

Current cryo-EM data

The *Bacillus halodurans* Cas4₂-Cas1₄-Cas2₂ complex is a hetero-octomeric protein complex bound to a DNA substrate, for a total size of ~200 kDa. A ~2500 image data set has been collected for a Cas4-Cas1-Cas2 complex bound to a 24-bp duplex DNA with 15 bp 3'-overhangs. Images were collected on a Glacios with K3 detector at the Iowa State Cryo-EM Facility, using a nominal defocus range of -1.2 to -2.5 μM at 0.45 $\text{\AA}/\text{px}$ and a total dose of 48.8 $\text{e}/\text{\AA}^2$ over 45 frames (Figure 1).

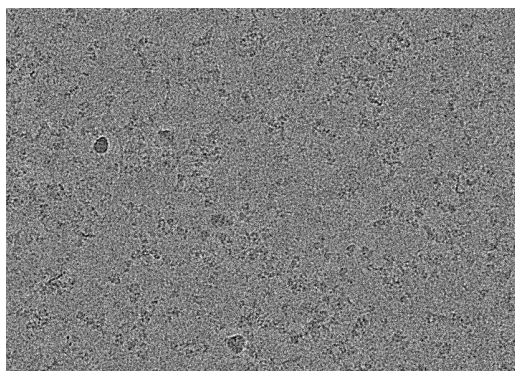


Figure 1: K3 image of Cas4-1-2 sample with defocus of 2 μM .

Data processing has been performed in both cryosparc and Relion. CTF estimates indicate high

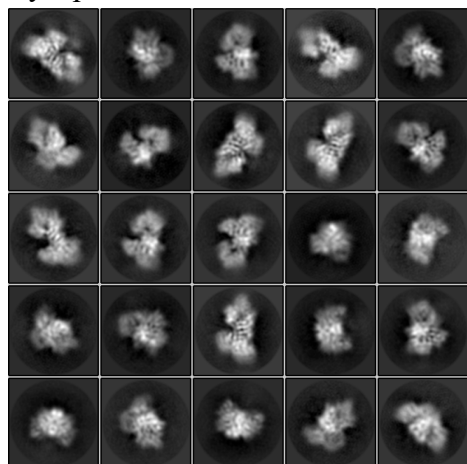


Figure 2: Representative 2D class averages from cryosparc.

image resolution, ranging from 2-5 \AA . Particle picking was performed using a combination of blob picking followed by template picking, yielding ~350,000 particles. 2D class averages revealed a wide distribution of views, with little evidence of preferred orientations (Figure 2).

3D classification and refinement has thus far yielded a 4.6 \AA map (Figure 3, middle, right). Within this structural model, we are most interested in the interactions between Cas4 and the other protein subunits, as well as how Cas4 recognizes a sequence within the 3'-single-stranded DNA overhangs. We have fit homology models of each of the protein subunits into the current map. The current map resolution enables unambiguous assignment of Cas4 orientation within the complex and we can see

clear density for single-stranded DNA interacting with Cas4. However, limited homology between the *B. halodurans* Cas4 sequence and the closest homolog for which a crystal structure is available will require building a Cas4 structural model based on cryo-EM density (Figure 3, left). The current resolution is not sufficient to begin structure building (Figure 3).

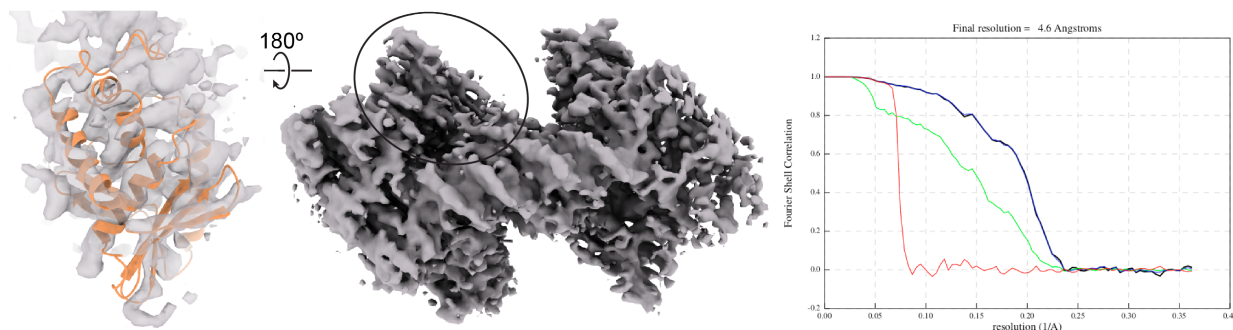


Figure 3: Current map based on data collected on Glacios + K3. The 4.6 \AA map is shown in the middle. The FSC curve for the final masked map is shown in blue on the right. Density for one of the Cas4 subunits is circled in the map. A close-up of this density with a homology model for the BhCas4 sequence fit in the density is shown on the left (based on PDB 4R5Q). The homology model diverges significantly from the density, indicating that a structural model will have to be built de novo, requiring a higher