Crystallization of ferredoxin from *Arthrospira platensis*

Cryo-crystallography and SFX data processing

Natalia Lyubaykina, Moscow Institute of Physics and Technology, Russia

Supervisors: Iosifina Sarrou, Oleksandr Yefanov

September 4, 2019

Abstract

Cryo-crystallography and SFX (serial femtosecond X-ray crystallography) crystallography are powerful complementary techniques for the determination of biological molecules' structures. The crucial point of structure determination is the stability of molecule during data collection. Soluble protein ferredoxin containing iron-sulfur cluster was investigated to establish, if cryo-crystallography and SFX gave different results for the oxidation state of iron to estimate the radiation damage in cryo-crystallography approach. The crystals of ferredoxin were obtained for both crystallography experiments. Data processing was conducted for several other proteins.

Contents

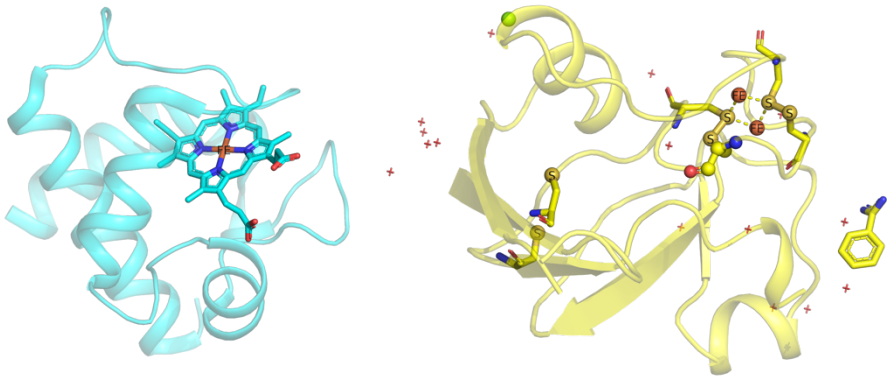
1. Introduction

The recent advances in X-ray crystallography led to the great increase of amount of structural data important for the understanding of fundamental biological processes. The structures of large membrane complexes photosystem I (PSI) and photosystem II playing the key role in photosynthesis are already available and enable building the model of photosynthesis[1]. For comprehensive model, it is important to obtain structural data of involved proteins from one organism.

This work was focused on ferredoxin. PSI, ferredoxin, cytochrome (Figure 1) and phycocyanin were obtained from natural source photosynthetic cyanobacteria *Arthrospira platensis.* PSI drives the light-induced reduction of ferredoxin on the stromal side of the thylakoid membrane together with the oxidation of plastocyanin or cytochrome c6 its lumenal side[2].

Ferredoxin is small protein containing iron-sulfur cluster, where iron can change oxidation state, accepting or discharging electrons. Cytochrome c6 is small protein containing covalently bond hem molecule, which utilizes iron for electron transport. The oxidation state of iron corresponds with functional state of these proteins, therefore, it is important to save it during diffraction data collection[3]. The iron oxidation state might probably change during cryo-crystallography experiments, while serial crystallography probably detects the iron oxidation state correctly. The aim of this work was to compare the results of both approaches for ferredoxin, learn experimental crystallization techniques and learn to process cryo-crystallography and SFX data.

Figure 1. Left – structure of cytochrome c6, PDB ID 1F1F. Right – structure of ferredoxin obtained by Iosifina Sarrou.



1.1. Cryo-Crystallography

The crucial point for X-ray crystallography is obtaining protein crystal of appropriate quality and size (about 100µm). Crystal is essential, because the diffraction of single molecule is too weak to be detected. The molecules in crystal are periodically arranged and identically oriented, it enable signal enhancing for definite orientations, its detecting and interpretation.

If the crystal was obtained, it places to the intensive monochromatic continues X-ray beam (wavelength is usually about 1 Å) and diffracted data is collected for different orientations of crystal regarding the beam. Crystal has to be exposed by X-ray quite long, which lead to the radiation damage – gradual degradation of protein molecules and decreasing of the amount of scattering protein molecules during the experiment. The cryo-cooling is used to reduce

the radiation damage and enable the collection of enough data (full dataset) for solving protein structure before complete degradation of molecules in crystal.

1.2. Serial Crystallography

Recently developed method of serial crystallography has already demonstrated its power for protein structure determination[4]. If the size of protein crystal is not big enough, the radiation damage does not allow to collect full dataset in classical crystallography approach even with cryo-cooling. Growing big crystals is a challenging and time-consuming especially for large complexes.
Serial crystallography uses significantly smaller crystals (from hundreds of nanometers to a few micrometers). This technique uses the bunch of crystals instead single one. The experiment is organized in different ways: the intense pulsed X-ray (produced by free electron lasers) falls on crystal stream, or the conventional synchrotron radiation is used to scan crystals fixed on some substrate (silicon crystal, special tape). As the result the great number of "snapshots" of several crystals in different orientations are collected and further proceeded together to get structural information.
Here in should be note, that serial approach might be realized with classical beam (continues one) in "scanning" mode. But the cutting-edge serial crystallography approach deals with pulsed X-ray radiation. Modern free electron lasers produce the femtoseconds long pulses of X-ray radiations, containing significant number of photons in one pulse, which enable collecting data from nanometers size protein crystals stream.

2. Methods

2.1. Crystallization of ferredoxin

The crystallization for ferredoxin and cytochrome was conducted by hanging drop method. Here the experiment of ferredoxin crystallization is described.

1. The 1st optimization tray was following:

1st tray	15 % PEG 3350	17 % PEG 3350	19 % PEG 3350	21 % PEG 3350	23 % PEG 3350	25 % PEG 3350
20 mM Zn-Acetate						
33 mM Zn-Acetate				Crystals		
46 mM Zn-Acetate						
59 mM Zn-Acetate						

   a. Ferredoxin with concentration 11.5 mg/ml was centrifuged 30 minutes at max speed.
   b. The drops contained 2µl reservoir and 1µl protein, 1µl reservoir and 1µl protein. Reservoir volume was 500µl.

3. The 2nd optimization tray was following:

2nd tray	19 % PEG 3350	20 % PEG 3350	21 % PEG 3350	22 % PEG 3350	23 % PEG 3350
25 mM Zn-Acetate					
30 mM Zn-Acetate					
35 mM Zn-Acetate					
40 mM Zn-Acetate			Crystals		Crystals

   a. Ferredoxin with concentration 50 mg/ml was centrifuged 30 minutes at max speed.
   b. The drops contained 2µl reservoir and 1µl protein, 1µl reservoir and 1µl protein. Reservoir volume was 500µl.

c. The crystals (thin needles) were worse than in the 1st tray. Crystals appeared in 33mM Zn-Acetate and 21% PEG 3350.

4. Besides, vapor diffusion crystallization (sitting drop) method was used, it surprisingly gave better results:
   a. Ferredoxin samples with concentration 50 mg/ml (f2) and 70 mg/ml (f_{new}) were centrifuged 30 minutes at max speed.
   b. Reservoir contained the 40µl of 40mM Zn-Acetate 21% PEG 3350 solution.
   c. This method gave better crystals in C1 well. The 2nd column was done with microseeds – the 10^{-7} dilution of batch crystallization of ferredoxin described below.

	1	2
A	1µl (f2)	1µl (f2)
B	2µl (f2) + 1µl (Reservoir)	2µl (f2) + 1µl (10^{-7} seeds)
C	1µl (f_{new}) - Crystals	1µl (f_{new}) - Crystals
D	2µl (f_{new}) + 1µl (Reservoir)	2µl (f_{new}) + 1µl (10^{-7} seeds)
E	1µl (f_{new}) + 2µl (Reservoir)	1µl (f_{new}) + 2µl (10^{-7} seeds)

2.2. Batch crystallization of ferredoxin

Batch crystallization means obtaining a lot of small crystals (from hundreds of nanometers to a few micrometers) in comparably big volume for serial crystallography. The batch for ferredoxin was done after the 1st tray.

The procedure was following:

1. Added to 10µl (ferrodoxin 11 mg/ml) by drops 20µl of 33mM Zn Acetate and 23% PEG 3350 in 0.5 ml epi.
2. Added to 10µl (ferrodoxin 12 mg/ml) by drops 30µl of 33mM Zn Acetate and 23% PEG 3350 in 0.5 ml epi.
3. Left samples at 20°C.

2.3. Data processing. XDS

XDS (X-ray Detector Software) was used for processing cryo-crystallographic data. XDS can processes a sequence of adjacent, nonoverlapping rotation images collected from a single-crystal at a fixed X-ray wavelength and recorded by a variety of imaging plate, CCD, pixel and multiwire area detectors, accepts fine sliced data as well as images covering a large oscillation range where each image in the data set covers the same positive amount of crystal rotation, allows arbitrary but fixed orientations of the detector and rotation axis, and only requires that incident beam and rotation axis intercept in one point in the center of the crystal, automatically derives reflecting range, spot width, crystal orientation, symmetry, and cell parameters from the data images.

The usual steps for the processing of cryo-crystallographic data was following:

1. Making a file with paths to all processing data files – files.lst.
2. Splitting data to the groups according samples (usually manually from experiment's logbook) and detector distances.
3. Launching for every group automatic bash script, which launches XDS for every dataset[5].

4. Understanding, why XDS did not index the data. Tring to adjust in XDS.INP following parameters:
    a. Checking if all data concluded – DATA_RANGE.
    b. Switching the FRIEDEL'S_LAW:
        i. FRIEDEL'S_LAW=TRUE if reflections h, k, l and -h, -k, -l are expected to have the same intensities.
        ii. FRIEDEL'S_LAW=FALSE if reflections h, k, l and -h, -k, -l could have different intensities due to anomalous scattering effects.
    c. Changing TRUSTED_REGION, INCLUDE_RESOLUTION_RANGE and parameters connected with number and intensity of pixels in the peaks.
5. Among other reasons, why XDS did not index the data, were experienced following:
    a. Wrong rotation axis written in XDS.INP.
    b. Wrong detector center (it was corrected in CrystFEL).
    c. Wrong detector distance, it might be corrected only during measurements.

2.4. Data processing. Crystfel

CrystFEL is software, which can index, integrate, merge, view and evaluate serial crystallography data. In this work CrystFEL was used for adjusting detector geometry.

1. The CrystFEL was launched without indexing to find peaks and form stream file for several dataset for one sample and one detector distance.

```
$ indexamajig -i list.lst -o output_file.stream -g geometry_file.geom --index=none -j 32 --peaks=peakfinder8 --min-snr=3.5 --min-pix=2 --max-res=1000 --threshold=1
```

2. Then found peaks were summarized with python script to the h5 file, their sum formed concentric rings.
3. The sum was visualized with proposed detector geometry and "simple" rings, characterizing geometry. In the calibration mode the "simple" rings were manually shifted till the matching with the concentric rings of sum.

```
$ hdfsee phyco_crystfel/P11_1_13-16/p1-powder.h5 -g crystfel/pilatus6M_300mm.geom --simple-rings=100,200,300,400
```

4. If the center of the concentric rings of sum was far away from the "simple" rings center, the new geometry file was saved and used for further data processing.

2.5. Data processing. Cheetah

Serial femtosecond inevitably produces a large amount of the "snapshots" – frames of protein stream. But there is no information from protein diffraction in some frames. Cheetah[6] is the software, which sort frames and choose only the ones, containing enough information for further data processing.

Cheetah was used for:

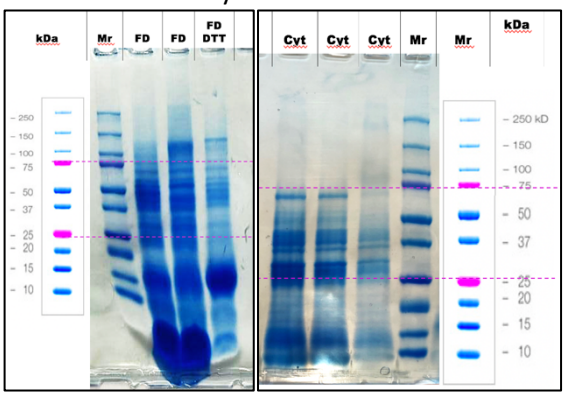
1. applying different masks to datasets to exclude the detector defects, ice or salt rings, other permanent trough dataset untrusted or empty detector regions,
2. estimation and subtraction of photon background, which measured separately in darkness – with closed detector,
3. hit finding and frame sorting,
4. identification and integration of Bragg peaks,
5. conversion of data for further Crystfel analysis.

3. Results

3.1. Ferredoxin Crystals

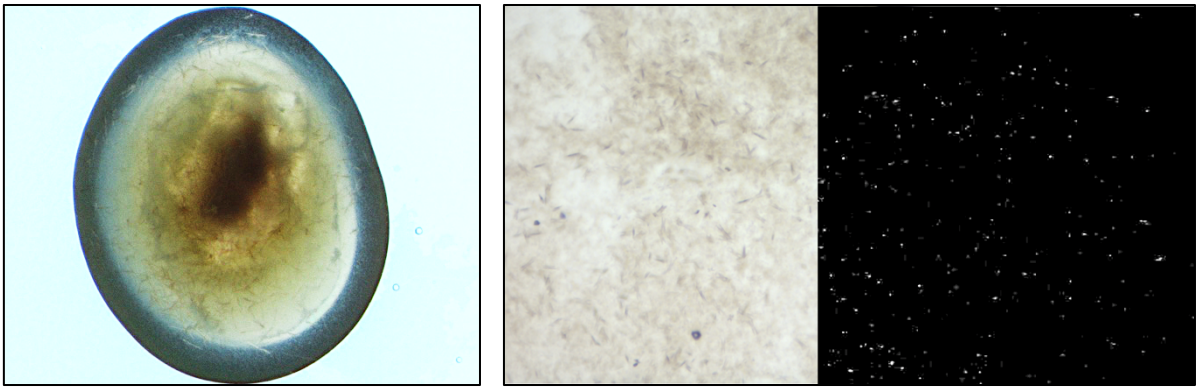
Ferredoxin and cytochrome c6 were obtained from host organism *Arthrospira platensis*. The quality of purified proteins was estimated on SDS-page (Figure 2). Ferredoxin size is 10.6 kDa. Cytochrome c6 size is 11.9 kDa. The denaturing buffer were added to protein samples and they were incubated 1 minute on 90°C. The proteins were stained with silver staining. Notable, not reduced ferredoxin "goes through" gel, the reducing agent DDT stabilizes denatured ferredoxin at its molecular mass.

Figure 2. SDS-pages. Left – for ferredoxin (FD) and ferredoxin with DTT (FD DTT), right – for cytochrome.



The 1st ferredoxin crystals (thin needles) were obtained on the 1st tray only in one drop 1µl reservoir and 1µl protein with the 500µl reservoir with *33mM Zn-Acetate and 21% PEG 3350* (Figure 3, left). Then the crystallization conditions were improved, and the bigger crystals appropriate for standard crystallography measurements were obtained in "sitting drop plate". The best crystal was observed in the ***1µl drop of ferredoxin with 70 mg/ml concentration with the 40µl reservoir with 40mM Zn-Acetate and 21% PEG 3350***. The small crystals were observed via SONICC technique in both ferredoxin batch (Figure 3, right).

Figure 3. Left – ferredoxin crystals, hanging drop, 500µl reservoir with 33mM Zn-Acetate and 21% PEG 3350. Right – ferredoxin batch visible light and SONICC images.



Ferredoxin crystals for appropriate for cryo-crystallography and SFX were obtained. There was not enough time for data collection during the available summer beam time. The data will be collected later.

3.2. Processing data from PETRA-2019-Sarrou-Mar-P11 Experiment

The geometry for this dataset was adjusted from data for phycocyanin.

1. For 200mm from the dataset:

/gpfs/cfel/cxi/scratch/data/2019/PETRA-2019-Sarrou-Mar-P11/raw/P11_11_p1_phyco/P11_11_p1_phyco_001/

Final geometry file:

/gpfs/cfel/cxi/scratch/user/pakhnovm/nlyubayk/crystfel/pilatus6M_200mm.geom

2. For 300 mm from dataset:

/gpfs/cfel/cxi/scratch/data/2019/PETRA-2019-Sarrou-Mar-P11/raw/P1126_3/P1126_3_001/P1126_3_001_00215.cbf

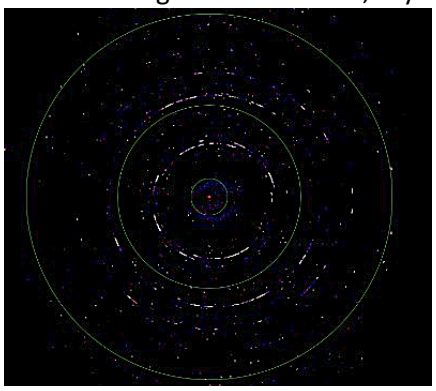
Final geometry file:

/gpfs/cfel/cxi/scratch/user/pakhnovm/nlyubayk/crystfel/pilatus6M_300mm.geom

As described above the data was splitting according samples and detector distances, checked geometry (Figure 4) and automatic bash script for every group was launched. The majority of datasets were not indexable. The unit cell parameters for one indexable dataset were approximately following:

140 Å / 280 Å / 780 Å / 90° / 90° / 90°

Figure 4. Checking detector centre, CrystFEL.



One unit cell parameter was comparable big (780 Å). It was counted, that the detector distance should be about 820 mm to separate peaks (if 1 pix of detector corresponds to 176µm, one separate peak on the screen corresponds to 6 pix, E = 12.4 keV). The majority of data were collected on the 500mm detector distance. Only one dataset P11_04_p4 collected on 800mm was indexed.

3.3. Processing data from PETRA-2019-Sarrou-Mar-P14 Experiment

A range of indexable datasets were collected taking to the account one big cell parameters.

1. Totally, 64 datasets were collected.
3. The 28 datasets were indexed from the first XDS launch.

4. XDS with fixed unit cell parameters was launch to index other datasets – only the 1st and 2nd datasets for puck_HH_011_7 were indexed.
5. XPARM was changed to GXPARM for indexed datasets to improve the quality of indexing. The quality became worse for the majority of dataset except:
    a. puck_599A_1_7300
    b. puck_599A_4, 1
    c. puck_599A_1_6979, 3
    d. puck_599A_1_3, 1

Finally, the dataset with the best resolution is

- ***puck_362A_3/green_1***
- space group number: 16.0
- unitcell parameters: 138.9482 / 224.7092 / 766.5823 / 90.0 / 90.0 / 90.0
- the resolution: 7.768

3.4. Processing LCLS data

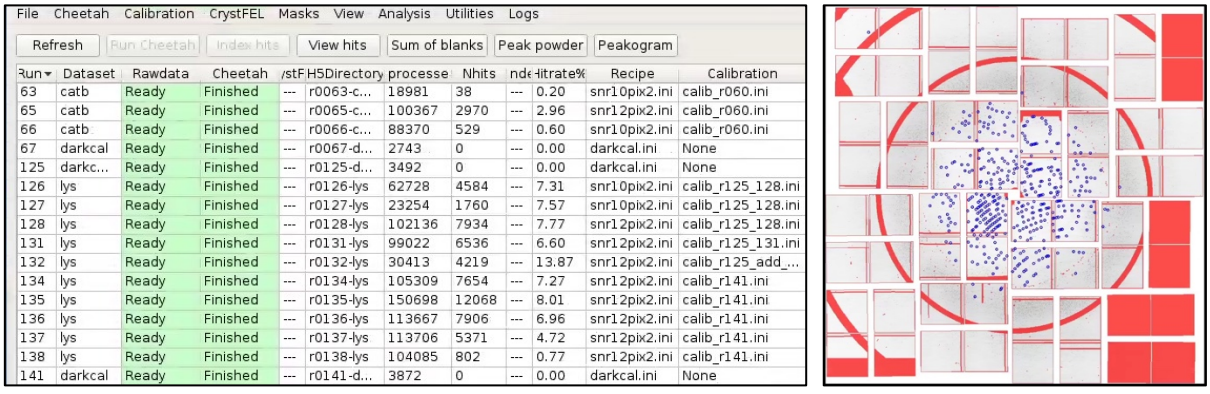
The methods of data processing are constantly developing. Here the data obtained in 2011 were processed again to show, that after 8 years is possible to reach better resolution from the same raw data due to more accurate procedure and better established detector geometry.

The serial crystallography data for lysozyme and cathepsin B was collected in 2011 at LCLS (Linac Coherent Light Source, Stanford). As the result the structures of with proteins were solved with 1.9Å (lysozyme[7]) and 2.1Å (cathepsin B[8]) resolution.

For the better data processing:

1. The appropriate dark measurements were found and checked for every run.
2. Masks considering detector features and salt/ ice rings were created and checked for all runs.
3. The parameters for pick finder were optimized.

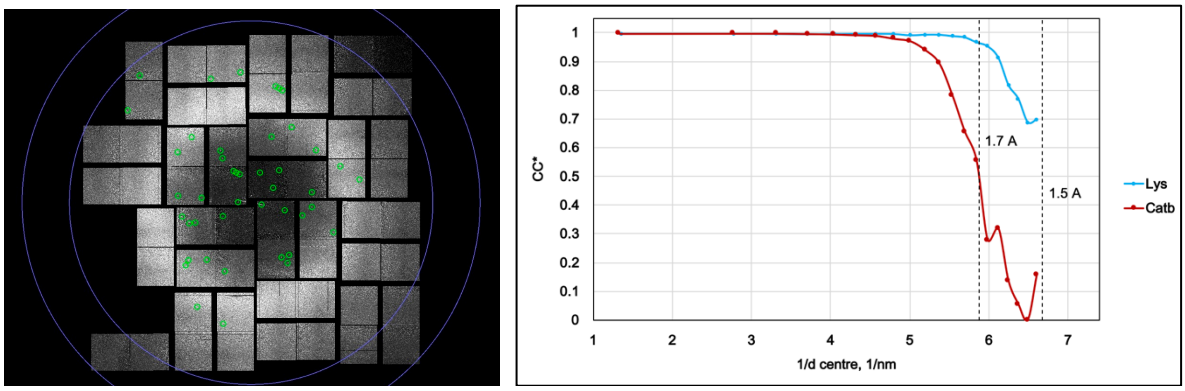
Figure 5. Left – Cheetah interface, showing a few of processed runs, right – the example of created mask.



Run	Dataset	Rawdata	Cheetah	/stF	H5Directory	processe	Nhits	nde	itrate%	Recipe	Calibration
63	catb	Ready	Finished	---	r0063-c...	18981	38	---	0.20	snr10pix2.ini	calib_r060.ini
65	catb	Ready	Finished	---	r0065-c...	100367	2970	---	2.96	snr12pix2.ini	calib_r060.ini
66	catb	Ready	Finished	---	r0066-c...	88370	529	---	0.60	snr10pix2.ini	calib_r060.ini
67	darkcal	Ready	Finished	---	r0067-d...	2743	0	---	0.00	darkcal.ini	None
125	darkc...	Ready	Finished	---	r0125-d...	3492	0	---	0.00	darkcal.ini	None
126	lys	Ready	Finished	---	r0126-lys	62728	4584	---	7.31	snr10pix2.ini	calib_r125_128.ini
127	lys	Ready	Finished	---	r0127-lys	23254	1760	---	7.57	snr10pix2.ini	calib_r125_128.ini
128	lys	Ready	Finished	---	r0128-lys	102136	7934	---	7.77	snr12pix2.ini	calib_r125_128.ini
131	lys	Ready	Finished	---	r0131-lys	99022	6536	---	6.60	snr12pix2.ini	calib_r125_131.ini
132	lys	Ready	Finished	---	r0132-lys	30413	4219	---	13.87	snr12pix2.ini	calib_r125_add_...
134	lys	Ready	Finished	---	r0134-lys	105309	7654	---	7.27	snr12pix2.ini	calib_r141.ini
135	lys	Ready	Finished	---	r0135-lys	150698	12068	---	8.01	snr12pix2.ini	calib_r141.ini
136	lys	Ready	Finished	---	r0136-lys	113667	7906	---	6.96	snr12pix2.ini	calib_r141.ini
137	lys	Ready	Finished	---	r0137-lys	113706	5371	---	4.72	snr12pix2.ini	calib_r141.ini
138	lys	Ready	Finished	---	r0138-lys	104085	802	---	0.77	snr12pix2.ini	calib_r141.ini
141	darkcal	Ready	Finished	---	r0141-d...	3872	0	---	0.00	darkcal.ini	None

It was shown, that it is possible to reach resolution 1.5Å for lysozyme (lys) and 1.7Å for cathepsin B (Catb) (Figure 6).

Figure 6. Left – resolution rings for 97 mm detector distance, right – the actual resolution.



The preprocessed data was indexed with Crystfel with new (comparably with 2011) geometry. The processing of these data was not finished:

- The detector geometry should be further optimized.
- Other indexing programs should be checked.
- The new indexing strategy should be checked with old detector geometry.
- The indexing without mask should be conducted to check the difference.

4. References

1. Liu, L. N. Distribution and dynamics of electron transport complexes in cyanobacterial thylakoid membranes. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 256–265 (2016).
2. Sétif, P., Mutoh, R. & Kurisu, G. Dynamics and energetics of cyanobacterial photosystem I:ferredoxin complexes in different redox states. *Biochim. Biophys. Acta - Bioenerg.* **1858**, 483–496 (2017).
3. Kekilli, D. *et al.* Photoreduction and validation of haem-ligand intermediate states in protein crystals by in situ single-crystal spectroscopy and diffraction. *IUCrJ* **4**, 263–270 (2017).
4. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73–78 (2011).
5. Galchenkova, M. 3D merging : getting more information from crystallography data. *Desy Summer Sch.* (2017).
6. Barty, A. *et al.* Cheetah: Software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Crystallogr.* **47**, 1118–1131 (2014).
7. Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science (80-.).* **337**, 362–364 (2012).
8. Redecke, L. *et al.* Natively inhibited trypanosoma brucei cathepsin B structure determined by using an x-ray laser. *Science (80-.).* **339**, 227–230 (2013).