exposure target. Movies were collected over 17 s exposure with an exposure rate of 3.2 e⁻/pixel/s, resulting in a total exposure of ~65 e⁻/Å² (0.96 e²/Å²/frame) and a nominal defocus range from −0.8 μm to −2.8 μm. The same exposure settings were used to collect 394 movies using image shift target navigation. Super-resolution images were Fourier-binned 2 × 2 (0.91 Å/pixel) before motion correction and dose-weighting using the MotionCor2 frame alignment program[10] as part of the Appion pre-processing workflow. Frame alignment was performed on 5 × 5 tiled frames with a B-factor of 100 applied. Unweighted summed images were used for CTF determination using CTFFIND4 (ref. 11). DoG picker[29] was used to automatically pick particles from the first 10 dose-weighted micrographs yielding a stack of 4,821 particles that were binned 8 × 8 (3.64 Å/pixel, 92 pixel box size) and subjected to reference-free 2D classification using an iterative topology-representing network-based classification followed by multi-reference alignment[30] in the Appion pipeline. The best four classes were then used for template-based particle picking against the each data set using FindEM[31]. Only micrographs fulfilling the following two criteria: (i) possessing a CTF estimate confidence of fit ≥90% and (ii) with resolution estimates to 5 Å or better at CC = 50% (359 total from the stage position data set; 283 total from the image shift data set) were used.

For the 20S proteasome data set collected using stage position, 153,429 particles were extracted from the dose-weighted summed images, further binned 4 × 4 (3.64 Å/pixel, 92 pixel box size), and subjected to reference-free 2D classification using RELION 2.0 (ref. 13). The particles from the best-aligned classes (106,581 particles) were then 3D auto-refined with D7 symmetry imposed using EMD-6287 as an initial model (low-pass filtered to 60 Å) to yield a ~7.4-Å resolution reconstruction (7.28-Å Nyquist). Particles were then re-centered using the refined particle offsets before being re-extracted unbinned (0.91 Å/pixel, 368 pixel box size) for 3D auto-refinement using a scaled output from the 4 × 4 binned refinement as an initial model. The refinement was continued using a soft-mask (3 pixel extension, 5 pixel soft cosine edge) generated from a volume contoured to display the full density.

The final resolution was estimated as ~3.3 Å (gold-standard FSC)[14,32,33] using phase-randomization to account for the convolution effects of a solvent mask on the FSC between the two independently refined half maps[33]. The refined particle coordinates were then used for local CTF estimation using gCTF[12] followed by re-extraction of particles using a 512-pixel box size to limit aliasing in the Fourier domain. Gold-standard 3D auto-refinement using the same soft mask yielded a ~3.1-Å reconstruction (gold-standard FSC) (**Supplementary Fig. 2**). These particles were then subjected to 3D classification (k = 4, tau fudge = 25) without angular or translational searches using the same soft mask. However, further refinement of each individual class did not yield an improvement in nominal FSC-reported resolution. In addition, further filtering of the final particle stack using a cutoffs based on the height of the probability distributions at their maximum (MaxValueProbDistribution) did not result in an increase in nominal

FSC-reported resolution. Finally, removal of all particles with an estimated defocus of greater than 1.5 µm yielded a reconstruction with the same reported resolution (~3.1 Å), indicating that particles illuminated at greater defoci did not substantially contribute high-resolution information (**Supplementary Fig. 3**).

In order to eliminate as many non-optically related discrepancies as possible between the stage movement and image shift collections, all the data were collected using sample grids prepared and frozen in parallel during back-to-back sessions on the Talos Arctica TEM with similar acquisition parameters (magnification, exposure rate, cumulative dose, and so on). Processing and refinement of the 20S proteasome data set collected using image shift was performed similarly as described for the 20S proteasome data set collected using stage position (**Supplementary Fig. 5**).

Rabbit muscle aldolase. 810 movies (44 frames/movie) of rabbit muscle aldolase were acquired over 11 s with a exposure rate of 5.1 e−/pixel/s, yielding a total dose of ~68 e−/Å^2 (1.55 e−/Å^2/frame), and a nominal defocus range from −0.8 µm to −1.4 µm. 2 × 2 Fourier-binned super-resolution images (0.91 Å /pixel) were motion correction and dose weighted using the MotionCor2 program[10] using 5 × 5 tiled frames with a B-factor of 100 applied. Thon rings in the FFT of motion-corrected aldolase movies were visible to ~3 Å. Unweighted summed images were used for CTF determination using CTFFIND4 (ref. 11). Weighted sums were used for automated template-based particle picking with FindEM[31]. Only micrographs with CC ≥90% and resolution estimates of 5 Å or better at CC = 50% (659 total) were used. 1,009,341 particles were extracted from these micrographs and binned 4 × 4 (3.64 Å/pixel, 92 pixel box size). Reference-free 2D classification in RELION 2.0 (ref. 13) was then used to sort out non-particles and poor-quality picks in the data.

A total of 652,869 particles corresponding to 2D class averages that displayed strong secondary-structural elements were selected for homogenous *ab initio* model generation using cryoSPARC[34] to eliminate potential model bias. The generated volume was low-pass filtered to 30 Å and used as an initial model for 3D auto-refinement in RELION with D2 symmetry applied. The refined particle coordinates were then used for local CTF estimation using gCTF[12] followed by re-extraction of particles binned 2 × 2 (1.82 Å/pixel). Due to the close proximity of neighboring particles, any re-centered particle within a 32-pixel range of another was considered a duplicate and subsequently removed. The resulting stack, consisting of 606,278 particles, was subjected to 3D auto-refinement using the 4 × 4 binned map obtained from the previous refinement as an initial model. The refinement was continued using a soft mask (5-pixel extension, 10-pixel soft cosine edge) generated from a volume contoured to display the full density. The final resolution was estimated as ~3.6 Å (Nyquist = 3.64 Å, gold-standard FSC)[14,32] using phase randomization to account for the convolution effects of a solvent mask on the FSC between the two independently refined half-maps[33]. These particles were then subjected to 3D classification (k = 6, tau fudge = 8) without angular or translational searches using the same soft mask. Particles contributing to the classes that possessed the best resolved side-chain and backbone densities were re-centered, re-extracted, and cleaned of duplicates (83,910 particles, 0.91 Å/pixel, 368-pixel box size). These particles were then 3D auto-refined using a soft mask. The final resolution was estimated as ~2.6 Å (gold-standard FSC) (**Supplementary Fig. 1**).

Local resolutions for all reconstructions were calculated using the bloc_res function in BSOFT[35].

Model building and refinement. For each of the final reconstructions, an initial model was subjected to a multi-model pipeline using methodologies similar to those described previously[36]. Briefly, for the 20S proteasome reconstructions, PDB 1YAR was stripped of all ligands (i.e., PA26 and waters) and all alternative conformations, all occupancies were set to unity, a single B-factor value was set for all atoms, and all Ramachandran and geometric outliers were corrected. For aldolase, PDB 6ALD was stripped of all cofactors and water molecules, with all occupancies set to unity and all Ramachandran and geometric outliers corrected. Initial docking of PDB 6ALD into the density revealed additional density beyond the Cβ of Ala146. Analysis of the UniProt[37] metadata associated with PDB 6ALD indicated this amino acid should be a lysine. Thus, Lys146 was corrected in each chain before refinement. These initial models were then refined into the EM density using the imposed symmetry and estimated resolution while adjusting the Rosetta weighting and scoring functions according to the estimated map resolution.

Each of the 100 Rosetta-generated models[38] were ranked based on the number of Ramachandran outliers (%, lower better), geometry violations (%, lower better), Rosetta aggregate score (value, lower better), and MolProbity clashscore[39] (value, lower better). The top 10 structures that scored the best across all categories were selected for real-space refinement using the Phenix refinement package[40]. Model–model agreement statistics were calculated using a previously described approach[36].

Data availability. The coordinates for the 20S proteasome stage position, 20S proteasome image shift and aldolase structures have been deposited in the Protein Data Bank (PDB) under accession codes 5VY3, 5VY4 and 5VY5, respectively. The corresponding EM density maps have been deposited to the Electron Microscopy Data Bank under accessions EMD-8741, EMD-8742 and EMD-8743, respectively. The data sets that support the findings of this study are also available from the corresponding author upon request. A **Life Sciences Reporting Summary** is available.

25. Herzik, M.A. Jr., Wu, M. & Lander, G.C. Setting up the Talos Arctica electron microscope and Gatan K2 direct detector for high-resolution cryogenic single-particle data acquisition. *Protocol Exchange* http://dx.doi.org/10.1038/protex.2017.108 (2017).
26. Dubochet, J. *et al. Q. Rev. Biophys.* **21**, 129–228 (1988).
27. Grassucci, R.A., Taylor, D. & Frank, J. *Nat. Protoc.* **3**, 330–339 (2008).
28. Asadabad, M.A. & Eskandari, M.J. in Modern Electron Microscopy in Physical and Life Sciences. (eds. M. Janacek & R. Kral) (InTech, 2016).
29. Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S. & Carragher, B. *J. Struct. Biol.* **166**, 205–213 (2009).
30. Ogura, T., Iwasaki, K. & Sato, C. *J. Struct. Biol.* **143**, 185–200 (2003).
31. Roseman, A.M. *J. Struct. Biol.* **145**, 91–99 (2004).
32. Scheres, S.H. & Chen, S. *Nat. Methods* **9**, 853–854 (2012).
33. Chen, S. *et al. Ultramicroscopy* **135**, 24–35 (2013).
34. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. *Nat. Methods* **14**, 290–296 (2017).
35. Cardone, G., Heymann, J.B. & Steven, A.C. *J. Struct. Biol.* **184**, 226–236 (2013).
36. Herzik, M.A., Fraser, J.S. & Lander, G.C. A multi-model approach to assessing local and global cryo-EM map quality. Preprint at *bioRxiv* https://doi.org/10.1101/128561 (2017).
37. Apweiler, R. *et al. Nucleic Acids Res.* **32**, D115–D119 (2004).
38. Wang, R.Y. *et al. eLife* **5**, e17219 (2016).
39. Chen, V.B. *et al. Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
40. Adams, P.D. *et al. Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

Corresponding Author: Gabriel C. Lander

Date: Aug 24, 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. Sample size

   Describe how sample size was determined.

   N/A

2. Data exclusions

   Describe any data exclusions.

   N/A

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   N/A

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   N/A

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   N/A

   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☒	☐	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☒	☐	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☒	☐	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☒	☐	The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted
☒	☐	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

   Describe the software used to analyze the data in this study.

   Appion, FindEM, CTFFind4, RELION, BSOFT, GCTF, Chimera, EMRinger, MotionCorr2

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

> N/A

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

> No eukaryotic cell lines used

b. Describe the method of cell line authentication used.

> No eukaryotic cell lines used

c. Report whether the cell lines were tested for mycoplasma contamination.

> No eukaryotic cell lines used

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

> No commonly misidentified cell lines were used

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

> no animals used

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> did not involve humans