Simulating cryo-EM maps using a Generative Adversarial Network (cryo-GAN)

The ability to perform a realistic simulation of a cryo-EM density map (at appropriate resolution) is crucial for the development and analysis of advanced cryo-EM based macromolecular structure modeling algorithms. In particular, such simulation is required for the recent machine learning based algorithms, which have to be trained on very large databases of "realistic" data. Such large databases are currently unavailable, and thus have to be simulated.

The currently used simulation methods, such as pdb2mrc from the Chimera package, usually perform averaging on atomic resolution structures from the PDB. It is well known, that the resulting simulative structures (at the appropriate resolution) are significantly more similar to the atomic resolution structures than experimental cryo-EM maps at the same resolution, and thus do not represent the same degree of difficulty for the tested algorithms.

**Cryo-GAN** exploits recent deep learning techniques for creating cryo-EM maps which are indistinguishable from experimental ones. It uses the VAE-GAN architecture which is a compound Generative Adversarial Network (GAN) with Variation AutoEncoder (VAE). Both are proven deep learning techniques for generating 2D and 3D images.

The resulting synthetic maps have a high resemblance to the experimental maps. An independent discriminator was trained to distinguish between experimental and synthetic maps. In our computational experiments the discriminator marks all voxels of a map generated by existing simulation (pdb2mrc) as non-real. Less than half of the voxels of a map generated by Cryo-GAN are marked as non-real, meaning the map is indistinguishable from an experimental one.