**AI/ML-based peak and hitfinder project for CXLS/CXFEL**

**Objective:** Use established AI/ML algorithms to find crystal diffraction patterns on the Eiger detector, that is continuously trained on real data and can find peaks/hits taking into account experimental factors (detector distance and X-ray energy) in addition to image data.

**Background:** Peak and hitfinding is a difficult task in serial crystallography, since it is influenced by experiment specific factors such as X-ray wavelength, sample to detector distance (flashlight analogy) as well as the studied crystal system and protein. The objective of this particular project is therefore to develop an AI/ML-based hitfinder, that can take these parameters into account when separating out real crystal diffraction patterns from noise, salt diffraction (few intense spots), of PEG diffraction (spotty rings).

**Goal 1:** Train on simulated data with 3 different detector distances, but keeping the X-ray energy and crystal system constant.

A screen shot of a computer

Description automatically generated

Simulate 10000 patterns with the parameters above.

Protein: Proteinase K, PDB entry 1IC6

Replace “PYP.cell” with the correct unit cell file described in the PDB (roughly tetragonal lattice, unique axis c, constants 67,67,107; 90, 90, 90).

Pointgroup should accordingly be 4/mmm

In pattern\_sim\_submit

NAME=$1

GEOM=Eiger4M.geom

CRYSTAL=PYP.cell

NUMBER\_OF\_PATTERNS=10000

INPUT= 1IC6.pdb.hkl

POINTGROUP=4/mmm

CRYSTAL\_SIZE\_MIN=10000

CRYSTAL\_SIZE\_MAX=10000

SPECTRUM=tophat

SAMPLING=7

BANDWIDTH=0.01

PHOTON\_ENERGY=7000

N\_PHOTONS=3e8

BEAM\_RADIUS=5e-6

TASKS=$2

PARTITION=$3

QOS=$4

**Step 1:** Simulate 10000 diffraction patterns of Proteinase K (pdb entry 1ic6) at **7keV** on the Eiger detector with **detector distances 0.1m, 0.2m, 0.3m**; (Other parameters the same as for )

**Step 2:** Superimpose diffuse water scattering using the same simulation parameters. (e.g. using reborn).

**Step3:** Train AI/ML hitfinder on simulated data providing both the image data as well as the detector distance as training paramaters. Test the AI/ML hifinder on a newly simulated data with detector distance 0.15m.

**Goal 2:** Repeat steps 1-3 with 3 different X-ray energies (6.5 keV, 7.5 keV, 8 keV), but keeping the detector distance and crystal system constant (same as for goal 1, with detector distance 0.2m).

**Goal 3:** Repeat steps 1-3, varying the crystal system. (0.2m, 7keV, pdb entries 1vds, 6yqg and 5hdd).