Automated cryoelectron microscopy of “single particles” applied to the 26S proteasome

Stephan Nickell, Florian Beck, Andreas Korinek, Oana Mihalache, Wolfgang Baumeister*, Jürgen M. Plitzko

Max-Planck-Institute of Biochemistry, Department of Structural Biology, Am Klopferspitz 18, D-82152 Martinsried, Germany

Received 10 April 2007; accepted 9 May 2007

Available online 21 May 2007

Edited by Thomas C. Marlovits

Abstract The 26S proteasome is a large molecular machine with a central role in intracellular protein degradation in eukaryotes. The 2.5 MDa complex, which is built from two copies each of more than 30 different subunits, is labile and prone to dissociation into subcomplexes. Hence it is difficult if not impossible, to obtain structurally homogeneous preparations and, as a consequence, it is very cumbersome to obtain large numbers of images of the holocomplex. In this communication, we describe an automated procedure for the acquisition of large data sets of cryo-electron micrographs. The application of this procedure to the 26S proteasome from Drosophila has allowed us to determine the three-dimensional structure of the complex to a resolution of 2.9 nm and the prospects for further improvements are good.

Keywords: High-throughput; Electron microscopy; Automated image acquisition; Protein degradation

1. Introduction

Cryoelectron microscopy (CEM) of “single particles” has become a powerful tool for structural studies of macromolecular complexes [1]. Arguably, the term “single particle” analysis is somewhat misleading since it involves the averaging over large data sets of individual molecular images, following their alignment and classification. Nevertheless, the amounts of material required for this type of analysis are minute compared to the quantities needed for X-ray crystallography or NMR. A further advantage of CEM single particle analysis is that it can cope with some heterogeneity of the sample; therefore the material needs not to be purified to exhaustion since some purification can be performed in silico using smart image classification tactics. Moreover, the fact, that the molecules are not confined to a lattice, as they are in 2D or 3D crystals, makes them easily accessible for substrates, ligands or other binding factors. Although it has been demonstrated in a few cases, that resolutions can be attained that are good enough to discern secondary structure elements [2–4] most studies so far have been at a lower resolution (1–3 nm) level. This is often sufficient for hybrid approaches, in which high resolution structures of components (subcomplexes, subunits, and domains) obtained by other methods are fitted into the lower resolution structures of large, multi-subunit complexes yielding pseudo-atomic models [5,6]. Key to the attainment of higher resolution is the availability of large, high quality data sets; this requirement can be difficult to achieve, in particular with samples displaying structural heterogeneity. Automated data acquisition procedures can greatly facilitate the electron microscopic recording of large data sets of consistent quality [7,8]. In this communication, we describe a high-throughput mode of data acquisition and its application to the 26S proteasome.

The 26S proteasome acts at the downstream end of the ubiquitin–proteasome pathway executing the proteolytic cleavage of intracellular proteins marked for destruction by the attachment of multiubiquitin chains. The 26S complex is a large multimeric assembly of more that 30 different subunits [9]. Two major subcomplexes jointly form the 26S holocomplex: the barrel-shaped proteolytic core complex (the 20S proteasome) and the (19S) regulatory complex, which associates with either one or both ends of the core complex [10,11]. While the structure of the 20S proteasome has been determined to atomic resolution [12] the structure of the holocomplex and its mode of operation are only dimly understood. The main role of the regulatory complexes is to prepare substrates for degradation in the 20S core complex. This involves the recognition and binding of ubiquitylated substrates, their deubiquitylation and the unfolding of substrates, which is a prerequisite for translocation into the 20S core [13,9,14]. Progress in determining the structure of the 26S proteasome has been hampered by the low intrinsic stability of the complex, which tends to dissociate during purification and sample preparation. Recently, we have used a tomographic approach to obtain a low resolution (4.6 nm) structure [15] of the complex embedded in ice, unbiased by the use of a starting model. Here, we report on further developments in the structural analysis, using single particle analysis. A large data set of double-capped complexes has been recorded in a fully automated manner and from this the structure of the 26S proteasome has been reconstructed to a resolution of 2.9 nm.

2. Materials and methods

2.1. Isolation and purification of 26S proteasomes from Drosophila

26S proteasomes were purified as described in detail previously [16,17]. Briefly, 0-16-h Drosophila embryos (Yellow white strain) were collected at 25°C from feeding plates. After dechorionation and homogenization, the extract was clarified by centrifugation and nucleic acids were removed by precipitation with 10% streptomycin sulfate.
The supernatant was fractionated with hydroxyapatite in a batch procedure, followed by anion-exchange chromatography (diethylaminoethyl cellulose, DE32, Whatman) and sucrose density gradient centrifugation (15–40% sucrose). At all stages, fractions were tested for their ability to hydrolyze Succinyl-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin (Suc-LLVY-AMC; Bachem), and only fractions corresponding to peaks of activity were used for further purification. Active fractions were examined by electron microscopy and only fractions with relatively large numbers of intact double-capped complexes were used for cryoelectron microscopy. Four-microliter droplets of buffer solution containing 26S proteasomes were applied to lacey carbon grids previously rendered hydrophilic in a plasma cleaner. Excess suspension was removed after 60 s by blotting with filter paper. After a short washing step, using a few microliters of distilled water, the grid was blotted again and plunged into liquid ethane for vitrification.

2.2. High-throughput data acquisition

All data were recorded in a fully automated manner using a Polara G2 (FEI, Eindhoven, The Netherlands) microscope equipped with a GIF 2002 energy filter. The electron microscope was operated in a zero-loss mode at an accelerating voltage of 300 kV; the total magnification on the CCD-plane was 8.900, corresponding to a 0.36 nm pixel size in the object-plane. For all samples, the defocus was set to values between 3.5 μm and 4.5 μm underfocus; at this setting the first zero of the contrast transfer function was between 2.6 nm and 3.0 nm.

For robust automated data collection, we have developed a three-step procedure comprising an initial scan of the central area of an EM grid at low magnification, the selection of positions suitable for imaging (typically one position corresponds to a single mesh of the grid) and the automated acquisition of 2D images at higher magnification.

The grid scan procedure is performed at very low magnifications to minimize the number of micrographs required for complete coverage (magnification: ~300×, object pixel-size: ~100 nm). A square of approximately 1.96 mm² is virtually plotted within a circle with a diameter of 2 mm and images are recorded sequentially within this square (Fig. 1(1)). This circle and the inscribed square are limited by the maximum travel range of the stage, which is typically ±100 μm in the x- and y-direction. It would be preferable to cover the maximum field of view with one single image; however, the magnification cannot be lowered sufficiently and currently available CCD arrays (2048 × 2048 pixels) are too small to image the whole grid. Moreover, when post-column energy filters are used, one has to take the additional post-magnification into account, which will further limit the field of view at low magnifications. The dimensions of the scanned region can be adjusted, if necessary, and smaller areas can be selected around predefined positions. The area mapped is a square with an edge-length defined by the number of images and the magnification. A typical map would be assembled from 13 × 13 micrographs, resulting in a total 169 images used to generate the map. The acquisition time is less than 15 min. The low-dose state, in which this first part of the automated acquisition should take place, can be selected within the grid scan setup and additionally one can specify whether or not the acquired images are to be saved separately after assembling the map.

To facilitate the generation of the whole map and to minimize displacement errors as is necessary for retrieving positions accurately, the stage has to be pre-calibrated in two different ways. The first calibration step is used to transfer the stage coordinates onto the CCD coordinates, while the second step converts relative CCD coordinates to stage coordinates to increase the accuracy of the stage positioning.

Multiple positions for image acquisition at higher magnification can be selected and at every chosen position the eucentric height is manually pre-adjusted and stored. This minimizes defocus changes due to deviations from grid flatness or variations in specimen thickness.

After identification of suitable areas the number of images to be recorded within one area can be defined; typically it is in the range of several hundreds. For the acquisition at high magnification (in this case M ~82500×, object pixel-size ~0.36 nm) the required low-dose states need to be pre-set. The single-particle data acquisition is fully embedded within the tomographic data acquisition software of the TOM toolbox[18] and for the sake of simplicity we have utilized the ‘Search’, ‘Focus’, ‘Tracking’ and the ‘Acquisition’ states to perform the different tasks. ‘Search’ is only used for the grid scan at very low magnification. The adjustment of the focus is done with an advanced version of the ‘Focus’ state, mainly to further increase the reliability of the final acquisition process and to ensure stability of the electron optical setup. Focus changes larger than ~2 μm are first compensated by adjustments along z, before refining by beam tilt method. Instabilities are common when the objective lens settings are changed (mainly due to hysteresis), this can be minimized or completely avoided by this approach.

Cross-correlation based routines are facilitated by the presence of high-contrast features. Since such features are normally missing in frozen-hydrated samples, we routinely generated pairs of holes in the ice
layer by concentrating the beam with the condenser system (Fig. 1(4)). The ‘Tracking’ state serves as a second ‘Acquisition’ state for the recording at full resolution ( unbinned) but with a different focus setting. The movement of the stage is done after focusing and before acquisition and the distance between every acquisition is adjusted with the beam diameter. The diameter is determined beforehand by measuring the curvature at the rim of the beam. Typically, two images are one beam diameter apart with an additional tolerance of 10% to account for stage inaccuracies. Moreover, with the knowledge of the beam diameter, accidental double exposure of potential acquisition areas can be avoided.

The position of the last acquired image is used for the adjustment of the focus. Overall movement of the stage is comparable to the grid scan procedure, with the only difference being, that at this high magnification images do not overlap. The trail of this acquisition scheme can be followed (overview at intermediate magnification, Fig. 1(3)) from the pattern of holes left behind by the focusing procedure.

Currently, our automated procedure does not discriminate between ‘good’ or ‘bad’ images during acquisition, it simply images the complete area of a selected mesh sequentially, stores the images after acquisition and continues with the next selected area, until all pre-selected positions have been addressed. A ‘defocus map’ of the entire mesh is created (Fig. 3a), which reflects the relative flatness of this region and provides a basis for a post-acquisition focus determination (see below). If the contrast is too low for a direct determination of the contrast transfer function it is approximated by means of a bicubic interpolation using the values determined for neighbouring images (Fig. 3a).

2.3. Image processing

After data acquisition, all electron micrographs and their corresponding power spectra were visually inspected. Images lacking 26S complexes or showing indications of sample drift in the power spectra were sorted out. Altogether 2417 micrographs out of a total number of 4181 micrographs were selected for further image processing. Particles showing the typical shape of the 26S proteasome, the 20S core flanked by two 19S complexes, were selected in an interactive manner and a particle stack totalling 16742 particles with a box-size of 160 × 160 pixels was created. To reduce computing time particles were resized to a box-size of 64 × 64 pixels and processed with EMAN [19], a software package for semi-automated single-particle reconstruction. After generating an initial model by a Fourier-common lines approach the three-dimensional reconstruction was iteratively refined until the observed changes became negligible. In a separate processing step, the contrast transfer functions (CTFs), respectively, the defocus values for all electron micrographs were determined by using a fitting routine implemented in the TOM toolbox [18], based upon a method described elsewhere [20]. The resulting CTFs were visually examined, outliers were identified and set to a weighted average of successfully determined defocus values of the surrounding micrographs. Using the individually calculated CTFs and the previously determined modulation

Fig. 2. (a) Automatically recorded electron micrograph of an ice-embedded 26S proteasome preparation. 20S core particles are mostly visible as ‘top’ views (ring structures) while 26S proteasomes appear exclusively as ‘side’ views with one or two 19S regulatory complexes attached to the core. In this study, only 26S complexes with two 19S cap complexes were analyzed. The scale bar corresponds to 100 nm. (b) Class averages of several distinct side views of the 26S proteasome. The scale bar corresponds to 50 nm.

Fig. 3. (a) Map of defocus values of the automatically acquired electron micrographs for one grid region. Defocus values were determined by a CTF fitting procedure. (b) Fourier-shell correlation function calculated from two separate sets of particles. The dashed line shows the resolution after applying a C2 symmetry to the set of data. The dotted line at a Fourier-shell correlation coefficient of 0.5 indicates a resolution of 2.9 nm.
transfer function (MTF) for the CCD device, the selected electron micrographs were de-convoluted by combining phase flipping [1] and MTF correction [21]; the frequency cut-off was set to 2.0 nm. Using the corrected particle images the reconstruction was further refined by applying a likelihood-based classification method [22] integrated in the Xmipp [23] software package. After a first refinement with a 10° search increment the angular increment was reduced to 5°; the refinement was iteratively repeated until the changes were negligible. In a final run the reconstruction was further refined using one time binned images (80 x 80 pixels) and a 5° angular search increment, until the model converged. C2-symmetrization was applied to the final reconstruction [15]. All image processing steps were carried out on Linux workstations. Image selection, particle picking and CTF/MTF correction was performed using the TOM toolbox running under Matlab.

Fig. 4. The left and center column show isosurface representations of the three-dimensional reconstruction of the 26S proteasome complex. The view direction corresponds to the C2-axis of the 20S core particle. The isosurface threshold was set to include a protein mass of 2.5 MDa. The crystal structure of the 20S catalytic core particle of yeast was low-pass filtered to a resolution of 2.0 nm (red mesh) and fitted by an extensive-search correlation algorithm. From top to bottom a rotation around the pseudo 7-fold axis of the 20S particle was applied resulting in seven different views of the complex. The center column shows isosurfaces of the half-cut three-dimensional reconstruction while in the right column central orthoslices of the density distribution (inverted) of the reconstruction are displayed.
3. Results

Drosophila embryos are a rich source of 26S proteasomes and yield biochemically “clean” preparations with a well-defined complement of subunits [24]. But in spite of all efforts to maintain the integrity of the complexes throughout purification and sample preparation, electron micrographs show a bewildering structural heterogeneity: 26S complexes, some single-capped others double-capped, coexist with 20S core and 19S regulatory complexes (Fig. 2a). At the chosen magnification of 82500× on a single image with a field of view of 740×740 nm less than 10 double-capped holocomplexes are found; this corresponds to less than 5% of the total recorded area. As a consequence, several thousand micrographs must be recorded to build up an adequate data set for a detailed structural analysis.

While the manual acquisition of that number of images would be very laborious we were able to take 4181 images within six days in a fully automated manner. Of these, 2417 (58%) met our quality criteria (presence of holocomplexes, ice thickness, focus value, acceptable specimen drift) and were processed further yielding a total data set of 16742 particles. Fig. 2b shows a gallery of selected class averages. The asymmetry of the regulatory complexes causes their appearance to be strongly dependent on their orientation around the long axis.

From the data set a three-dimensional reconstruction has been obtained using an angular reconstitution approach. The nominal resolution is 2.9 nm according to the Fourier-shell correlation using a cross-correlation value of 0.5 as a conservative estimate (Fig. 3b); when the less stringent 0.3 criterion is applied, resolution is ~2.25 nm; in any case the curve indicates that the map contains significant information beyond 2.9 nm.

In Fig. 4, the three-dimensional structure of the 26S complex is displayed in exactly the same manner as in a previous report describing its structural analysis by cryoelectron tomography (Fig. 4 in [15]). In spite of the difference in resolution (4.6 nm vs. 2.9 nm) the agreement between the two structures obtained is very good. Since the 26S complex contains the 20S core complex and since high resolution structures from X-ray crystallography are available for 20S complexes from various organisms [25–27], all very similar to each other, these structures can be used as internal standards in assessing and benchmarking the quality of the whole structure. In Fig. 4, we have superimposed the crystal structure of the 20S complex from yeast [26] onto the EM map. The yeast map has been low-pass filtered to a resolution of 2 nm (red mesh) and fitted computationally. Obviously, the agreement is excellent, confirming that the structure is indeed very accurate. The new structure, shown in Fig. 4 clarifies several aspects of the molecular architecture of the regulatory complexes of the 26S proteasome but a detailed discussion is beyond the scope of this article.

4. Discussion

Single-particle cryoelectron microscopy is a powerful tool for studying large molecular complexes and for labile assemblies, such as the 26S proteasome, perhaps the only viable approach. Unfortunately the method is slow; in particular the data collection can be very time consuming. Even when resolution targets are modest the data acquisition can take weeks or even months putting pressure on expensive infrastructure. Therefore, there is a strong incentive to develop automated procedures for data acquisition enabling the operation of microscopes in a high-throughput mode [7,8]. In fact, the recording of images of consistent quality involves blindly following a prescribed sequence of repetitive operations, a work better suited for a computer than a human operator [6].

The automated data acquisition procedure described in this communication has proven to be reliable and efficient. It has allowed us to collect a large data set of more than 4000 images in less than a week and still there is room for improvement. Throughput could be increased dramatically if large area CCD cameras with multiple port readout capabilities were available or if “settling times” of the specimen stage and thus the waiting time between subsequent recordings could be further reduced. The procedure is simple and robust, and it is a distinct advantage that only two different imaging conditions (e.g. two different magnifications) have to be set up initially. During the final acquisition no changes in magnification are necessary and major changes of the optical system are avoided. This, in turn, avoids hysteresis effects, which otherwise make it necessary to perform time-consuming iterative alignment and normalization procedures.

The automated acquisition procedure can run for several days/weeks, provided that the specimen stays cold, remains contamination free and that the grid at hand exhibits sufficient areas for an in-depth investigation. The Tecnai Polara instrument is very well suited for this mode of operation since specimens can be kept cold for 24 h, before liquid nitrogen has to be refilled. The internal stage-holder system using small grid carrying cartridges having no external parts provides greatly enhanced stability. Direct user interaction with the delicate machinery, as required with conventional side-entry systems, is completely avoided. Nevertheless, conventional side entry holders can be used as well for automated acquisition with the only restriction that every 3–4 h the dewar of the holder has to be refilled and the system has to be recalibrated. One has to take into account the drift behavior of the holder after stage movements, which can require a waiting period of several minutes. The procedure is adaptable to any of the various specimen supports being used as it neither requires special grid designs nor coating styles (e.g. Quantifoil).

The automated acquisition procedure has enabled us to collect a large data set of 26S proteasome particles – in spite of the low abundance of intact holocomplexes. The resulting structure represents a significant improvement over previous structures and it is by all criteria very trustworthy. Further improvements can be expected from careful classifications taking into account local differences in subunit occupancy or other dynamic phenomena. For bringing resolution to the subnanometer level it will be necessary to collect data sets 10- or perhaps 100-fold larger than the present one. Using automated procedures and with larger and faster detectors in place this is not an unrealistic goal.

Acknowledgements: This work was supported by the 3D Repertoire Grant, the 3D EM Network of Excellence Grant and the High throughput-3DEM Grant all within the Research Framework.
Programme 6 (FP6) of the European Commission. We thank Sjors Scheres and José Maria Carazo for support with the Xmipp Program Package, Reiner Hegerl for help with the CTF/MTF correction routines and Dennis Thomas for critical reading.

References


