Published in final edited form as:

J Struct Biol. 2015 November ; 192(2): 196–203. doi:10.1016/j.jsb.2015.09.006.

The Influence of Frame Alignment with Dose Compensation on the Quality of Single Particle Reconstructions

John M. Spear^a, Alex J. Noble^b, Qing Xie^c, Duncan R. Sousa^d, Michael S. Chapman^c, and Scott M. Stagg^{a,e,*}

^a Institute of Molecular Biophysics, 91 Chieftan Way, Florida State University, Tallahassee, FL 32306-4380.

^b Department of Physics, 77 Chieftan Way, Florida State University, Tallahassee, FL 32306-4350.

^c Department of Biochemistry & Molecular Biology, School of Medicine, Oregon Health & Science University, Portland, OR 97239-3098.

^d Department of Biological Science, Florida State University, 319 Stadium Drive, Tallahassee, FL 32306.

^e Departments of Chemistry and Biochemistry, 95 Chieftain Way, Florida State University, Tallahassee, FL 32306-4390.

Abstract

As direct electron detection devices in cryo-electron microscopy become ubiquitous, the field is now ripe for new developments in image analysis techniques that take advantage of their increased SNR coupled with their high-throughput frame collection abilities. In approaching atomic resolution of native-like biomolecules, the accurate extraction of structural locations and orientations of side-chains from frames depends not only on the electron dose that a sample receives but also on the ability to accurately estimate the CTF. Here we use a new 2.8 Å resolution structure of a recombinant gene therapy virus, AAV-DJ with Arixtra, imaged on an FEI Titan Krios with a DE-20 Direct Electron detector to probe new metrics including relative side-chain density and ResLog analysis for optimizing the compensation of electron beam damage and to characterize the factors that are limiting the resolution of the reconstruction. The influence of dose compensation on the accuracy of CTF estimation and particle classifiability are also presented. We show that rigorous dose compensation allows for better particle classifiability and greater recovery of structural information from negatively charged, electron-sensitive side-chains, resulting in a more accurate macromolecular model.

*Corresponding author. sstagg@fsu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Accession numbers

The AAV-DJ-Arixtra map was deposited to the EMDB under accession number xxxx.

Keywords

Direct electron detector; Radiation damage; Beam-induced motion; Dose compensation; Atomic resolution; Cryo-electron microscopy

1. Introduction

The introduction of direct electron detection devices (DDD) has precipitated a revolution in 3DEM. The combination of their high detective quantum efficiency (DQE) with their high frame rates has allowed for the determination of several single particle reconstructions to better than 3 Å resolution (Bartesaghi et al., 2015; Fischer et al., 2015; Grant and Grigorieff, 2015; Jiang et al., 2015; Melody G Campbell, 2015), and many reconstructions with high enough resolution to model atomic structures with high confidence (DiMaio et al., 2015). The advantage of the high DQE of DDDs is that they outperform film and CCD cameras in terms of contrast and signal-to-noise ratios at all spatial frequencies (McMullan et al., 2014), and this in-turn allows for better alignment and classification. The advantages of the high frame rate are that it allows for the compensation of beam induced motion and for compensation of specimen damage due to the ionizing radiation of the electron beam (Campbell et al., 2012; Li et al., 2013; Scheres, 2014; Wang et al., 2014).

In spite of the ability to compensate for radiation damage, a common observation among the high-resolution single particle reconstructions that have been determined thus far is that the density values for acidic residues are low relative to the other residues (Bartesaghi et al., 2014; Fromm et al., 2015; Jiang et al., 2015). It has been suggested that acidic residues are more sensitive to radiation damage than other residues (Bartesaghi et al., 2014). Another possibility is that their electron scattering factors are lower than the other residues (Mitsuoka et al., 1999). Several methods have been used to mitigate the effects of beam damage. One approach is to align and classify particles on the lower-noise high-dose images then reconstruct based on the early frames that have suffered less beam damage (Liao et al., 2013). Another approach is fitting relative B-factors to reconstructions from individual frames before combining them into a map containing all frames (Scheres, 2014). Yet another is to combine the frames into one image after applying a low-pass filter to each that is proportional to the cumulative electron dose the specimen has experienced, then summing all the frames (Grant and Grigorieff, 2015; Wang et al., 2014). In the current work, the filtering resolution is set by calibration to the dose-dependent attenuation of catalase electron diffraction (Wang et al., 2014), while it is also possible to similarly calibrate against the resolutions of prior single particle reconstructions (Grant and Grigorieff, 2015). Here, we test the effects of the advances in frame alignment with dose compensation on other aspects of single particle reconstruction, including CTF estimation, particle classification (Euler angle assignment), and examine the impact on resolution and side-chain densities.

2. Materials and methods

2.1. Expression and purification

Virus-like particles of AAV-DJ were expressed in insect cells from a baculovirus construct as previously described (Lerch et al., 2012). Empty capsids were purified as before using three rounds of CsCl density gradient ultracentrifugation, followed by heparin affinity chromatography, eluting with a NaCl gradient. Capsids were then diluted in 50mM Hepes, 25mM MgCl₂, 25mM NaCl, pH= 7.4.

2.2. Cryo-EM sample preparation

3 μ l of 0.6 mg/ml AAV-DJ was applied to Quantifoil® (Jena, Germany) R2/2 200 mesh copper grids that were rendered hydrophilic by glow discharge in 75/25 percent Ar/O. The grid was hand blotted and further incubated with 3 μ l of Arixtra (pharmaceutical preparation) for 15 seconds at a concentration of 5.7 mM. Arixtra was dissolved in ultra pure water. After addition of Arixtra, the grids were vitrified in liquid nitrogen-cooled ethane using an FEI Vitrobot (FEI, Hillsboro OR) with the following parameters: blot force = 1, blot time = 3 seconds, total blots = 1, humidity = 100 percent, temperature = 4°C. Vitrified samples were stored in liquid nitrogen until further use.

Kv β 2 was expressed and purified as previously published with no modifications (Weng et al., 2006). For cryo-EM sample preparation, the final sample was dialyzed overnight at 4°C against glycerol-free buffer (20 mM Tris pH 8.0, 300 mM KCl, 1 mM beta-mercaptoethanol) in a 5000 molecular weight cut off (MWCO) dialysis cassette (Thermo, USA). 3 μ l of Kv β 2 at 0.9 mg/ml was applied to Quantifoil® R2/2 200 mesh grids that were pre-treated in the same fashion as mentioned above. Grids were vitrified in liquid nitrogen-cooled ethane using an FEI Vitrobot (FEI, Hillsboro OR) with the following parameters: blot force = 2, blot time = 2 seconds, total blots = 1, humidity = 100 percent, temperature = 4°C. Vitrified samples were stored in liquid nitrogen until further use.

2.3. Image acquisition

Two AAV-DJ Arixtra datasets were obtained from the same grid with differing defocus ranges. A final exposure magnification of 29,000 was maintained for all images. All images experienced a similar total dose (66 e⁻/Å²) and were acquired using an FEI Titan Krios (FEI, Hillsboro OR) using Legikon (Suloway et al., 2005) for image collection. The images were recorded on a DE-20 direct electron detector (Direct Electron, San Diego CA) with the dose fractionated across 45 frames leading to a dose rate of 1.5 e⁻/Å² per frame. The first dataset was collected at a defocus range between 1.5 - 3.0 μ m and the second set at 0.75 - 1.75 μ m. Defocus estimates were continually made throughout the image acquisitions to ensure that the defocus range did not drift outside of the defined parameters using the automated contrast transfer function (CTF) estimator ACE (Mallick et al., 2005) and CTFFIND3 (Mindell and Grigorieff, 2003). The two datasets yielded a total of 1428 images were acquired from a total of 40 hours of data collection. Of those images 377 were discarded due to contamination or mistargeting of the exposure.

2.4. Dose compensation and frame alignment

Frame alignment and dose compensation was performed using the DE_process_frames-2.5.1 software that comes with Direct Electron cameras. Briefly, the algorithm makes rolling averages for sets of frames, aligning them to the sum of all frames, and iterating the alignment a user-defined number of iterations. Dose compensation is achieved by applying a low-pass filter to individual frames where the cutoff frequency for a given frame is dependent on the cumulative dose for that frame. The specific cutoff frequencies were calibrated from the fading of spots of images of catalase crystals (Direct Electron personal communication) determined similarly to that described in Baker *et al.* (Baker et al., 2010). A scaling factor can optionally be applied to make the filtering more or less aggressive, and this was used to generate stacks with four different dose compensation schemes. The individual per frame low-pass filtering schemes for each stack are given in Fig. S1. The DE_process_frames software was wrapped into the Appion pipeline so that if a user has access to the DE_process_frames software, frame data can be processed in a high-throughput manner. For the datasets here, full frames were first aligned and dose compensated, then after particles were picked, individual particles were realigned and dose compensated.

2.5. Image processing

Of the two AAV-DJ Arixtra datasets, the first image set yielded 503 full-frame image exposures, producing 65,978 particles. The second image set totaled 548 full-frame image exposures and produced 54,188 particles. Particle picking was completed semi-automatically using the template picker FindEM (Roseman, 2003) within Appion (Lander et al., 2009). The template used for picking was a rotational average of the unaligned average of a stack of manually picked particles. Particles over carbon were manually deselected. Initial Eulers were generated using EMAN1 (Ludtke et al., 1999) by refining for 2 iterations with an angular increment of 1° starting from an initial model of AAV that had been low-pass filtered to 20 Å. The Euler angles were then further refined with 9 iterations of FREALIGN (Grigorieff, 2007) refinement. An inverse B factor was applied to the final FREALIGN maps using EMBFactor (Fernández et al., 2008) after refinement was complete.

2.6. Atomic modeling and fitting

The effective EM magnification, an envelope correction, and the resolution of a low-pass filter were least-squares refined using RSRef (Chapman et al., 2013) by optimizing the agreement between the density of the EM reconstruction and the atomic model. The starting model was derived from a prior AAV-DJ structure at 4.5 Å resolution (Lerch et al., 2012). It was manually adjusted using Coot (Emsley and Cowtan, 2004) and then refined by simulated annealing torsion angle optimization against the AAV-DJ-Arixtra complex map. This was done using a real-space objective function calculated by RSRef embedded in CNS v1.3 (Brünger et al., 1998).

Arixtra was refined against a difference map between the Arixtra complex and native. Prior to difference map calculation, the reconstructions were scaled in reciprocal space using EMAN1 to limit the impact of differences in power spectra. They were also scaled in real-space, putting each on an absolute scale with reference to the atomic model using RSRef

(Chapman et al., 2013). The model for the Arixtra ligand was refined against this difference map and then fixed for a final round, refining the virus structure against the reconstruction of the complex.

2.7. ResLog analysis

ResLog analysis was performed as described in Stagg, 2013 (Stagg et al., 2013). In brief, a number of increasingly small randomly sampled subsets of particles were selected from the full particle stack and reconstructed using the orientations determined for the full particle stack. The resolution of the reconstructions from the individual subsets was then plotted against the *log* of the number of particles contributing to the reconstruction.

2.8. Voxel density analysis and MSMS

The 'Values at Atom Positions' option in the 'Volume Data' tool within UCSF Chimera (Pettersen et al., 2004) was used in conjunction with the AAV-DJ subunit atomic model and the reported EM density maps to obtain relative, interpolated voxel density values for each atom. A Python script was written to sum the voxel densities of side-chain atoms (C β and beyond) for each residue, normalized to the number of side-chain atoms, relative to their residue's backbone atoms and relative to the average backbone atoms of the entire subunit.

Michel Sanner's Molecular Surface (MSMS) (Sanner et al., 1996) was used to model the solvent-accessible surface area (SASA) of the AAV-DJ-Arixtra subunit atomic model. First the pdb model of an individual subunit was fit into the EM density. Next, using the measure, symmetry, and fit commands in UCSF Chimera, an atomic model of the whole virus was generated. A pdb of one AAV-DJ-Arixtra subunit with its nearest neighbors was supplied to MSMS where a SASA was generated for a simulated water molecule of radius 1.5 Å. The SASA was then mapped onto the atomic model and per-residue voxel density and SASA analysis was performed for each EM density.

Simulation of the maximum number of usable particles—The maximum number of usable particles per micrograph for a 100 Å diameter particle was estimated as described in Shrum *et al.* (Shrum et al., 2012). Briefly, positions of particles on a micrograph with the same area as the DE-20 were simulated assuming random positions with an increasing number of particles per micrograph and a pixel size of 1.215 Å. Particles were eliminated if they came within 5% of the particle diameter of each other. To allow for defocus-dependent signal delocalization, particles were also eliminated if they came within 75% of the particle diameter of the image border.

3. Results

We set out to characterize practical factors that limit the resolution of single particle reconstructions on data collected on a DE-20 DDD. To this end, we collected datasets on adeno-associated virus serotype DJ (AAV-DJ) that we had previously determined to 4.5 Å resolution, and AAV-DJ in association with Arixtra which is an analog of cell surface glycosaminoglycans to which the virus attaches. These datasets were probed for factors such

as the number of particles contributing to the reconstruction, dose compensation, and CTF estimation - factors that might be limiting the resolution of 3D reconstructions.

3.1. ResLog analysis

Recently several cryo-EM structures have been published where relatively high resolution was achieved with thousands or tens of thousands of particles (Grant and Grigorieff, 2015; Lu et al., 2014; Voorhees et al., 2014). However, it is unclear with those studies if higher resolution reconstructions could be achieved simply by adding more data. In other words, is the number of particles limiting the resolution of the reconstructions. For our AAV-DJ-Arixtra samples, we were able to achieve 4.5 Å resolution with only 290 particles (17,400 asymmetric units (ASU), Fig. 1A,D). At that resolution, bulky side-chains are clearly visible, and the backbone could be traced with the benefit of homology models to resolve ambiguities in the topology of the protein backbone (Fig. 1D, arrow). However, in order for an *ab initio* model to be built, one must be able to unambiguously trace the chain and distinguish similar residues. A rule of thumb in crystallography is that resolutions better than ~3.5 Å are needed to model backbone topology without ambiguity. We and others have previously established that there is a linear relationship between spatial frequency (1/resolution) and the logarithm of the number of particles that are contributing to the corresponding EM map, a so-called ResLog plot (LeBarron et al., 2008; Liao and Frank, 2010; Stagg et al., 2013). This trend was present for all datasets that we have examined (Stagg et al., 2013) and is still true with the new data collected on the DE-20 (Fig. 1B). The linear trend continued no matter how many particles were included in the map, and we were able to resolve features at 2.8 Å resolution by using 107,454 particles (6,447,240 ASU, Fig. 1C).

3.2. Frame weighting and dose compensation

It has been demonstrated that DDD frames can be aligned, dose compensated, and summed so that the individual frames are combined into one image where the high resolution features are not washed out with higher and higher dose (Grant and Grigorieff, 2015; Wang et al., 2014). In this approach, which is implemented in the DE_process_frames-2.5.1 software, successive frames are low-pass filtered to a resolution that is a function of the cumulative dose such that when the frames are summed, the high frequency terms are on a more even weight with the low frequency terms (Fig. S1). One of the advantages of this approach is that all of the frames are used at all stages of the reconstruction. This approach greatly enhances the contrast of the particles (Fig. 2A). This is especially true for smaller particles, such as from data we collected of Kvβ2, a 360 kDa octomer that are practically invisible at low defocus without dose compensation but are easily distinguished after compensation (Fig. S2). Alignment and compensation can have two different effects on 3D reconstructions: 1) it can affect particle classification during refinement and 2) it can affect the relative weighting of low-resolution terms and high-resolution terms in the reconstructions. We tested the effect of dose compensation on classification by independently refining a stack of AAV-DJ particles that were frame aligned but not compensated and a stack of the same particles that were both aligned and compensated. The refinements were then compared using ResLog plots. We previously showed that the y-intercept of a ResLog plot was an indicator of particle classifiability and the slope was an

indicator of data quality (Stagg et al., 2013). The compensated stack had both a higher intercept and higher slope (Fig. 2D) indicating that the dose compensation improved both particle classifiability and the preservation of high-resolution information.

The impact of dose compensation on resolution was tested independently of refinement by reconstructing stacks of particles with different degrees of weighting of the dose compensation and with fixed Euler angles and shifts. Four different weighting schemes were used. The most aggressive weighting scheme started low-pass filtering with frames that had a cumulative dose of $8.7 \text{ e}^-/\text{\AA}^2$, and the last scheme applied no frame weighting (Fig. S1). The schemes were named according to the dose compensation multiplier that was passed to the DE frame alignment and compensation script (i.e. dose0.5, dose1, dose2, and dose4, Fig. S1). All schemes used the same per frame shift to compensate for any specimen motion. The four resulting stacks were reconstructed without refining the Euler angles and shifts that had been calculated for the dose1 data, and the reconstructions were compared using their respective FSC plots (Fig. 3A). This resulted in a trend where the uncompensated data had the lowest FSC correlations across all frequencies followed by dose2 and dose1. However, the most aggressive compensation, dose0.5, resulted in a reconstruction that had higher FSC than the other schemes at frequencies beyond $0.31 \text{ \AA}^{-1}$ ($3.2 \text{ \AA}$). A similar trend was observed when the Eulers and shifts for dose0.5 and dose4 were refined. Dose4 had the worst FSC across all frequencies, dose1 had the highest FSC for frequencies lower than $0.31 \text{ \AA}^{-1}$, and dose0.5 had the highest FSC for frequencies greater than $0.31 \text{ \AA}^{-1}$ (Fig. 3B). Moreover, dose4 had worse resolution than when it was reconstructed using the Eulers from dose1, which agrees with our earlier result that the uncompensated data are less classifiable than the compensated data.

Given the observation that dose compensation made such a big difference for single particle refinement, we hypothesized that it could improve CTF estimation. We tested the CTF estimation software ACE (Mallick et al., 2005) and CTFFIND (Mindell and Grigorieff, 2003), to establish if the use of compensated images in defocus estimation yielded improved resolution. Defocus was separately estimated for compensated and uncompensated images using ACE and CTFFIND, then reconstruction were made using the Eulers angles and shifts from the dose1 reconstruction. In contrast to the recent CTF challenge results (Marabini et al., 2015), ACE gave higher FSC values perhaps because astigmatism was practically undetectable in our images. We suggest that the nonastigmatic estimates from ACE likely gave better results because CTFFIND always estimated some amount of astigmatism. When the dose compensated images were used for defocus estimation, the resolution for the reconstruction using CTFFIND was substantially worse while the ACE reconstruction was slightly better. This result can be explained by differences in the algorithms for ACE and CTFFIND and the amount of carbon film impinging on our images, which for most of the images in the AAV-DJ-Arixtra data takes up ~30% of the image. Since the carbon film is not damaged by the beam, dose compensation limits the amount of signal that would contribute to the Thon rings in a Fourier transform and thus makes CTF estimation worse for CTFFIND which uses the full power spectrum when estimating CTF parameters. ACE, on the other hand, estimates CTF parameters from a defined resolution band that is based on the recorded defocus of the microscope during data collection. The resolution band in our case incorporated only the first few Thon rings in an image transform, which are in the least

noisy part of the spectrum. It is possible that ACE performs better with the compensated images in part because the lower frequency terms on which ACE is estimating CTF are relatively upweighted as a result of dose compensation. Finally, a new algorithm in Appion that chooses between different defocus estimates based on correlation with observed data (Sheth et al., 2015) produced a reconstruction with resolution intermediate between those using ACE and CTFFIND estimates (Fig. 3C, blue). This is consistent with the observation that the algorithm used values from ACE and CTFFIND about evenly.

3.3. Voxel density at negatively charged residues

As with other high-resolution EM reconstructions, we observed that acidic side-chains had lower relative density values than other side-chains (Bartesaghi et al., 2014; Fromm et al., 2015; Jiang et al., 2015). It has been proposed that negatively charged groups are more sensitive to beam damage than other moieties (Bartesaghi et al., 2014). To test this hypothesis, we evaluated the relative density of acidic residues by comparing the four reconstructions that were generated from differently dose compensated particle frames (dose0.5, dose1, dose2, and dose4 described earlier) with fixed Euler angles and shifts. We calculated the density of each residue's side-chain relative to its backbone carbons and compared this against the solvent accessible surface area (see Methods). As expected, in general, residues that were highly solvent accessible had lower relative density than residues that were less accessible (Fig. 4A) due to their greater degrees of freedom. When the relative densities for all residues for the four differently compensated volumes were compared, a trend was noticed where the average density value for the reconstruction from the dose0.5 particles was significantly higher than the dose4 map ($p=0.003$, two tailed Students t-test) (Fig. 4B). The acidic side-chains had substantially lower density values than the rest of the residues (Fig. 4B). However, increasingly stringent dose compensation resulted in a trend where the average density values for acidic residues increased more rapidly than the rest of the side-chains (Fig. 4B). This shows that indeed acidic residues are damaged at a faster rate than other residues and some of that damage can be recovered by stringent dose compensation. The same trend was present for the acidic ligand, Arixtra, that was present in this sample. The Arixtra density was improved in a similar manner to the acidic side-chains with increasingly stringent dose compensation. It should be noted, however, that the improvement in the Arixtra density could be the result of the greater freedom of a non-covalent adduct to move upon radiation damage in addition to the increased sensitivity of the acidic moieties in the Arixtra molecule.

Interestingly, in spite of the improvement in the quality of side-chain densities with dose compensation, the acidic side-chains could still be confidently modeled in each of the differently compensated volumes. We examined, in detail, the densities for $\sim 1/3$ of the acidic side-chains. Maps were put onto an absolute scale with reference to the (entire) atomic model. The dose0.5, dose1 and dose4 maps were then compared in Coot by matching their contours and recording the corresponding sigma levels. There was no significant difference in the densities when compared this way. This indicates that the density is still present for the acidic residues in all of the maps, but dose compensation limits the selective build-up of other features at higher doses, leaving them on a more even scale (Fig. 4C,D).

3.4. Characterization of the DE-20 resolution and throughput

In addition to reconstructing volumes, we acquired data to characterize the performance of our DE-20 detector. We collected a DQE curve using the beam stop and the FindDQE software by Ruskin *et al* (Ruskin *et al.*, 2013). This showed that the DQE was relatively flat across lower frequencies and starts falling off around 75% of Nyquist (Fig. 5). A plot of the estimated spectral signal to noise ratio was calculated using the EMAN2 e2ctf program (Tang *et al.*, 2007) on an image of carbon at the same magnification and dose used for our AAV data collection and with 1.0 μm defocus. This showed that CTF ripples could still be observed at 2.8 $\text{\AA}$ (85% Nyquist) and beyond, but the CTF was not sampled well beyond this resolution with the current pixel size (Fig. 5). The poor sampling of the CTF at high frequencies is likely a contributing factor for what is limiting the resolution of the current reconstruction to 2.8 $\text{\AA}$. It is likely that a higher resolution reconstruction could be achieved by collecting data at a higher magnification or interpolating the images to have a smaller pixel size.

One of the advantages of integrating direct detectors over counting detectors is that they have a higher throughput due to their shorter image acquisition time. For the AAV-DJ Arixtra data, we collected 1429 exposures over the course of 40 hours of data collection on a single grid. This allowed us to image 725 holes spread over 21 squares. Each $65 \text{ e}^-/\text{\AA}^2$ exposure was associated with 45 frames that were aligned and compensated using the frame alignment and compensation script provided by Direct Electron. This script was wrapped within the Appion processing pipeline for automated frame processing, and a new utility was added so that frame stacks can be processed by submitting multiple single processor jobs to a high-performance computing cluster. This resulted in a speedup of processing so that frames could be processed almost at the same rate that data were acquired.

4. Discussion

Here we have evaluated the effects of frame alignment with dose compensation, and summation on various aspects of single particle reconstructions including CTF estimation, particle classification, resolution, and the relative densities of acidic residues and other side-chains. We found that frame alignment and dose compensation have positive effects on all aspects of 3D reconstruction except for CTF estimation, where its influence is dependent on the algorithm used for CTF estimation. One of the biggest effects of dose compensation was a noticeable increase in contrast (Fig. 2B,C), and this resulted in a large improvement in particle classifiability. ResLog analysis showed that reconstructions that were refined from dose compensated particles had a substantially higher slope and y-intercept, which we have previously shown correlates to better-classified particles (Stagg *et al.*, 2013) (Fig. 2D). This conclusion was validated by showing that a reconstruction of uncompensated particles reconstructed to 3.3 $\text{\AA}$ resolution, but the resolution improved to 3.2 $\text{\AA}$ when the uncompensated particles were assigned the Eulers and shifts from a refinement from compensated particles (Fig. 3A,B). The resolution was further improved to 2.8 $\text{\AA}$ when dose compensated particles were used for the both refinement and reconstruction. This demonstrates that in addition to their better contrast, an advantage of the dose compensated particles is that they retain their high-resolution features. One caveat to this, however, is that

with fixed Euler angles, the most aggressive compensation, dose0.5, resulted in a reconstruction that had higher FSC than the other schemes at frequencies higher than 0.31 Å⁻¹ (3.2 Å), but lower FSC at lower frequencies than reconstructions from less aggressively compensated particles. It is possible that the worse FSC at lower frequencies could mean that model building is easier with the less aggressively compensated particles.

It is noted that a newer version of the DE frame alignment script is available that uses the dose compensation scheme described in Grant and Grigorieff (Grant and Grigorieff, 2015), but this was not available for the work described here. We anticipate that it would give results that are similar to the dose0.5 compensation scheme used here.

It has been shown that acidic residues in high-resolution single particle reconstructions have weaker relative densities compared to the other side-chains (Bartesaghi et al., 2014; Fromm et al., 2015; Jiang et al., 2015). A benefit of the dose compensation scheme used here is that some of the lost density can be recovered by using rigorous compensation (Fig. 4). We showed that dose compensation has a dual benefit of 1) increasing the density for all side-chains relative to the backbone, and 2) increasing the relative densities of acidic side-chains more than the others, thus putting them on a more even scale (Fig. 4). Thus, aggressive dose compensation can have the benefit of improving model building for all side-chains and aid in interpretation of structures where precisely determined side-chain rotamers are required. In spite of the dose compensation, the acidic residues had lower density values relative to the rest of the side chains even in the most aggressively dose compensated reconstruction. It is unclear whether the acidic residues can be resolved even better by collecting with a lower dose rate and more aggressive dose compensation or whether the remaining differences simply result from the lower electron scattering of negatively charged atoms (Mitsuoka et al., 1999).

The prospects for high-resolution structure determination of other samples depends strongly on the resolution that is required to answer the biological question of interest. If one wanted to reach the same resolution as the AAV-DJ-Arixtra reconstruction, the following conditions would be required. We estimate that for a camera that covers the same area as the DE-20, for a 100 Å diameter particle with a pixel size of 1.215 Å, the maximum usable number of particles in any given micrograph is ~300 (Fig. 6). If the particle were asymmetrical, then to get the ~6,000,000 asymmetric units that we had in our highest-resolution AAV-DJ-Arixtra reconstruction, 20,000 images would need to be acquired requiring a whopping 34TB of disk space for frames. Using our data acquisition rate, this would require 571 hours or 24 days of data collection. A caveat to that estimate is that the resolution of our AAV-DJ-Arixtra map was approaching the Nyquist limit, which meant that we were at the tail end of the detection efficiency of the camera, and resolution is degraded by poor sampling of the CTF (Fig. 5). It is likely that higher resolution could be achieved more efficiently with data collected at higher magnification (or by resampling the current images with a smaller pixel size). It has been shown that high-quality atomic models can be made from reconstructions with better than 4.5 Å resolution (DiMaio et al., 2015). We were able to determine a 4.5 Å map of AAV-DJ-Arixtra with only 17,400 ASU. For a 100 Å diameter particle, this would require only 1 hour of data collection. It is clear that one must carefully consider what resolution is required to make meaningful biological interpretations, and balance that against

the cost of microscope time and the ability to store and process data when planning for a single particle reconstruction project.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Benjamin Bammes and Michael Spillman for helpful discussions about processing DE-20 frames. We thank Omar Davulcu for his help in performing the statistical analysis of the density values in Coot. This work was supported by R01GM086892 (SMS) and R01GM66875 (MSC).

References

- Baker LA, Smith EA, Bueler SA, Rubinstein JL. The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. *J Struct Biol.* 2010; 169:431–437. [PubMed: 19958834]
- Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S. Structure of -galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proc Natl Acad Sci U S A.* 2014; 111:11709–11714. [PubMed: 25071206]
- Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, et al. Electron microscopy. 2.2 Å resolution cryo-EM structure of -galactosidase in complex with a cell-permeant inhibitor. *Science.* 2015; 348:1147–1151. [PubMed: 25953817]
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr.* 1998; 54:905–921. [PubMed: 9757107]
- Campbell MG, Cheng A, Brilot AF, Moeller A, Lyumkis D, et al. Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure.* 2012; 20:1823–1828. [PubMed: 23022349]
- Chapman MS, Trzyna A, Chapman BK. Atomic modeling of cryo-electron microscopy reconstructions--joint refinement of model and imaging parameters. *J Struct Biol.* 2013; 182:10–21. [PubMed: 23376441]
- DiMaio F, Song Y, Li X, Brunner MJ, Xu C, et al. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat Methods.* 2015; 12:361–365. [PubMed: 25707030]
- Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
- Fernández JJ, Luque D, Castón JR, Carrascosa JL. Sharpening high resolution information in single particle electron cryomicroscopy. *J Struct Biol.* 2008; 164:170–175. [PubMed: 18614378]
- Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, et al. Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature.* 2015; 520:567–570. [PubMed: 25707802]
- Fromm SA, Bharat TAM, Jakobi AJ, Hagen WJH, Sachse C. Seeing tobacco mosaic virus through direct electron detectors. *J Struct Biol.* 2015; 189:87–97. [PubMed: 25528571]
- Grant T, Grigorieff N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *Elife.* 2015; 4
- Jiang J, Pentelute BL, Collier RJ, Zhou ZH. Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature.* 2015; 521:545–549. [PubMed: 25778700]
- Lander GC, Stagg SM, Voss NR, Cheng A, Fellmann D, et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol.* 2009; 166:95–102. [PubMed: 19263523]

- LeBarron J, Grassucci RA, Shaikh TR, Baxter WT, Sengupta J, Frank J. Exploration of parameters in cryo-EM leading to an improved density map of the E. coli ribosome. *J Struct Biol.* 2008; 164:24–32. [PubMed: 18606549]
- Lerch T, O'Donnell J, Meyer N, Xie Q, Taylor K, et al. Structure of AAV-DJ, a Retargeted Gene Therapy Vector: Cryo-Electron Microscopy at 4.5 Å Resolution. *Structure.* 2012; 20:1310–1320. [PubMed: 22727812]
- Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods.* 2013; 10:584–590. [PubMed: 23644547]
- Liao HY, Frank J. Definition and estimation of resolution in single-particle reconstructions. *Structure.* 2010; 18:768–775. [PubMed: 20637413]
- Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature.* 2013; 504:107–112. [PubMed: 24305160]
- Lu P, Bai X-C, Ma D, Xie T, Yan C, et al. Three-dimensional structure of human α -secretase. *Nature.* 2014; 512:166–170. [PubMed: 25043039]
- Ludtke SJ, Baldwin PR, Chiu W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol.* 1999; 128:82–97. [PubMed: 10600563]
- Mallick SP, Carragher B, Potter CS, Kriegman DJ. ACE: automated CTF estimation. *Ultramicroscopy.* 2005; 104:8–29. [PubMed: 15935913]
- McMullan G, Faruqi AR, Clare D, Henderson R. Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy.* 2014; 147:156–163. [PubMed: 25194828]
- Melody G Campbell DVACCSPBC. 2.8 Å resolution reconstruction of the Thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. *Elife.* 2015; 4
- Mindell JA, Grigorieff N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol.* 2003; 142:334–347. [PubMed: 12781660]
- Mitsuoka K, Hirai T, Murata K, Miyazawa A, Kidera A, et al. The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution. *Journal of Molecular Biology.* 1999; 286:861–882. [PubMed: 10024456]
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004; 25:1605–1612. [PubMed: 15264254]
- Roseman AM. Particle finding in electron micrographs using a fast local correlation algorithm. *Ultramicroscopy.* 2003; 94:225–236. [PubMed: 12524193]
- Ruskin RS, Yu Z, Grigorieff N. Quantitative characterization of electron detectors for transmission electron microscopy. *J Struct Biol.* 2013; 184:385–393. [PubMed: 24189638]
- Sanner MF, Olson AJ, Spehner JC. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers.* 1996; 38:305–320. [PubMed: 8906967]
- Scheres SH. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife.* 2014; 3:e03665. [PubMed: 25122622]
- Sheth LK, Piotrowski AL, Voss NR. Visualization and Quality Assessment of the Contrast Transfer Function Estimation. *J Struct Biol.* 2015
- Shrum DC, Woodruff BW, Stagg SM. Creating an infrastructure for high-throughput high-resolution cryogenic electron microscopy. *J Struct Biol.* 2012; 180:254–258. [PubMed: 22842049]
- Stagg SM, Noble AJ, Spilman M, Chapman MS. ResLog plots as an empirical metric of the quality of cryo-EM reconstructions. *J Struct Biol.* 2013
- Suloway C, Pulokas J, Fellmann D, Cheng A, Guerra F, et al. Automated molecular microscopy: the new Leginon system. *J Struct Biol.* 2005; 151:41–60. [PubMed: 15890530]
- Tang G, Peng L, Baldwin PR, Mann DS, Jiang W, et al. EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol.* 2007; 157:38–46. [PubMed: 16859925]
- Voorhees R, Fernández I, Scheres SW, Hegde R. Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution. *Cell.* 2014; 157:1632–1643. [PubMed: 24930395]

- Wang Z, Hryc CF, Bammes B, Afonine PV, Jakana J, et al. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nat Commun.* 2014; 5:4808. [PubMed: 25185801]
- Weng J, Cao Y, Moss N, Zhou M. Modulation of voltage-dependent Shaker family potassium channels by an aldo-keto reductase. *J Biol Chem.* 2006; 281:15194–15200. [PubMed: 16569641]

Fig. 1.

Comparison of volumes reconstructed with differing numbers of particles. A) FSC curves of reconstructions with increasing numbers of particles. B) ResLog plot showing resolution ($FSC_{0.143}$) as a function of numbers of particles contributing to the map. C). Density from a volume reconstructed with 107,454 particles. D) Density from a volume that was independently refined and reconstructed from 290 particles. The arrow indicates a region where there is an ambiguity in tracing the protein backbone.

Fig. 2. Effect of dose compensation on particle contrast. A) Aligned/uncompensated particle with 15e⁻/Å² dose. B) Aligned/uncompensated particle with 60e⁻/Å² dose. C) Aligned/compensated particle with 60e⁻/Å² dose. D) ResLog plots comparing refinements with aligned/uncompensated particles (blue) with aligned/compensated particles (red).

Fig. 3.

Comparison of reconstructions with differently dose compensated particles. A) FSC curves of reconstructions from aggressively compensated particles (dose0.5) to uncompensated particles (dose4) where the particles from the different sets had fixed Euler angles and shifts. B) FSC curves of reconstructions of differently dose compensated particles where the Euler angles and shifts were independently refined. C) FSC curves comparing reconstructions where Euler angles and shifts were fixed but CTF estimates were estimated in using ACE and CTFFIND with compensated and uncompensated images. The dotted line indicates FSC_{0.143} for all panels.

Fig. 4.

Analysis of the affects of dose compensation on relative side-chain density and side-chain SASA. (A) Scatterplot of the relative side-chain density of each AAV-DJ residue vs. side-chain SASA and subplots of acidic residues for dose1 (see Fig. S3 for analyses of reconstructions from all dose compensation schemes). (B) The mean and 95% confidence interval (error bars) of all residues but acidic residues (blue), and acidic residues (yellow) across all reported doses. Average density values of Arixtra atoms relative to the average residue backbone density is shown in gray. (C) Density for neighboring arginine and glutamic acid residues for the dose0.5 volume (D) Density for the same residues in C but with density from the dose4 volume. The dose0.5 volume has more even density values for both the glutamate and arginine.

Fig. 5.

DQE and SSNR for the DE-20 camera. The DQE was estimated using the method described in Ruskin *et al.* (Ruskin et al., 2013) (blue). The frequencies were sampled on the same scale as the data collected for the reconstructions. The SSNR was estimated using the e2ctf tool in EMAN2 on a $65 \text{ e}^-/\text{\AA}^2$ image of carbon collected at the same magnification as the data collected for the reconstructions.

Fig. 6.

Estimate of the maximum number of 100 Å diameter particles that can be imaged on a camera with the same area as the DE-20, and the same pixel size as the AAV-DJ-Arixtra data (1.215 Å/pix). A) Estimate of the total number of usable particles on a micrograph as a function of the total number of particles on a micrograph. Bars represent ± 1 standard deviation. B) Example of the particle distribution with 800 particles per micrograph 300 of which are usable.