Less data for higher resolution in-situ with M

Greg Kafetzopoulos

6678149

Minor research project

Less data for higher resolution in-situ with M

Greg Kafetzopoulos (6678149)

Layman summary

Imagine if we could look inside the cell and actually see its tiny machinery function. That would help very much in understanding how life works and maybe it could shed some light on some of its malfunctions such as diseases. Well actually we can, with a method called cryo electron tomography (CET). In CET we rotate the sample to image it in different angles which provides enough information to create a 3D map of it. Therefore, we can create maps where we can see cells viruses and organelles in 3D.

These maps are a bit blurry so we can barely see big molecules. Nevertheless, we want to have a clearer look at these molecules to understand their function. So what we usually do is find many instances of the same type of molecule in a map and average their structure. Through this process called sub tomogram averaging (STA), we can combine the weak signal of many blurry molecules into one well resolved 3D molecule map.

But we can go even further with a new method called refinement. Refinement tracks and correct motions that happened in the sample while imaging it. During imaging the whole sample moves a little bit and the electron beam which we used to take the images creates local movements. By correcting these alterations, we can resolve our 3D- map even better.

In this research I used STA and refinement to create 3D maps of ribosomes bound to the endoplasmic reticulum (ER). Ribosomes are large biomolecules which synthesize proteins. Proteins in turn are required for most of the cell's functions therefore ribosomes have an essential role to cell life. Defective ribosomes lead to disease whereas ribosome drugs can be designed to eliminate pathogens such as bacteria. Therefore, unraveling the structure and functions of the ribosome will aid the treatment of diseases and the design of new pharmaceuticals. Ribosomes are also a great target for CET because they are relatively big molecules.

I used ribosome samples to test the performance of a refinement software vs a conventional STA software. I observed that refinement outperformed the conventional approach. The refined ribosome maps revealed much more detail than the ones from STA. This indicates that the field is progressing and new biological insights from this advance are to be expected.

CET is also very effective in locations where other methods can't reach such as under biological membranes. To study molecules residing under membranes most techniques need to destroy them. However, this also alters the molecules under study. With CET we can just image such targets. Here I observed that using refinement on molecule machineries under the ER membrane performed better than the conventional approach using less images for the same quality of map. Overall my research suggests that refinement will allow us to produce

more detailed molecule maps from less samples than with previous approaches. That will help us reach and study molecular machineries which support life.

Abstract

Cryo Electron Tomography (CET) has advanced the capabilities of structural biology research, allowing the structural characterization of large macromolecular complexes in their native environment. In the past decade major efforts have been made in order to increase the resolution feasible with CET. The main strategy to achieve this is averaging of identical molecules within 3D maps of the sample. This combines information present in the data, thus increasing the resolution. This process is termed sub tomogram averaging (STA) and is commonly combined with classification of the subtomograms into groups of distinct structures.

The mentioned processes are supported by various tilt series processing software available such as PyTOM, RELION and EMAN2. Recently this pipeline has been supplemented with a final refinement shown to further improve the resolution through improving the tilt series alignment.

Here I compared PyTOM as a representative of the conventional STA & classification approach to Warp-M, a pipeline entailing a final-multi particle refinement in M. I reconstructed endoplasmic reticulum (ER)-derived ribosomes with both software packages, using two datasets of different ice thickness. In all the cases Warp-M outperformed PyTOM in ribosome resolution whereas the ice thickness negatively affected the final resolution in both software. The highest resolution was observed in the Warp-M thin ice dataset in which refinement yielded an 8.2 Å structure using 2,317 subtomograms. This was improved to 5.6 Å by increasing the number of subtomograms to 7,323. Finally, Warp-M was also tested in reconstructing complexes residing in the ER lumen. To this end I classified the previous set and selected 1,664 TRAP & OST bound ribosomes, which yielded a 9.5 Å reconstruction of the TRAP & OST complex bound ribosome.

Introduction

Cryo Electron Tomography (CET) is a novel technique in the field of molecular imaging which has evolved significantly during the past decade. It allows the characterization of molecular structures at increasingly high resolution, from dense and diverse molecular environments such as cells, organelles and biomolecules.[1–4] CET allows the determination of structures in their native environment in contrast to single particle analysis (SPA). SPA is a cryo electron microscopy method shown to reach atomic resolution but requires purification of the target molecule. [5–7] Therefore, CET has advanced significantly the capabilities of cryo electron microscopy enabling observations in situ.

During image acquisition, samples are tilted around an axis vertical to the incoming electron beam in order to capture the objects of interest from different angles. These 2D tilt series are then computationally aligned to correct translations and rotations of the image as

well as magnification differences[8]. Tilt series alignment (TSA) is usually performed via tracking gold bead (fiducial) markers along each tilt image. Then the aligned images are back projected[9] to produce a 3D volume of the sample, called a tomogram. Thus tomogram reconstruction and therefore its resolution relies on the performance of the tilt series alignment (TSA). Furthermore, sub-volumes of interest, known as subtomograms, can be selected from the tomogram. Subtomograms could be molecular complexes or whole organelles.

The raw 2D images acquired with CET are usually noisy due to low dose of electrons per tilt which in turn limits the resolution of the tomogram[10]. To compensate for this, a variety of software have been developed such as EMAN2[11], Relion[12] and PyTOM[13]. The computational processing of tomographic data is purposed to produce a subtomogram reconstruction with increased signal to noise ratio (SNR), and thus resolution, in comparison to the tomogram. In previous years, the main means to improve the resolution of subtomograms was STA. In STA high resolution signal distributed in the tomogram(s) is incorporated in a single reconstruction (average). This is accomplished by iteratively aligning and averaging subtomograms depicting the same type of molecule. This process significantly increases the SNR of the subtomogram average and therefore its resolution.

STA is usually combined with classification, a process to categorize the particles into populations with different structural features, which can in turn increase the resolution of the subtomogram average. STA coupled with classification has been previously shown to yield a ribosome average at a resolution of ~9 Å using a total of 17,600 subtomograms in PyTOM[14]. Overall, with the STA& classification strategy it is currently feasible to obtain structures of resolution as high as 15 Å and reaching resolutions better than 8 Å is considered exceptional.[15]

The mentioned routine has been enriched with further stages of processing such as refinement and post-processing which have led to significant increases in final resolution. Refinement is a process in which a subtomogram reconstruction is projected on the tilt series. These high SNR projections are used to track the movement of particles which allows the optimization of their position and orientation. Refinement can improve the tilt series alignment, the motion correction as well as the subtomogram alignment. [16]

Recently published refinement software such as M[17], EMAN2[11] and emCLARITY[15] are representatives of this current advance in tilt series processing. EMAN2 utilizes a per particle per tilt strategy in which the orientation of each tilt image of each particle is refined. Using EMAN2, refinement yielded a ribosome structure of 9.3 Å, using a set of 3,000 particles[11]. emCLARITY groups proximal particles in patches in order to register their movement and improve the tilt series alignment. With emCLARITY an 8.6 Å ribosome structure has been refined from an initial set of 3,090 particles[15]. Finally, M, also utilized here, uses a multi-particle approach. In this, Particle trajectories are also tracked per particle and per tilt however the calculation of grid deformation occurs simultaneously for all the particles within the tilt series. This deformation model corrects global motion of the sample such as stage translation and rotation. The model also accounts local non-uniform alterations modelled on 2D grid representations of the tilt images (image space). Furthermore, it depicts deformation

such as shearing or bending in the z dimension via a 3D grid representation of the sample (volume space). The multi particle deformation model utilizes less parameters to address sample alterations that other refinement approaches which optimize the alignment of each particle separately. M also corrects for further parameters such as per tilt defocus and high order aberrations on the full tilt series. M's multi particle refinement approach has allowed the refinement of a 3.5 Å ribosome-antibiotic complex structure using 17,890 particles inside whole bacterial cells [17,18].

Ribosomes are bio-molecular complexes with the role of translation. Translation in eukaryotes is facilitated both in the cytoplasm by soluble ribosomes and in the rough compartment of the ER, by membrane bound ribosomes. In the ER newly synthesized peptides are inserted into the lumen upon translation where they are folded and subdued to modifications such as glycosylation. Also, in the lumen, peptides destined to be transmembrane proteins are inserted into the ER membrane. [19,20,21] These processes are performed by the translocon. Translocon machineries are essential for mammalian cell life as they are responsible for the translation, processing and quality control of a significant proportion of proteins required for secretion[22], signaling[23] and various other processes. The translocon is a dynamic transmembrane complex whose invariant constituent component is the Sec61 protein conducting channel. Sec61 is bound to ER membrane ribosomes, allowing unfolded peptide chains to be inserted into the lumen or the ER membrane, in case of transmembrane proteins[22,24]. Apart from the mentioned Sec61, the translocon is variable in composition as it entails transient and stable interactions with a variety of accessory factors, which compose different translocon assemblies within the cell. [22] A prominent form in mammalian cells contains the TRAP complex, which facilitates signal peptide insertion to the ER lumen in a substrate-dependent manner and OST, which catalyzes peptide glycosylation.[3,25,26] The translocon is an ideal target for CET and a difficult case for most other structural methods such as NMR and X-ray crystallography[27], because they form large macromolecular assemblies which span on the extra-membrane, intra- membrane and submembrane space and support various transient interactions.

In this study I aimed to compare the performance of a conventional classification & STA protocol, to a pipeline entailing refinement. To this end I computationally processed tilt series depicting endoplasmic reticulum derived vesicles (microsomes) obtained from HEK293F cells. These were used to characterize the structure of ER membrane-associated ribosomes. I tested the performance of two tilt series processing pipelines PyTOM[13] and Warp-M[17,28], in reconstructing a ribosome map.

Materials & Methods

Tilt Series Data

HEK WT cell endoplasmic reticulum microsomes were prepared as described by Walter et. al[29]. Tilt series were collected using a Talos Arctica electron microscope. The

acquisition tilt range was -51-51° with an increment of 3°. The pixel size of the raw micrographs was 1.724 Å/pixel.

PyTOM pipeline

I used PyTOM[13] to process the raw tilt series in my datasets mentioned above. Motion Corr2[30] plug-in was used to align frame tilts in order to compensate for beam induced motion (BIM). Consequently, fiducials were picked automatically and filtered manually prior to TSA to increase its performance. CTF parameters were determined for the tilt series using IMOD's CTFplotter[31] via its PyTOM plug-in and CTF correction through phase flipping was performed. I then proceeded with TSA and tomographic reconstruction. PyTOM's Weighted Back Projection was used to reconstruct tomograms at a Bin factor 8. PyTOM calculates residual error mean as a measure of the TSA's performance. The residuals were used to manually discard poorly aligned tilt series-tomograms. To identify ribosomes in the tomograms I performed template matching using a 40 Å resolution template of the mammalian ribosome. The picked particles were then subdued to further manual selection. I then used the selected ribosomes to batch reconstruct subtomograms at a Bin factor 2.

To obtain an average ribosome structure I aligned the subtomograms in real space using Glocal. A spherical mask was utilized to cancel unnecessary noise. Finally AutoFocus classification was performed using the aligned particle rotations and coordinates. This classification was done to discard non-ribosome subtomograms still present in the dataset. The discarded particles correspond to the 10 % of the sub tomograms. I then aligned the remaining selections with Glocal. Finally, Fourier shell Correlation was used to estimate the resolution of the averages obtained from Glocal. For the FSC I used tight masks created in PyTOM, to avoid noise influencing the correlation (Fig S1a).

Warp-M plus PyTOM pipeline

Warp-M[17,18] is a recently published processing pipeline shown to produce high resolution subtomogram reconstructions[18]. This pipeline, unlike PyTOM, uses a variety of different software available to perform different stages of the tilt series processing. These are namely Warp[28], IMOD[31], Relion[12] and M in the published version of this pipeline. Here I slightly modified this routine by including PyTOM functions as well.

As in the published pipeline[18,28], I preprocessed the raw tilt series in Warp, a process which contains an initial CTF estimation and motion correction and outputs tilt movie averages. These averages are then converted to MRC stacks in order to be aligned in IMOD. In more detail IMOD's etomo is used for additional pre-processing of the stacks, such as X-ray depletion and for the alignment of the frames per tilt in the tilt series. A preliminary Coarse alignment was performed and subsequently fiducials were selected and tracked along the tilt series. A final fine alignment was also used to correct tracking errors and increase the alignment's accuracy. IMOD calculates the residual error mean for the alignment model which was used to reject tilt series. CTF was re-estimated in Warp for the aligned tilt series and tomograms were reconstructed in the same software at a Bin value of 20 Å/pixel.

At this stage I diverged from the published routine[17,18], using PyTOM's template matching and batch extraction functions to select ribosomes in my tomograms. I used the same 40 Å template as in the previous chapter, for the template matching. Upon selecting the particles, I had to convert their data (e.g. micrograph name coordinates and rotations) in a STAR format readable by Warp and M. To that end I used pl_merge_name.py, and extractXYZ_mod.sh in order to extract particle coordinates, rotations and corresponding tomogram name from PyTOM's xml files containing TM selections. PyTOM and WarpM use different Euler's angles to describe rotations. Therefore, I used PL2Star_mod.m to convert the previously obtained rotations from ZXZ to ZYZ. The mentioned data were combined and formatted to a STAR file by pl2star.sh.

I then extracted sub-tomograms using acquired STAR files in Warp, the sub tomograms were reconstructed at a Bin factor 2. These were submitted to an initial 3D refinement in Relion in order to obtain an average, with the function 3D auto-refine. Refinements in Relion were performed without CTF correction. The unfiltered average half maps produced by the process were then submitted as reference for additional multi-particle refinement with M. The subtomograms were also classified in Relion in order to deplete particle classes not corresponding to ribosomes. After this "denoising" classification the selected ribosomes were submitted once more to M for refinement. M refinement is accompanied by FSC estimation of resolution (Fig S2) and each refinement was performed in as many rounds necessary to reach a plateau in resolution. This was 3-4 rounds of refinement in all the cases.

M uses tights masks for FSC, created from the submitted maps. Here I also performed FSC for Warp-M's reconstructions in PyTOM. This was done to ensure comparability in the calculated resolutions. I used tight masks created in PyTOM in all FSCs performed in PyTOM for Warp-M reconstructions. (Fig S1).

Results

Warp-M notably outperforms PyTOM

I aimed to assess the performance of PyTOM vs Warp-M, thus I used each pipeline to process tilt series of the same data set. Ribosomes obtained by template matching were adjusted to 3,279 particles for both pipelines through excluding tomograms. However, the selected particles, and therefore corresponding tomograms, weren't identical amongst Warp and PyTOM. The particles were then subjected to classification to discard false-positives and poorly aligned particles. In PyTOM, an Auto Focused classification was performed, excluding 10% of the dataset whereas in Warp-M, classification was performed using Relion's 3D-classification through which 224 of the 3,279 particles were rejected. After classification I re-adjusted the number of particles in the two datasets to 2,317, to ensure comparability. STA in PyTOM using the selected 2,317 particles from 23 tomograms yielded a reconstruction of 12.8 Å (Fig 1a,e). In Warp-M refinement of an equal number of particles from 18 tomograms
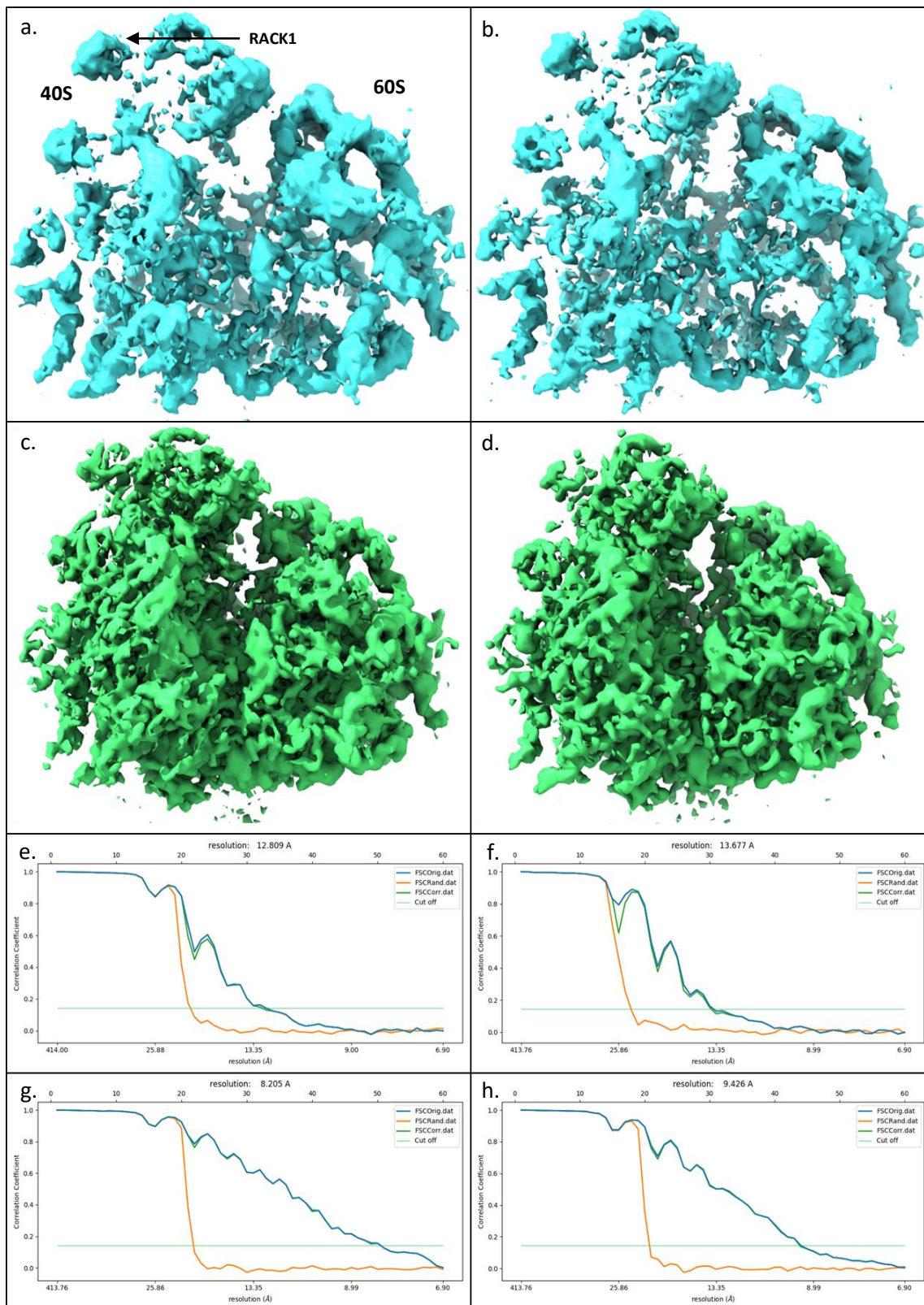
Figure 1 PyTOM vs WarpM reconstructions

a. 12.8 Å Subtomogram average of the PyTOM thin ice dataset, yielded after classification and selection of 2317 particles. Depicted in a is the front ribosome view, 40S and 60s correspond to the small and large ribosomal subunit. The peripheral protein RACK1 (receptor of activated C kinase 1) appears more intense than central elements b. 13.7 Å subtomogram average of the PyTOM thick ice dataset, obtained upon Auto Focus classification and selection of 2317 particles.

c. Front view of the 8.2 Å reconstruction produced with the Warp-M pipeline from 2,317 particles, selected with classification from the thin ice dataset. d. Front view of the 9.4 Å average reconstruction obtained with the Warp M pipeline from 2,317 particles, selected with classification from the thick ice dataset. e,f,g,h. PyTOM thin, PyTOM thick, Warp-M thin, Warp-M thick FSC curves accordingly.

yielded a reconstruction of 8.2 Å. (Fig 1c,g). Therefore, I could observe that the Warp-M pipeline significantly outperformed PyTOM achieving a higher resolution reconstruction. Furthermore, peripheral parts of PyTOM's reconstruction such as the RACK1 complex (Fig 1a) and peripheral 28S rRNA, were more intensely depicted than elements in the center such as 28S rRNA strands. (Fig S3a,b). This effect was not observed in M's reconstructions.

Finally, I obtained a 7.5 Å reconstruction in Warp-M using the classified set prior to adjustment (3,055 particles). (Fig S4a,b) This particle number was used to compare Warp-M's performance with recent published reconstructions using refinement as will be discussed following.

Ice thickness decreases the resolution

I further assessed the performance of Warp-M and PyTOM in datasets of different ice thickness. My goal was to determine the impact of ice thickness on resolution. Therefore, I processed an additional dataset, thicker than the previous, hereby mentioned as thin. I processed the thick set as previously. Upon tomographic reconstruction, template matching and selection of 3,279 particles in each case, I calculated the tomogram thickness for the thick and the thin datasets. Ice thickness was measured on the superset of tomograms selected in PyTOM and Warp-M, for the thick and thin ice tomogram datasets. On average the first was 235 ± 34 nm on the Z axis and the second 184 ± 29 nm. (Fig 2b).
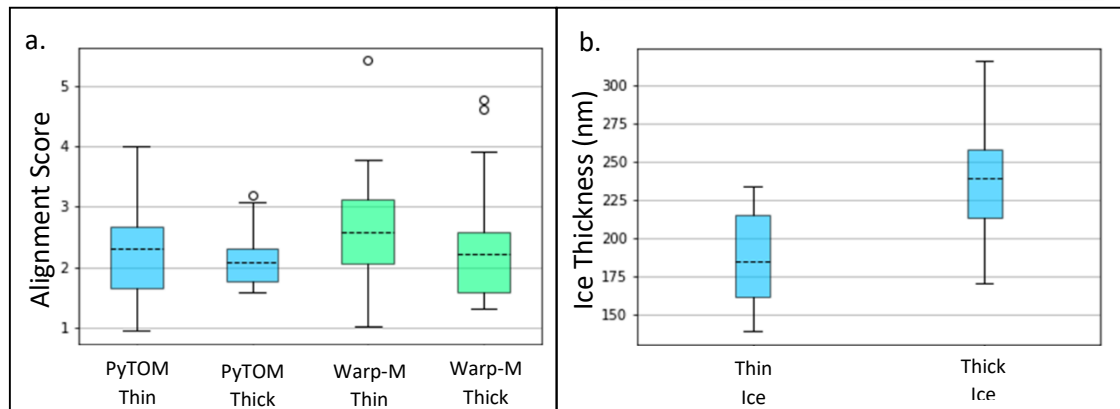
Figure 2 Thick & thin ice dataset analysis

a. Distribution of tilt series alignment scores (Residual error means) for the PyTOM (blue) and WarpM (green) datasets of 3279 subtomograms. The black line corresponds to the mean value for each dataset and the error bars indicate the range of values. b. Tomogram ice thickness distributions in nm for the thin ice and thick ice datasets.

To ensure that the resolution of the final reconstruction is affected by ice thickness and not differences in alignment I compared the TSA performance amongst the datasets. Hence I compared corresponding alignment scores or residual error means. This metric is the RMSD of the coordinates of fiducials in tilt images vs coordinates calculated via the alignment model. Thus it assesses the accuracy of the alignment.

The average alignment score for the tilt series selected in PyTOM's thin dataset was 2.3 ± 0.8 nm and 2.1 ± 0.5 nm for the thick set. The Warp-M alignments had average residual error means of 2.6 ± 1 nm and 2.2 ± 0.9 nm for the thin and thick sets accordingly. (Fig 2a). Therefore, the alignment scores did not differ significantly amongst the datasets. Upon classification of the thick sets I adjusted the selected particles to reach 2,317 as in the thin, this selection corresponded to 18 tomograms for the PyTOM set and 29 for the Warp-M set. PyTOM's reconstruction upon STA was 13.7 Å (Fig 1b,d,f,h) whereas Warp-M yielded a structure at 9.4 Å resolution. Therefore, in both software the resolution was limited by the sample thickness.

Ribosome reconstruction at 5.6 Å with Warp-M

Previously I observed that Warp-M performed better than PyTOM in terms of resolution. Also ice thickness was an important factor limiting the potential resolution of the subtomogram average. I then proceeded with processing the Warp–M thin set and incorporated all the particles available in the dataset, reaching a total of 8,288 subtomograms. As previously I classified this set and discarded 965 particles corresponding to false-positive hits or poorly aligned particles. The refinement of this set resulted to a reconstruction of 5.6 Å. (Fig. 3a-c, S5) In this map I could fit a 80S molecular ribosome model obtained from a 3.4 Å SPA reconstruction[32]. The fitting of the molecular model illustrated structural details resolved in my 5.7 Å map. In this I could clearly distinguish single α-helices as tubular densities, beta sheets and rRNA. (Fig. 3d-f)

OST bound ribosome reconstruction at 9.5 Å with Warp-M

These results signified Warp-M as a robust pipeline capable of producing high resolution subtomogram reconstructions from CET data. My focus so far was the reconstruction of a full ribosome and averages were centered to the cytosolic part of the ribosome-translocon complex. Nevertheless, the luminal compartment of such complexes is usually involved in important biological processes.

Therefore, my next goal was to test Warp-M's performance on ER luminal complexes. For this I focused on the OST complex. Hence I performed a 3D classification with Relion. In detail 7,323 tomograms previously chosen from the thin set were further classified focused on the luminal area of the translocon complex. 4,960 particles lacked the density of the ER membrane and thus this class can be assigned to a population of soluble ribosomes, 699 corresponded to an unidentified translocon and 1,664 subtomograms corresponded to OST-translocons. (Fig. 4a-c) The last were used for refinement in M, in which I refined whole ribosomes. The initial goal behind this was to align the subtomograms before centering them to the OST. However even at this stage I obtained a 9.5 Å reconstruction of a full ribosome
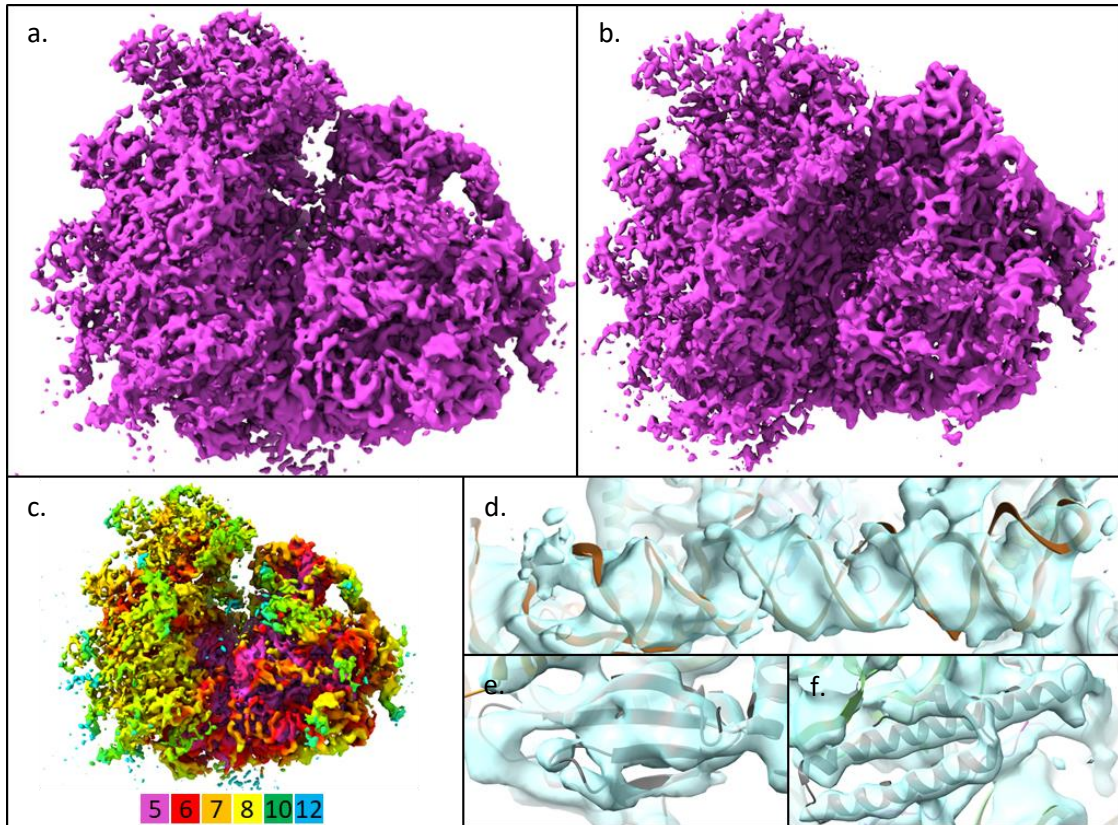
**Figure 3 5.6 Å reconstruction with Warp-M**

a,b. Front and top view of the 5.6 Å average obtained upon classification of 8266 thin ice particles and selection of 7323 with Warp-M. c Resolution distribution in the surface of the previous map (colour scale in Å). d. 28S rRNA molecular model fitted in the 5.6 Å map. e,f. beta-sheet and alpha-helix from the ribosomal proteins uL6 and uL29 accordingly, fitted in the previous map.

along with the OST complex. (Fig. 4d-f) The local resolution in the TRAP-OST region was 12.8 Å. (Fig. 4g) Further processing using subtomograms centered on the translocon did not improve the resolution. This could be explained by the low particle number resulting to low signal.

Discussion

In this study I tested the performance of PyTOM and Warp-M at producing a subtomogram average from *in-situ* ribosome samples. I found that Warp-M produced higher resolution averages than PyTOM with a 2,317 particle set. This observation was made on different sample thicknesses. Tomogram thickness was found to negatively affect the resolution of reconstructions in both software. Furthermore, PyTOM reconstructions had more intense peripheral structural elements in comparison to central components of the ribosome.

Increased intensity on specific domains of cryo EM reconstructions has been attributed to insufficient CTF correction even in cases were phase flipping was utilized[33]. With conventional phase flipping, spatial frequencies close to contrast zeros in the CTF curve have
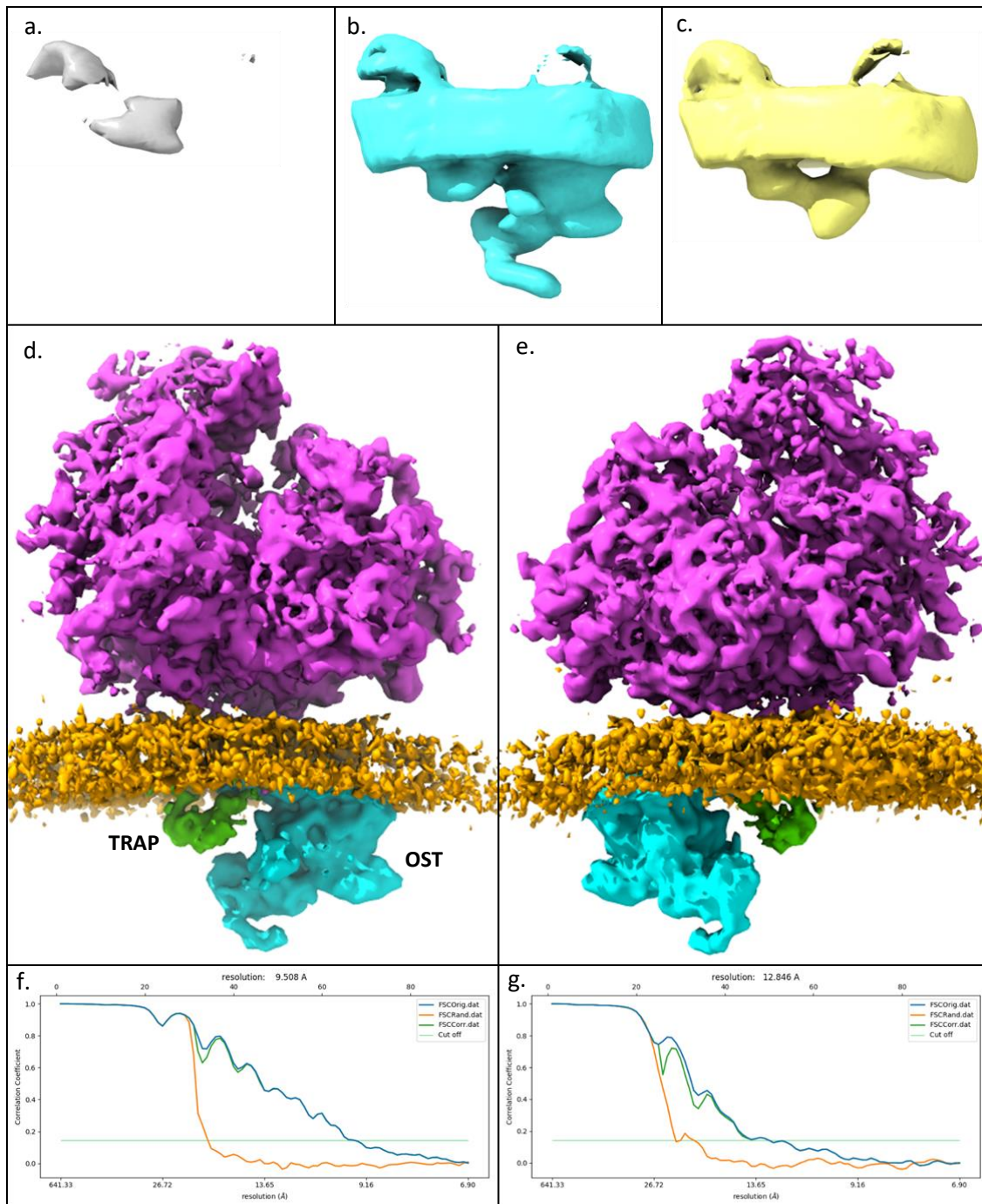
Figure 4 Whole ribosome OST refinement.

a,b,c. Reconstructions obtained by a 3d classification of 7323 particles focused on the luminal part of translocon. a. 4960 particles were classified as soluble as the luminal signal could not produce a reconstruction. b. 1664 particles were found to contain OST& TRAP bound translocons. c. 669 particles were classified as unidentified translocons. d,e. Depicted here is the 9.5 Å average obtained upon classification of 7323 particles, focused on the OST complex. (Purple; ribosome, Orange; ER membrane, Green; TRAP complex, Blue; OST complex) This structure was produced by the selection and refinement of 1664 particles that contained the OST. In the image the OST complex has been filtered to 13 Å. f. FSC curve of the whole ribosome–translocon complex determined at a global resolution of 9.5 Å. g. FSC estimated the local resolution for the TRAP & OST complex at 12.8 Å.

low amplitudes. Especially high (resolution) frequencies as CTF deteriorates. Thus the information entailed in these frequencies has minimal contribution in the reconstructions. This can lead to artefacts as well as differences in intensity of the reconstructed domains. This effect was not observed in M's reconstructions where a Wiener-like filter is applied in CTF refinement[18]. Through such filtering all the frequencies apart from zero contrast values (theoretically) contribute equally to the reconstruction[34].

The 7.5 Å ribosome reconstruction obtained with the M thin set using 3,055 particles vastly exceeded the reported performance of conventional STA & classification with Relion, yielding a 13 Å ribosome reconstruction from 3120 particles[35]. It also outperformed recently published reconstructions obtained from EMAN2 and emCLARITY refinement of equivalent numbers of subtomograms.[14,15]

Increasing the number of particles to 7,323 yielded a reconstruction of 5.6 Å. This result also exceeds the performance of emCLARITY, in a study reporting a ~7 Å 70S ribosome reconstruction from a dataset of 9,923 ribosomes.[36,37] Furthermore, my findings are consistent with previous research illustrating a 3.5 Å ribosome reconstruction by Warp-M using 17,890 particles[18].

The comparison of conventional STA& classification to refinement protocols led to the conclusion that CET subtomogram averages can significantly improve in resolution through refinement. In refinement much more control points can be used (e.g. particles) to assess the global and local motion of the sample than conventional global alignment using fiducials. M specifically is a pioneer in the field of refinement as it composes movements of all particles into a single model accurately depicting different alterations in the whole sample. In contrast, other refinement software such as EMAN2 and emCLARITY optimize the movement of single, or patches of particles. Additionally, M combines particle and sample movement optimization along with refinement of CTF parameters in a single process. Thus it optimizes various stages of tilt series processing simultaneously whereas emCLARITY, for example, has different processes for refining global and particle alignment [11,15,18].

Following I pursued the reconstruction of the ribosome bound, OST & TRAP translocon. OST & TRAP are submembrane complexes of the ER lumen. The lumen is a difficult environment to obtain structural information on proteins as most structural techniques would require the isolation of the complex with the use of detergents. However, such manipulation would also abolish transient interactions present as well as affect the structures present in the isolate. CET nevertheless is able to capture such complexes in near native constitution as imaging occurs in situ[38,39]

Focusing on the luminal compartment of the translocon complex I obtained a TRAP and OST bound ribosome reconstruction of 9.5 Å global resolution and 12.8 Å resolution for the OST and TRAP complexes. In a previous study PyTOM was able to produce a ~9 Å average of the ribosome in which the local OST and TRAP resolution were 12-15 Å[14] . However as mentioned previously the utilized pool of particles of that research was more than tenfold compared to the 1664 particles used here.

Thus I can conclude that Warp-M's processing and multi-particle refinement allows high resolution reconstructions using a small fraction of the data necessary with approaches previously employed. The increase in resolution and decrease in required particles will allow the characterization of structures and interactions in a more realistic timeframe. This upsurge

in performance of tilt series processing will accelerate the characterization of essential and previously un-reachable biomolecular machineries, such as the translocon complex, *in-situ*.

References

1. Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Current Opinion in Structural Biology* (2019) doi:10.1016/j.sbi.2019.05.021.
2. Fischer, T. D., Dash, P. K., Liu, J. & Waxham, M. N. Morphology of Mitochondria in Spatially Restricted Axons Revealed by Cryo-Electron Tomography. *PLoS Biol.* (2018) doi:10.1101/273052.
3. Pfeffer, S. *et al.* Dissecting the molecular organization of the translocon-associated protein comp1. Pfeffer, S. et al. Dissecting the molecular organization of the translocon-associated protein complex. Nat. Commun. (2017) doi:10.1038/ncomms14516.lex. *Nat. Commun.* (2017) doi:10.1038/ncomms14516.
4. Schur, F. K. M. *et al.* Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* (2015) doi:10.1038/nature13838.
5. Jonić, S., Sorzano, C. O. S. & Boisset, N. Comparison of single-particle analysis and electron tomography approaches: An overview. *Journal of Microscopy* (2008) doi:10.1111/j.1365-2818.2008.02119.x.
6. Nakane, T. *et al.* Single-particle cryo-EM at atomic resolution. *Nature* (2020) doi:10.1038/s41586-020-2829-0.
7. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* (2020) doi:10.1038/s41586-020-2833-4.
8. Verguet, A., Messaoudi, C., Sorzano, C. O. S. & Marco, S. Alignment of Tilt Series. in 183–207 (2018). doi:10.1007/978-3-319-68997-5_7.
9. Radermacher, M. Weighted Back-Projection Methods. in *Electron Tomography* 91–115 (Springer US, 1992). doi:10.1007/978-1-4757-2163-8_5.
10. Pfeffer, S. & Förster, F. Structural Biology in Situ Using Cryo-Electron Subtomogram Analysis. in 237–259 (2018). doi:10.1007/978-3-319-68997-5_9.
11. Chen, M. *et al.* A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat. Methods* (2019) doi:10.1038/s41592-019-0591-8.
12. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* (2018) doi:10.7554/eLife.42166.
13. Hrabe, T. *et al.* PyTom: A python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J. Struct. Biol.* (2012) doi:10.1016/j.jsb.2011.12.003.
14. Pfeffer, S. *et al.* Structure of the native Sec61 protein-conducting channel. *Nat. Commun.* (2015) doi:10.1038/ncomms9403.
15. Himes, B. A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat. Methods* (2018) doi:10.1038/s41592-018-0167-z.
16. Pyle, E. & Zanetti, G. Current data processing strategies for cryo-electron tomography and subtomogram averaging. *Biochemical Journal* (2021) doi:10.1042/BCJ20200715.
17. Tegunov, D. High-resolution in situ imaging of biological samples with Warp and M . *Acta Crystallogr. Sect. A Found. Adv.* (2020) doi:10.1107/s0108767320097743.
18. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle1. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. Nat. Methods (2021) doi:10.1038/s41592-020-01054-7. cryo-EM refinement with M. *Nat. Methods* (2021) doi:10.1038/s41592-020-01054-7.
19. Pfeffer, S. *et al.* Structure and 3D arrangement of endoplasmic reticulum membrane-associated ribosomes. *Structure* (2012) doi:10.1016/j.str.2012.06.010.

20. Jomaa, A. *et al.* Structure of the quaternary complex between SRP, SR, and translocon bound to the translating ribosome. *Nat. Commun.* (2017) doi:10.1038/ncomms15470.

21. Braunger, K. *et al.* Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science (80-.).* (2018) doi:10.1126/science.aar7899.

22. Gemmer, M. & Förster, F. A clearer picture of the ER translocon complex. *Journal of Cell Science* (2020) doi:10.1242/jcs.231340.

23. Hurt, C. M., Ho, V. K. & Angelotti, T. Systematic and Quantitative Analysis of G Protein-Coupled Receptor Trafficking Motifs. in 171–187 (2013). doi:10.1016/B978-0-12-391862-8.00009-0.

24. Ménétret, J.-F. *et al.* Single Copies of Sec61 and TRAP Associate with a Nontranslating Mammalian Ribosome. *Structure* **16**, 1126–1137 (2008).

25. Fons, R. D., Bogert, B. A. & Hegde, R. S. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* (2003) doi:10.1083/jcb.200210095.

26. Nguyen, D. *et al.* Proteomics reveals signal peptide features determining the client specificity in human TRAP-dependent ER protein import. *Nat. Commun.* (2018) doi:10.1038/s41467-018-06188-z.

27. Razi, A., Britton, R. A. & Ortega, J. The impact of recent improvements in cryo-electron microscopy technology on the understanding of bacterial ribosome assembly. *Nucleic Acids Research* (2017) doi:10.1093/nar/gkw1231.

28. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* (2019) doi:10.1038/s41592-019-0580-y.

29. Walter, P. & Blobel, G. [6] Preparation of microsomal membranes for cotranslational protein translocation. in 84–93 (1983). doi:10.1016/S0076-6879(83)96010-X.

30. Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

31. Mastronarde, D. N. & Held, S. R. Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* (2017) doi:10.1016/j.jsb.2016.07.011.

32. Voorhees, R. M., Fernández, I. S., Scheres, S. H. W. & Hegde, R. S. Structure of the Mammalian Ribosome-Sec61 Complex to 3.4 Å Resolution. *Cell* **157**, 1632–1643 (2014).

33. Cong, Y. & Ludtke, S. J. Single Particle Analysis at High Resolution. in 211–235 (2010). doi:10.1016/S0076-6879(10)82009-9.

34. Jiang, L., Liu, Z., Georgieva, D., Kuil, M. E. & Abrahams, J. P. A novel approximation method of CTF amplitude correction for 3D single particle reconstruction. *Ultramicroscopy* **110**, 350–358 (2010).

35. Bharat, T. A. M. & Scheres, S. H. W. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat. Protoc.* **11**, 2054–2065 (2016).

36. Ni, T. *et al.* High-resolution in situ structure determination by cryo-electron tomography and subtomogram averaging using emClarity. *Nat. Protoc.* (2022) doi:10.1038/s41596-021-00648-5.

37. Eisenstein, F., Danev, R. & Pilhofer, M. Improved applicability and robustness of fast cryo-electron tomography data acquisition. *J. Struct. Biol.* **208**, 107–114 (2019).

38. Lučić, V., Rigort, A. & Baumeister, W. Cryo-electron tomography: The challenge of doing structural biology in situ. *J. Cell Biol.* **202**, 407–419 (2013).

39. McGilvray, P. T. *et al.* An ER translocon for multi-pass membrane protein biogenesis. *Elife* (2020) doi:10.7554/ELIFE.56889.

Supplementary Information
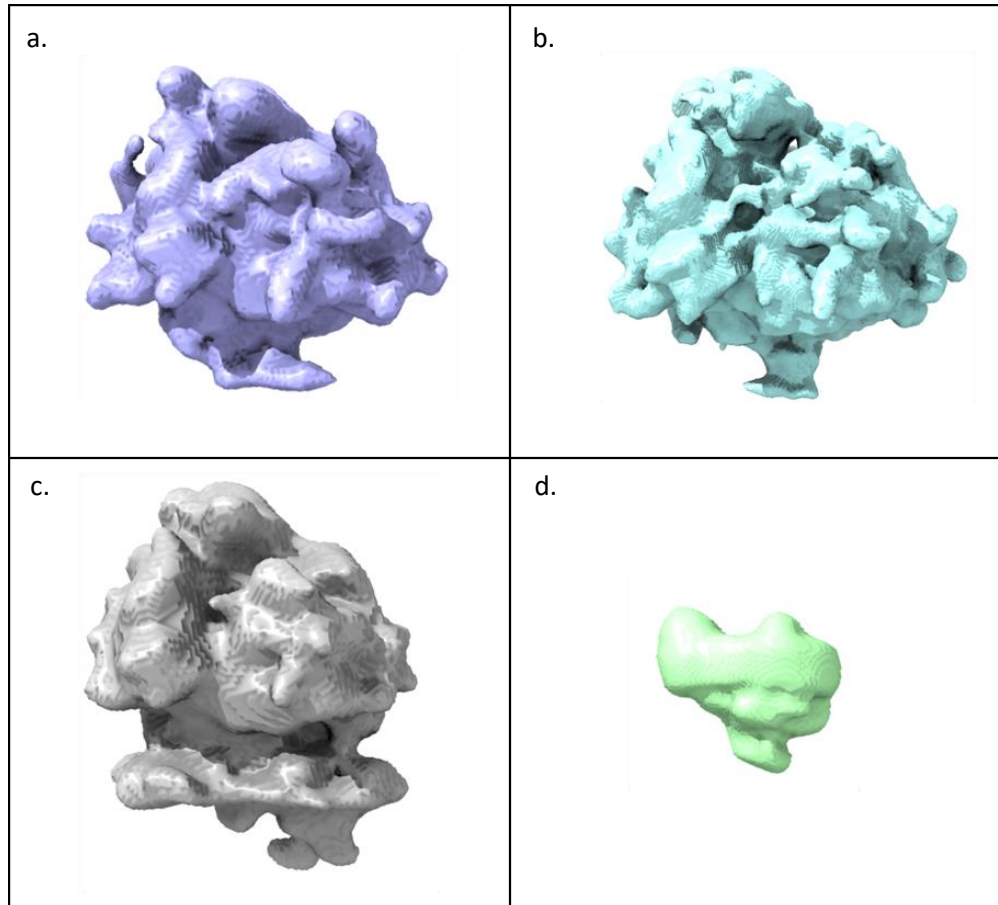
**Figure S1 Tight masks for FSC in PyTOM**

a. Mask used for the FSCs of; PyTOM thin and thick reconstructions and Warp-M thin and thick reconstructions utilizing 2317 particles each, selected upon classification. Mask used for FSC of Warp-M reconstructions utilizing 7323 particles of the thin dataset; e.g. the 5.6 Å reconstruction obtained using the full tilt range as well as the reconstructions obtained with 30° and 21° tilt range restriction. c. Mask used for FSC of the whole ribosome–OST, translocon reconstruction obtained with Warp-M using 1664 particles. d. Mask for the calculation of local TRAP and OST resolution in the previous reconstruction. All masks and corresponding FSCs were performed in PyTOM.
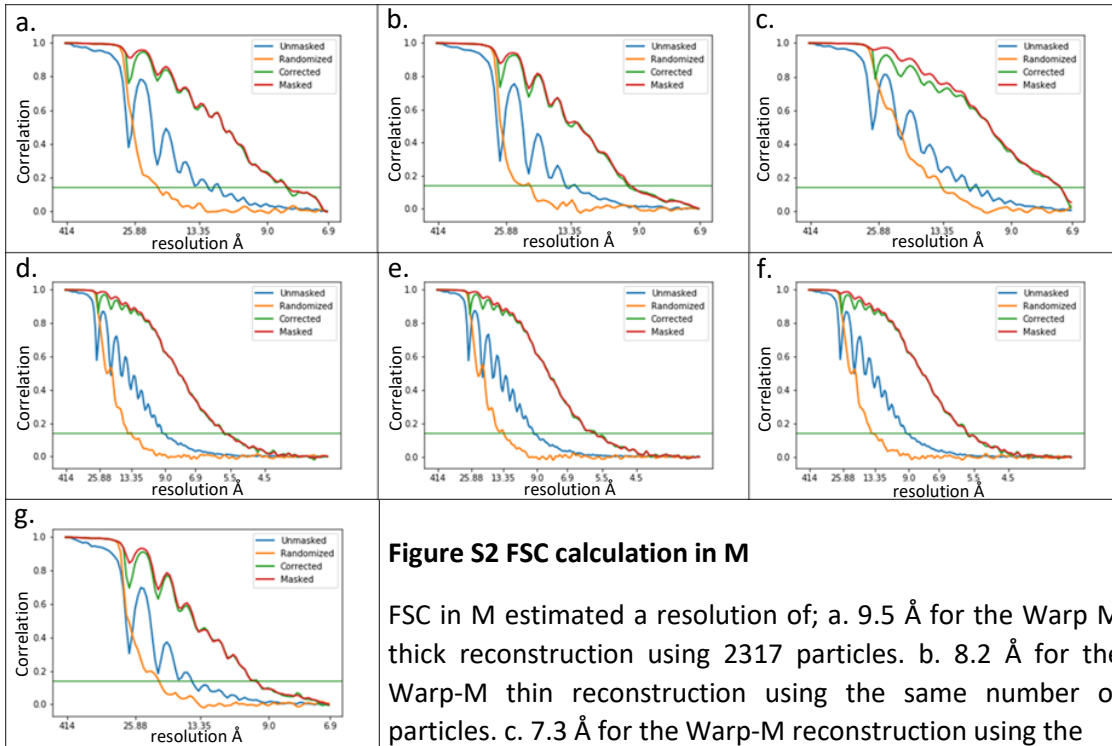
Figure S2 FSC calculation in M

FSC in M estimated a resolution of; a. 9.5 Å for the Warp M thick reconstruction using 2317 particles. b. 8.2 Å for the Warp-M thin reconstruction using the same number of particles. c. 7.3 Å for the Warp-M reconstruction using the 3055 particles obtained upon classification of the Warp-M thin dataset (prior to equalization). d,e,f. 5.8 Å for the Warp-M reconstructions utilizing 7323 particles of the thin dataset. (d. full tilt range, e. 30° tilt restriction f. 21° tilt restriction. g. 9.6 Å global resolution for the whole ribosome-translocon reconstruction obtained with Warp-M using 1664 particles.
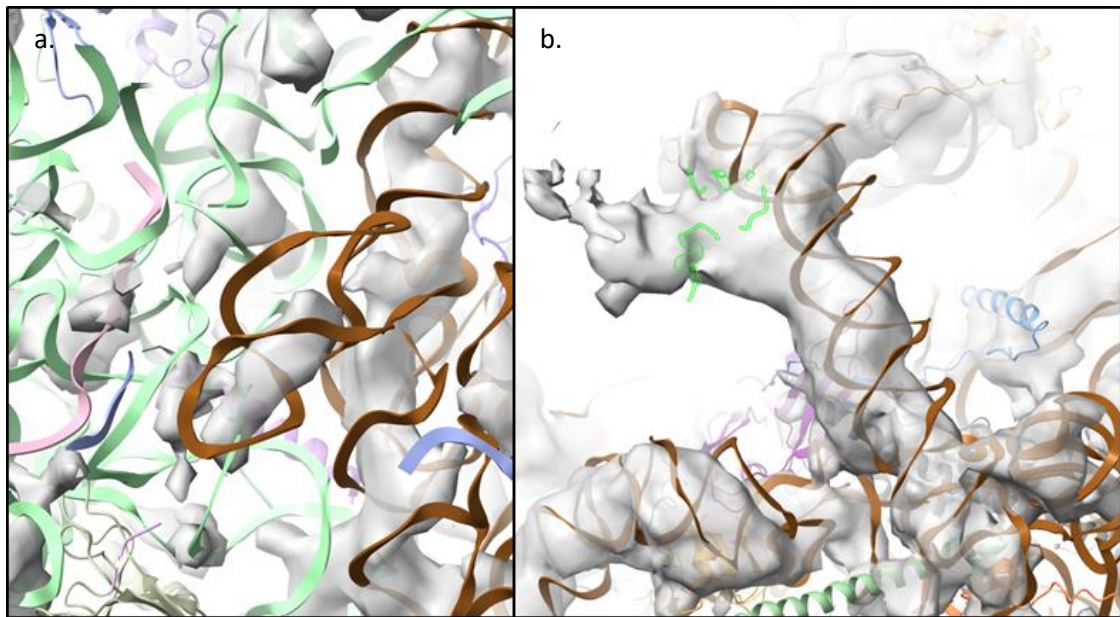
Figure S3 Peripheral vs Central elements in PyTOM reconstructions

a.28S rRNA (orange) in the center of the ribosome reconstructed using the PyTOM thin dataset of 2317 particles. In the volume I have fitted an 80S ribosome molecular model previously used for the 5.6 Å Warp-M reconstruction. b. 28S rRNA in the periphery of the same ribosome reconstruction viewed at the same contour in Chimera. Structural elements in the periphery of PyTOM's reconstructions were more intense than corresponding central
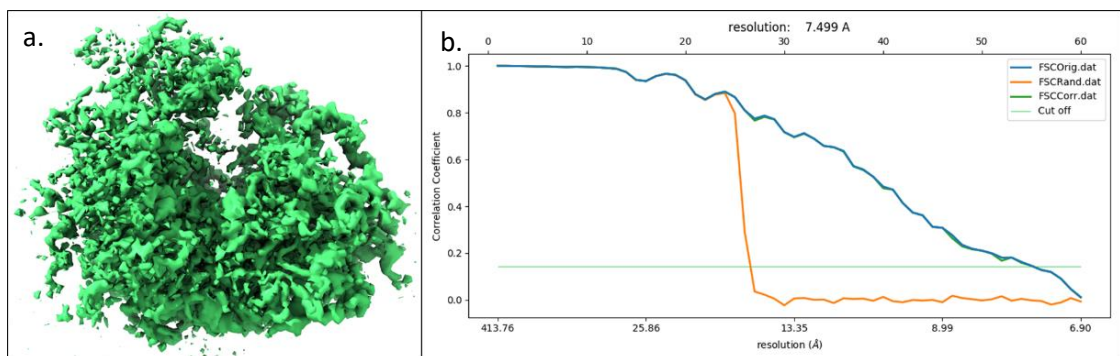

Figure S4 Warp-M Reconstruction at 7.5 Å from 3055 particles.

a,b. Front ribosome view from the Warp–M reconstruction obtained from refining the classified thin set of 3055 particles. b. FSC estimation of resolution at 7.5 Å for the previous reconstruction.
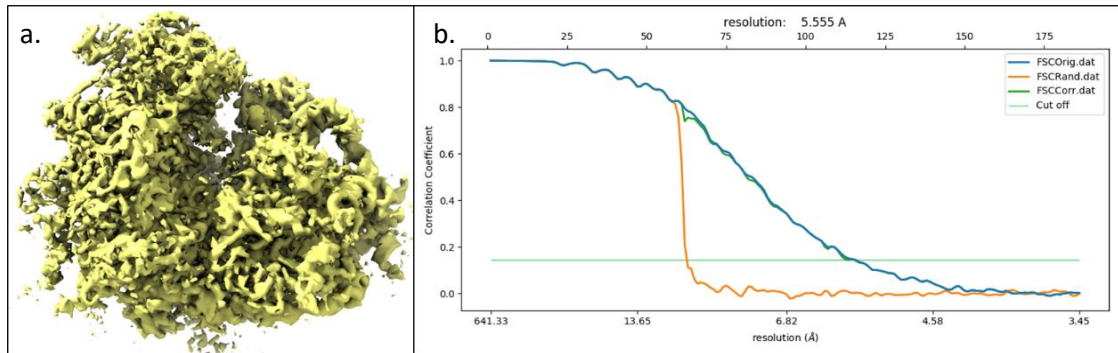
Figure S5 Warp-M 8288 particles processing.

a.Warp-M reconstruction at 7.1 Å obtained using the unclassified dataset of 8288 particles. b. FSC determination of resolution at 5.7 Å for a reconstruction obtained using 7323 particles of the Warp-M thin set.
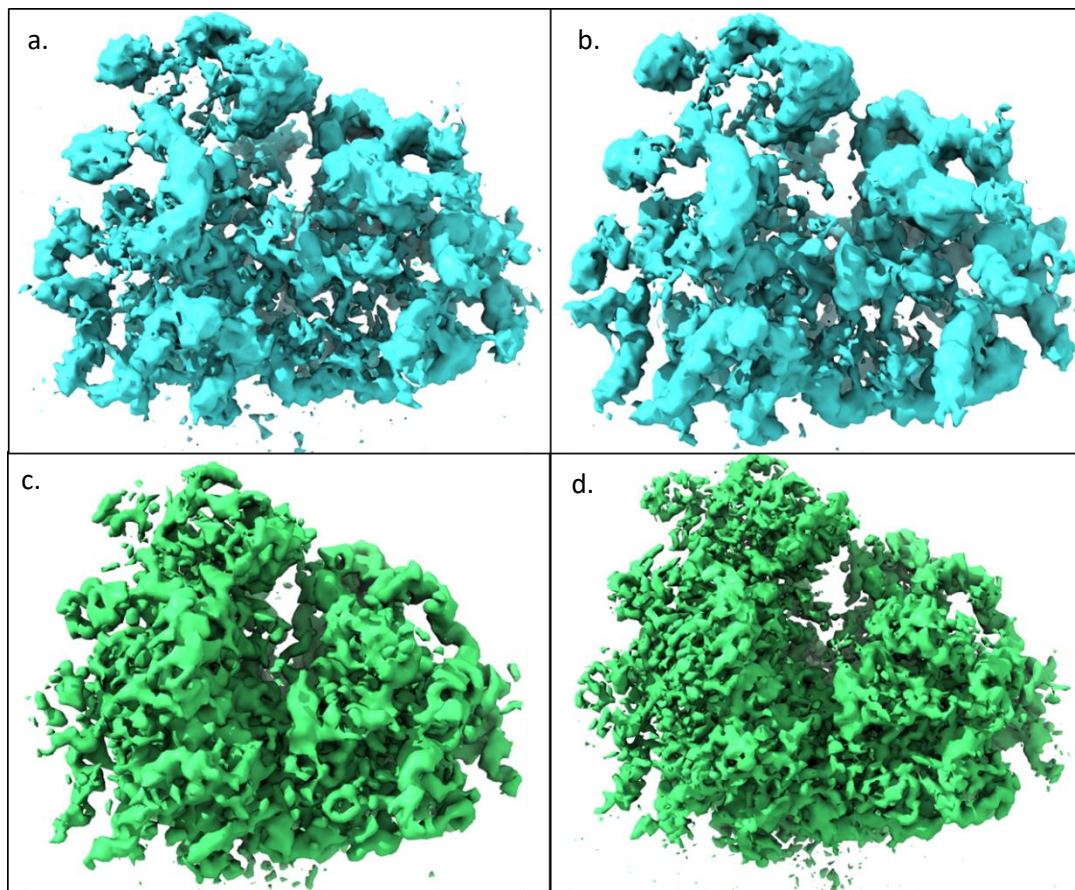
Figure S6 PyTOM vs Warp-M unclassified.

a-d. Reconstructions utilizing the full particle set after template matching (3279 particles). a. PyTOM thick set reconstruction at 13.8 Å b. PyTOM thin set reconstruction at 13.4 Å. c Warp-M thick reconstruction at 10 Å. d. Warp-M thin reconstruction at 7.5 Å.
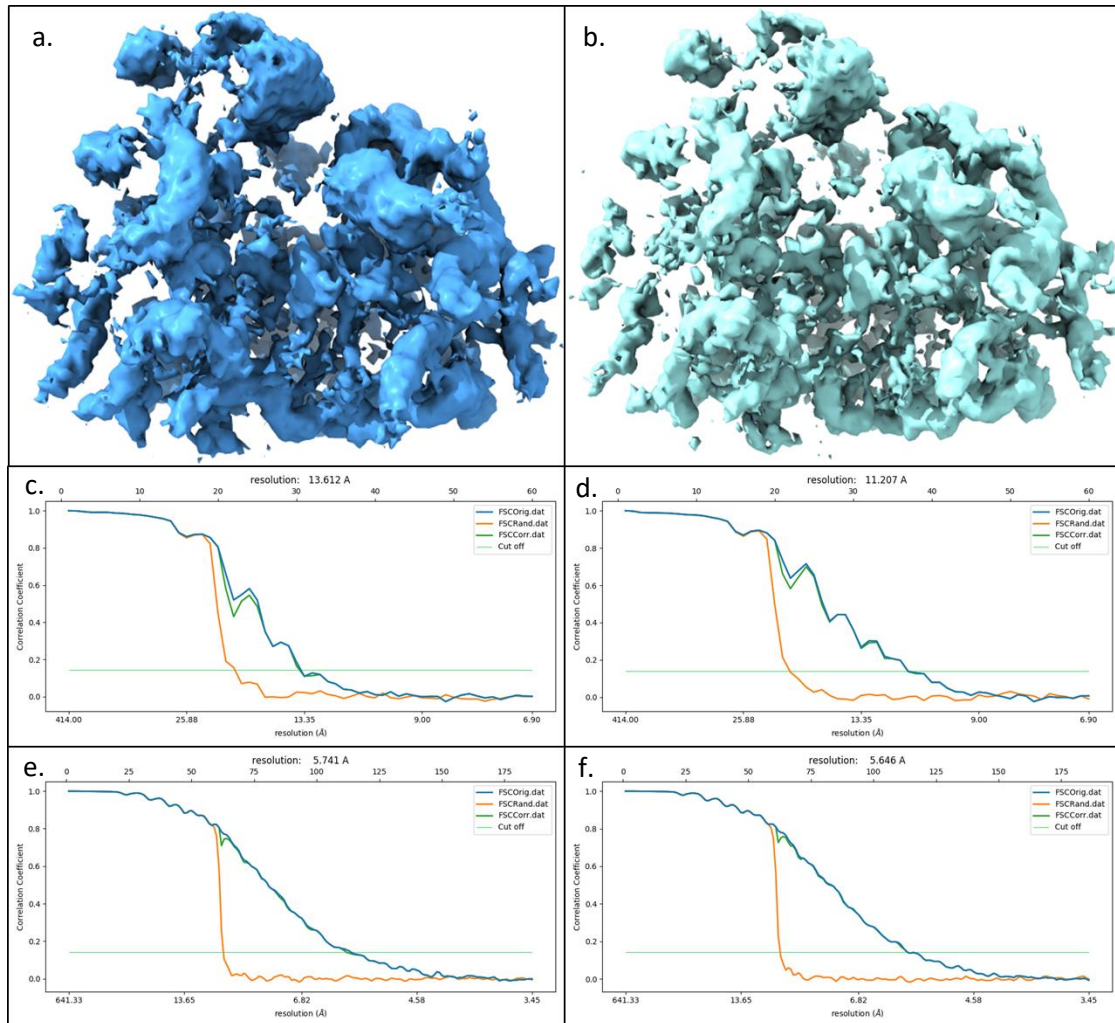
Figure S7 Tilt Range restriction

Electron beam radiation during imaging damages the sample. This effect is more prevalent at higher tilts within a tilt series where the radiation is accumulated from previous images. Therefore, restriction of the utilized tilt range can increase resolution of reconstructions by excluding radiation damaged particle tilts. a,b, Tilt range restriction significantly increases the resolution of the sub tomogram average in PyTOM. a. Sub tomogram average from PyTOM, utilizing 3660 particles of the thin set, at the full tilt range (-51-51°) determined at 13.6 Å (c) b. Sub tomogram average from PyTOM using the previous set but with a tilt range restriction at -21-21°. d. Resolution was significantly increased with -21-21° restriction and was determined at 11.2 Å. e,f Tilt range restriction does not affect resolution in Warp-M this could be due to M's compensation of radiation damage by weighting particle tilts by an exposure-dependent function[15]. e,f M refinement using the thin 7323 particle set with a tilt range restriction of -30-30° and -21-21° accordingly both reconstructions were determined approximately 6 Å as well as in the full tilt range (-51-51°) reconstruction (Fig S5b).