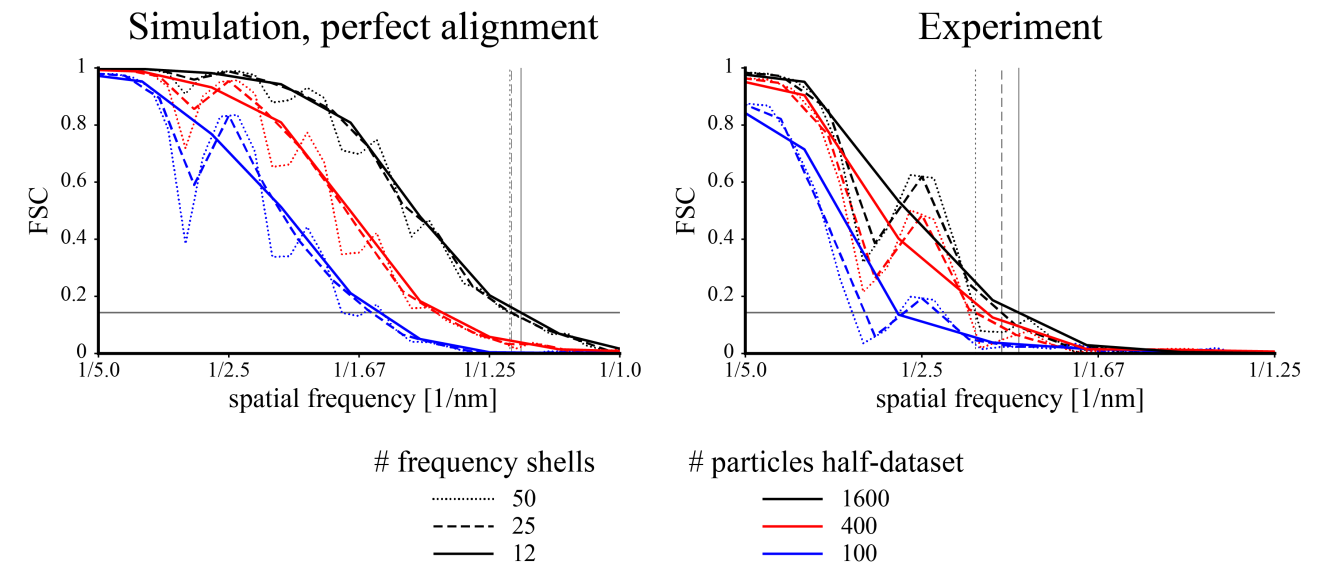
In Fig. 4. we observe strong oscillations in the FSC due to the zero-crossings of the CTF. Here we demonstrate that the visibility of these dips depends on the width of the frequency shells used for calculating the FSC. Consequently, this influences the found resolution by thresholding the FSC curve.



**Fig. S1.** Fourier Shell Correlation (FSC) for different numbers of frequency shells and numbers of particles for one simulated dataset and one experimental dataset. Both datasets are processed using TCTF correction using a varying defocus estimate. The simulated data has a perfect tilt-series alignment whereas the experimental data has a non-perfect alignment.

Dips in the FSC in subtomogram averaging occur if the particles are recorded with a similar defocus. Even though each tilt-series exhibits a certain variation in defocus and each tilted projection contains macromolecules imaged with a different defocus, this defocus spread does not guarantee that the SNR around the zero-crossings is similar to spatial frequencies without zero-crossings. Only when a large number of particles is averaged or when tomograms are intentionally acquired using different defoci, can these dips be avoided.

Nevertheless, Fig. S1 shows that when the number of frequency shells used to calculate the FSC is too low it is impossible to judge whether these dips disappeared due to under sampling of the FSC or that the SNR was indeed high enough. Furthermore, choosing too few frequency shells can lead to an artificial increase in reported resolution.

For the aforementioned reasons we calculate the FSC using 50 frequency shells, making the oscillations in the FSC more apparent. This number of frequency shells results in a frequency shell width that is close to the voxelsize of the Fourier transformed subtomograms in this study (size of the subtomograms was 1283). Fig. S1 shows that due to these oscillations it is possible that the FSC crosses the threshold multiple times. We use the spatial frequency where the FSC crosses the threshold for the first time to determine the resolution.

Different choices in terms of number of frequency shells, type of threshold and whether to use the first or last crossing of the threshold lead to different reported resolutions. The focus of this study is to quantify the influence of the different processing steps that lead to a reconstruction. Therefore, we primarily use FSC and the derived resolution to make a relative comparison between experiments and different simulations. In that sense, the choice of resolution criteria and the exact computation of the FSC does not influence our findings.