	Experiment title: Structure analysis of a dimeric cryptochrome by cryo-EM	**Experiment number**: MX2337
Beamline: CM01	**Date of experiment**: from: 21.10.2020 to: 23.10.2020	**Date of report**: 26.08.2021
Shifts: 6	**Local contact(s)**: KANDIAH Eaazhisai	*Received at ESRF:*

Names and affiliations of applicants (* indicates experimentalists):

Dr. Elmar Behrmann, University of Cologne

Dr. Eva Wolf, University Mainz

Dr. Gayathri Jeyasankar, University of Cologne *

Report:

The beam-line allocation of Wednesday 21st of October till Friday 23rd of October 2020 yielded a total of 10,779 exposure frame-movies. Initially, we tried to follow the processing pipeline we had established for the data collected in-house. This yielded a total of ~2.4 million particles from 10,759 high quality summed and dose-weighted micrographs. With the compute infrastructure available to us in late 2020, we had to divide this particle set into twenty equally sized groups and process each subset of particles individually in Relion3.1 at the stage of 2D classification. 2D classification confirmed the high accuracy of crYOLO used for identifying particle coordinates, as ~2 million remained in the high-quality dataset. However, 2D classes predominantly showed top views of the dimer with as little as 10% front views and no side views present. Resulting 3D reconstructions did not reach high resolutions and were strongly anisotropic showing elongated features in the z direction.

At this point, we relocated our laboratory to a different institute in another city – which together with the ongoing pandemic resulted in a forced break that lasted almost half a year. Hoping that we would have benn able to continue working on the project, and hoping that – given the high amount of particle images – we refrained from submitting a report with the aforementioned results.

Recently, we could finally install a HPC GPU node at our new institute (equipped with a dual A100 GPU) allowing us to re-process the ESRF dataset from experiment MX2337. In order to avoid potential bias due to the previously used selection and classification strategy we employed cryoSPARC this time for the 3D reconstruction. Limiting total motion of the movies to 20 pixels, maximum in-frame motion to 2 pixels, and the CTF fit to better than 7.5 Å resulted in a subset of 8,619 micrographs that were further processed. Using the blob picker implemented we created an initial selection of putative particles that was then subjected to

unsupervised 2D classification into 50 classes. 9 of those showed clear top and front views of the protein, though regretfully no parameters could be found that unveiled side views of the dimer. Using these classes as templates, close to 4 million putative particles were identified in the dataset, of which ~1.5 million were part of 2D classes that showed distinct protein features (Figure 1).

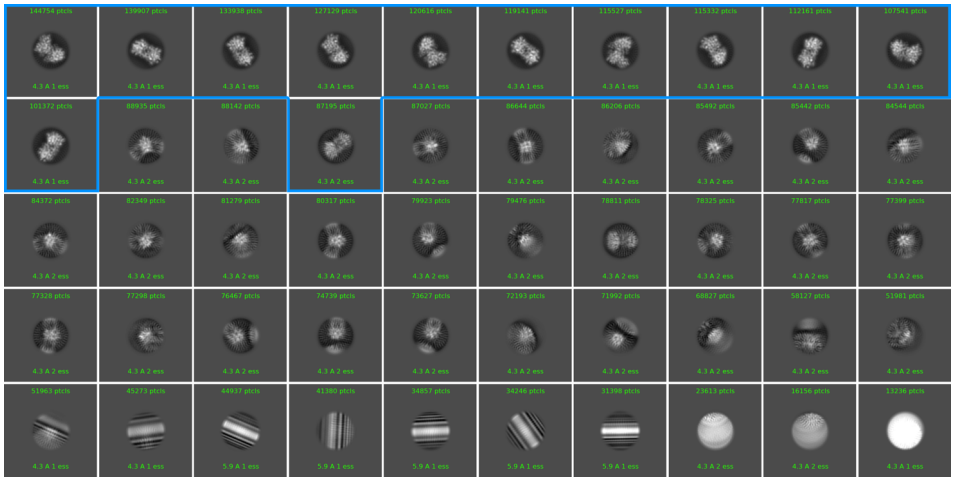

Figure 1: Particles are either oriented in front or top view direction. Different from our preceding approach, we could observe both front and top views in the dataset. However, side views were not observed in the 2D class averages. Classes marked with the blue outline were selected for further processing steps.

To increase processing speed, we re-extracted the particles in the selected classes with re-centering and binning from the original 0.65 Å/px sampling rate to a sampling rate of 1.06 Å/px. As expected, an initial 3D model calculated ab-initio from these particles showed the same dimer as observed before. Heterogeneous 3D refinement specifying 5 classes using the same initial model as the starting point revealed a single garbage class that accounts for 222,316 particles and contains a significant amount of clearly artificial particles. Recombining the remaining 4 classes and using non-uniform 3D refinement yielded a 3D reconstruction at a nominal resolution of 3.5 Å (Figure 2).

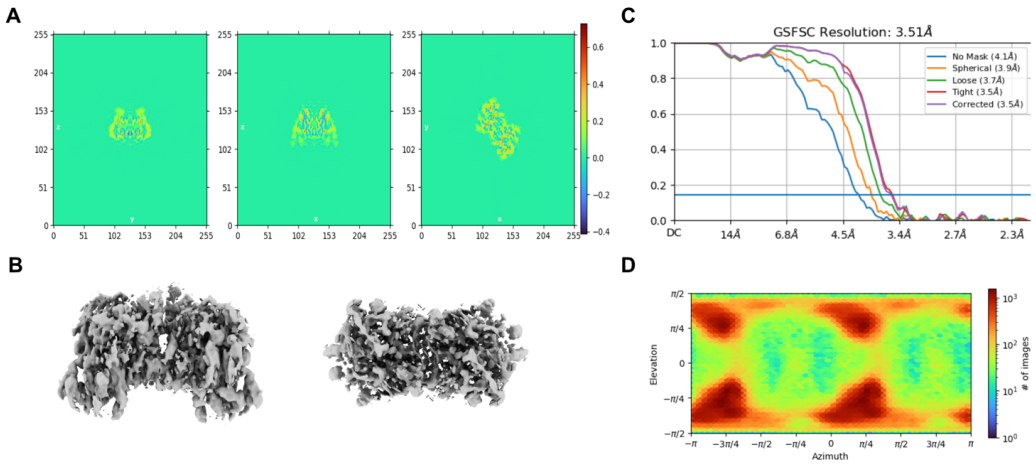

Figure 2: The resulting structure is strongly anisotropic. (A) Perpendicular slices through the volume and **(B)** surface representation in front and top view orientation of one of the most promising results in 3D refinement. Despite the presence of two perpendicular view directions in the dataset, the resulting structure is strongly anisotropic, although **(C)** the overall resolution is good. **(D)** A plot of the view directions reveals a severe bias towards a single azimuth (the second peak is due to the C2 symmetry) and a single elevation band.

Again, the resulting structure was strongly anisotropic, showing side-chain features in the x/y-plane but strongly distorted in the z-plane. The viewing distribution shows particles distributed either at a narrow band of the 'poles' of the plot or together with a narrow populated region at an azimuth of π/2. While the powerful GPU architecture allowed us to test various approaches to further investigate the data (i.e. omitting the 2D classification and sorting directly in 3D), we did not find any strategy that overcame the strong anisotropy.

Hence, we are preparing cryo-samples employing UltrAUfoil grids that will hopefully enable us to collect a dataset at a tilt of 40 degrees to overcome the limited views due to the preferential orientation of the protein.