Report for MX2203 and MX2322
Proposal Title: The Institute of Structural and Molecular Biology (ISMB) application for beam time, comprising groups from Birkbeck College, UCL and the Queen Mary University of London

Overall summary

Although only one publication was generated from data collected at ESRF during the previous session (considering also the long shutdown), our BAG was quite active and overall, we have collected more than 2,000 data sets over the past year. Most of the groups were focusing on ligand and inhibitors screening, still we have determined as well new structures. We are expecting several PDB depositions and publications over the next year.
Below we are providing 3 highlights of our research as also a more detailed summary of the BAG activities.

Research Highlights

1. **Kozielksi Lab: more than 1200 datasets mainly on Massif-1 (MX2203 & MX2322)**

Fragment-based screening of SARS-CoV-2 nsp1 reveals ligand-binding sites and cross-inhibition with other medically relevant coronaviruses

School of Pharmacy: Shumeng Ma, Shymaa Damfo, Jiaqi Lou, Shozeb Haider and Frank Kozielski; ESRF: Matthew Bowler

Non-structural protein 1 (nsp1), a unique viral protein, has been shown to be a crucial virulence factor causing host mRNA degradation and suppressing interferon (IFN) expression as well as host antiviral signaling pathways. In view of the essential role of nsp1 in the coronavirus (CoV) life cycle, it is regarded as a potential target for antiviral agent discovery. No inhibitors of nsp1 have been reported do far.

Our project aims at identifying fragment hits as a starting point for the development of more potent analogues by fragment-based screening of SARS-CoV-2 nsp1 via x-ray crystallography. Subsequently, ligands binding sites will be identified and characterised, followed by chemical optimisation of hits using structure-based drug design.

We expressed, purified and crystallised SARS-CoV-2 nsp1 and determined its structure improving the recently reported resolution from 1.6 Å to 1.05 Å. We are hoping to collect data to below 1.0 Å. The space group is $P4_32_12$ with one molecule in the asymmetric unit. The crystals diffract routinely to below 1.7 Å resolution and are stable in the presence of DMSO for several hours, excellent conditions for fragment-based screening using x-ray crystallography at ESRF beamline Massif-1. After optimising soaking and cryo-conditions we collected datasets for ca. 600 nsp1-fragments soaks and ca. 40 native crystals for the ground state in batches of 200 crystals in automated mode. Automated data collection parameters were further optimised between batches to increase data quality. Although PandAA analysis is not fully finished yet, we identified, verified and confirmed ten nsp1-fragment complexes so far. This number is at the lower end of success rates reported for fragment-based screening, however, it should be noted that nsp1 is a relatively small protein of 10 kDa (Figure 1).

Figure 1: Structure of SARS-CoV-2 nsp1 at 1.05 Å resolution. Helices are coloured in cyan and strands are shaded in purple. Water molecules are shown as red circles.

Currently we are characterising the nsp1-fragment interactions using orthogonal biophysical methods such as thermal shift assays (TSA) and microscale thermophoresis (MST) assays. In particular, MST allows us to determine K_d values and rank the fragment hits. We have also expressed and purified nsp1 from SARS and MERS and set up MST assays, allowing us to detect potential cross-inhibition with other CoVs of medical relevance. In parallel we initiated "SAR-by-catalogue" and started testing commercially available fragment analogues as a first step to improve potency of the fragments.

We hope to be able to contribute to the development of potent inhibitors targeting SARS-CoV-2 nsp1 and to the better understanding of this important protein [1].

Roberts Lab: A new protein complex was collected on ID23-1 8/12/2021 (MX2203)

Structure of the TCTEX-TCTEX1D2 heterodimer from the dynein-2 complex

Dynein-2 is a 1.4 MDa multi-subunit protein complex found in the primary and motile cilia of eukaryotic cells, where it is responsible for retrograde intraflagellar transport. Mutations in dynein-2 cause human ciliopathies including skeletal dysplasias and retinal degeneration. We previously determined the cryo-EM structure of the dynein-2 complex to 4.5 Å resolution (Toropova et al., 2019), revealing that the two identical motor subunits are contorted into different conformations by a sub-complex of associated subunits. However, the peripheral subunits were flexibly attached and their structure could not be determined. Here, we recombinantly expressed these two key subunits, TCTEX and TCTEX1D2, and determined their crystal structure to 2 Å resolution using ESRF beam time. The structure shows that TCTEX and TCTEX1D2 form a heterodimer, shedding light on how the symmetry of the homodimeric dynein-2 subunits is broken. This is functionally important, as the asymmetric structure of dynein-2 is tailored to its regulatory partner, the IFT-B complex. The structure also raises new questions about the function of the TCTEX1D2 N-terminal β-strand within dynein-2.

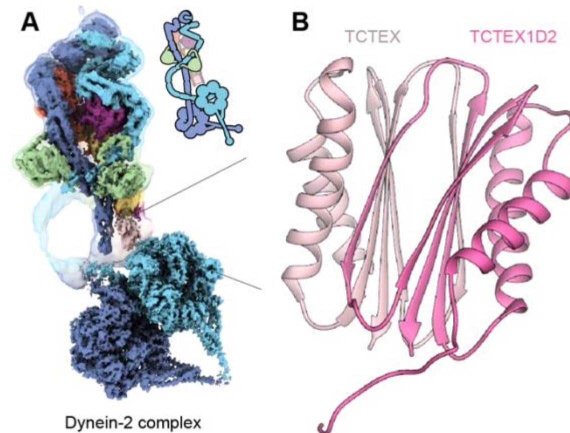

Figure X - Structure of the TCTEX-TCTEX1D2 heterodimer from the dynein-2 complex.
(**A**) Cryo-EM map of the dynein-2 complex (Toropova et al., 2019). The two identical motor subunits (shown in different shades of blue) adopt distinct conformations. (**B**) 2 Å resolution crystal structure reveals that the TCTEX and TCTEX1D2 subunits form a heterodimer which may help to break the symmetry of the dynein-2 complex.

Structure of the dynein-2 complex and its assembly with intraflagellar transport trains.
Toropova K, Zalyte R, Mukhopadhyay AG, Mladenov M, Carter AP, Roberts AJ.
Nat Struct Mol Biol. 2019 26(9):823-829. PMID: 31451806

Savva Lab: A new protein complex was collected on ID30B (8 July 2021, MX2322)

Distant sequence homology reveals the conservation of uracil-DNA glycosylase inhibition in novel permutations of SCC*mec* (manuscript in preparation)

Biologically irreversible stoichiometric inhibition of uracil-DNA glycosylase activity has evolved as an anti-restriction response in some viruses: Since the action of uracil-DNA glycosylases on single stranded regions of DNA will lead to chain breaks and replication poisoning, its inhibition by virus-encoded proteins will prevent this type of antiviral restriction. This survival strategy has been horizontally acquired by the transposable SCC*mec* cassette that underlies methicillin resistance in human and animal pathogens, typified by MRSA. Ung-inhibition ensures the integrity and efficiency of translocation in the form of ssDNA between cells of SCC*mec* DNA. A variety of classes of SCC*mec* are known, and all are characterised by encoding the DUF950 protein, also known as SAUGI, which is a potent Ung-inhibitor in the family of Ugi proteins. While investigating distant homology in the DUF950 family and its structurally conserved relationship to the bacteriophage encoded Ugi proteins, we discovered DUF950 proteins in new bacterial species, and novel permutations of SCC*mec*. This discovery will contribute to our understanding of the origins of transposable antibiotic resistance between different species of bacteria. The 2.7 Å structure of DUF950 from one of these species was recently determined from ESRF ID30B data and additionally sheds light on evolution and sequence plasticity in the family of Ugi proteins.

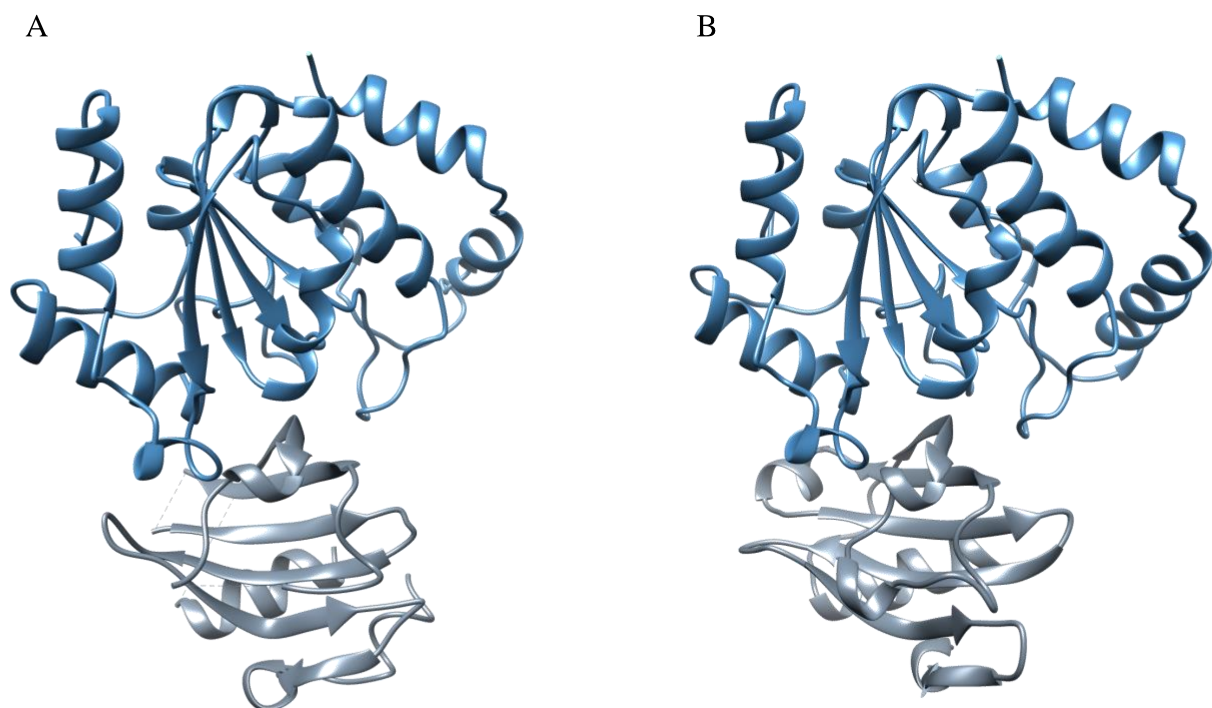

Figure 1: (A) The newly discovered DUF950 in complex with uracil-DNA glycosylase from Staphylococcus aureus. (B) The DUF950 known as SAUGI in complex with uracil-DNA glycosylase from Staphylococcus aureus (PDB: 3WDG).

Additional reports from usage of beamlines:

1. Irving group, used a significant amount of beam time during the past year

Determinants of conformational stability and instability in alpha-1-antitrypsin and the serpin superfamily/Designing diagnostic tools for alpha-1-antitrypsin deficiency

Data collection at ESRF since September 2020

Target	PDB	Beamline	Res.	Change
Fab2C1-AAT complex		ID23-2	1.85	The first high resolution structure of this complex; previous attempts have given 4-12A at best
Fab9C5-AAT complex		ID23-1	-	Screening to improve beyond current 4A dataset
AAT E342K		ID23-1	>10	Screening of crystals for usable diffraction
AAT I340W			>7	Screening of crystals for usable diffraction
NEUS		ID23-2	1.95	Identification of high resolution crystallisation conditions; protein underwent proteolysis in the drop

(1) Crystals were screened of a homologue of AAT that readily adopts a polymerisation-prone intermediate conformation. The best target diffracted to 1.95Å but the protein had been proteolytically clipped in the crystallisation drop. We are continuing to screen conditions. Another variant that has been engineered to prevent this intermediate transition was also screened but suitable diffraction has not yet been obtained.

(2) We have developed a toolkit of monoclonal antibodies that are able to recognise this protein selectively in different conformational states. As well as informing on the structural changes that result in aggregation and loss of activity, these antibodies have significant potential as clinical diagnostic tools. The 9C5 monoclonal antibody is used routinely in our lab in immunoassays, cryo-EM and has potential as a clinical diagnostic tool. Unsuccessful attempts were made to obtain a dataset of a complex between Fab9C5 and AAT with a resolution better than the current ~3.9A.

(3) The 2C1 monoclonal antibody is selective for the pathogenic polymer form of AAT. It has been used in many studies, and has been used to demonstrate the presence of polymers in the circulation of all individuals with the predominant causative allele of AAT deficiency. It has significant potential as a clinical diagnostic tool. Its selectivity also is expected to shed insight into the structural changes that occur during polymerisation. The only previous data obtained were for Fab 2C1 in complex with AAT diffracted at best to 4A with 16 copies present in the asymmetric unit. A new crystal form yielded a 1.85A dataset (see figure), providing unequivocal high-resolution detail of the basis for the recognition of AAT

polymers. These data, as well as revealing changes that occur in the protein during polymerisation, are now being used in a small molecule screening campaign with the aim of developing a non-invasive clinical imaging reagent.

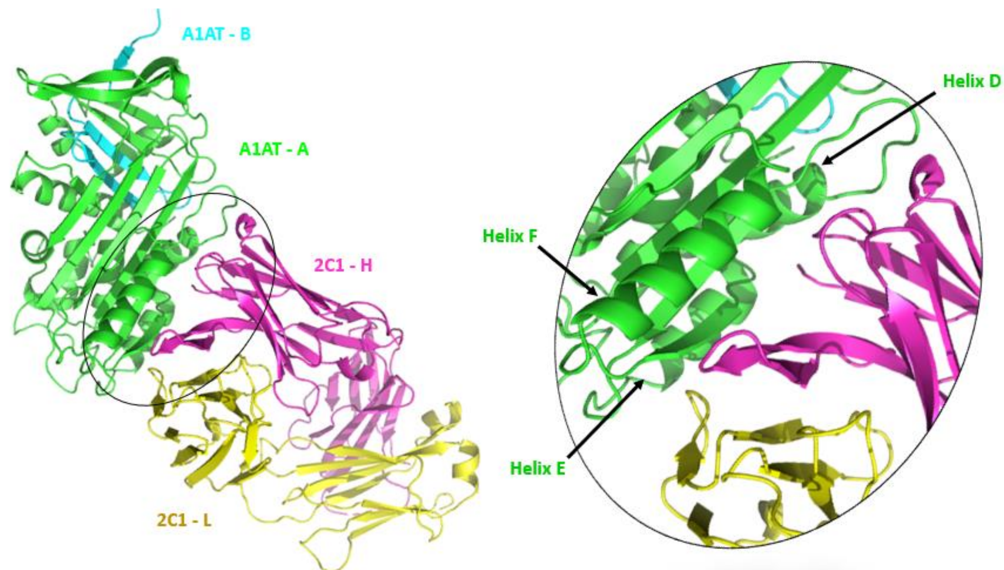

Figure. The structure of AAT in complex with Fab_{2C1} solved in $P2_12_12$ and refined to 1.85Å. AAT appears as green and cyan; the Fab heavy and light chains are magenta and yellow, respectively. The interface is fully defined in the electron density. The final R_{work} and R_{free} were 17.8% and 20.5%, respectively.

2. Morten Team (Part of Towers group UCL)

The team was regularly using the beam-lines during the past year

The work of the Morten team is focusing on new inhibitors for the HIV capsids. In a second project they are looking for interactions of the SARS-CoV-2 protein ORF6 (p6) with human proteins in order to develop efficient drugs for the treatment of the virus infection:

For HIV-1(M) capsids we are currently obtaining reproducible crystals that diffract down to a resolution of 1.9 Å. We are currently attempting to obtain structures with HIV-1(M) CA with several novel capsid inhibitors as well as peptide co-factors.

For our SARS-CoV-2 ORF6 project we have obtained peptides corresponding to the C-terminus region of ORF6 as this part has shown to be critical for ORF6's antagonistic effect. Furthermore, we co-crystallized ORF6 with the importin KPNA2. These crystals diffracted down to 2-3 Å, but we did not observe complexes. For this project we are focusing on sorting the protein-protein interactions with Crosslinking mass spectrometry (XL-MS) before proceeding with crystallography.

3. Gouge Team

The team was regularly using the beam-lines during the past year

Co-Enzyme A (CoA, Figure 1A) as a novel metabolic integrator possessing antioxidant activities in human cells. We are particularly interested in two proteins that were shown to be CoAlated in cells in response to oxidative stress, Prdx6 and NME1. These two proteins are regulated by

CoAlation and are prime target in cancer therapy. Prdx6 is able to reduce and repair oxidised lipid, protecting cells against ferroptosis. NME-1, a tumour suppressor, is down-regulated in cancer.

We recently discovered that Prdx6 can react with CoA in cells, which in turn modulates its activity. The goal of this project is to characterise by crystallography the binding site of CoA in Prdx6. We co-crystallised Prdx6 with CoA, or soaked CoA in apo Prdx6 crystals. We could collect 8 datasets between 2 and 2.6 Å. Unfortunately, we didn't observe any clear density that could be attributed to CoA in the composite omit map (Figure 1B). However, there is some residual density, suggesting partial occupancy. We managed to derivatise, in a test tube, Prdx6 with CoA (covalent bond between the thiol of CoA and the reactive cysteine of Prdx6). We screen about 25 derivatised crystals but didn't collect any data due to the poor diffraction (>6Å). We are now focusing our efforts on obtaining the structure of Prdx6 with the reactive pantetheine tail that is less likely to perturb the crystal packing.

NME-1 is a NDP kinase that can transfer the gamma phosphate from ATP to any NDP. We identified CoA a regulator of NME-1 through covalent and non-covalent binding in the active site. We collected 4 datasets of NME-1/CoA complex non-covalently bound, with resolutions ranging from 1.6 to 2.6 Å (Figure 1C). We are currently working on obtaining CoA covalently bound to NME-1.

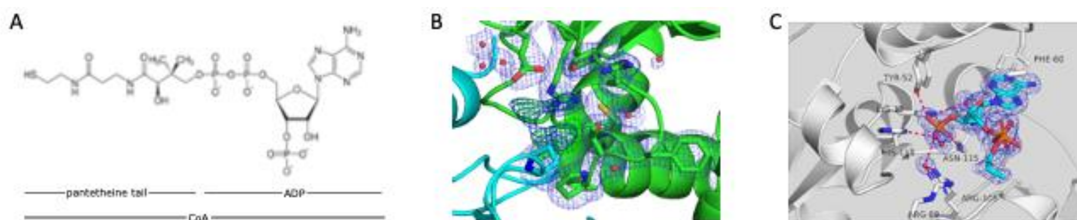

Figure 1: A. Chemical structure of CoA. B. Close-up view of Prdx6 active site. A representative omit map shows the presence of residual density (contouring at 1.2σ). C. Close-up view of NME-1 active site with CoA bound non-covalently (contouring at 1.2σ).

4. Wallace group
The Wallace group was one of the major users during past year

The Wallace Group's work is focused on elucidating the nature of the transmembrane drug interactions with voltage-gated sodium channels. These studies include comparisons of native and mutant sodium channels, where the mutations were designed to block the fenestrations and drug binding. Several data sets were collected at the ESRF to screen for drug complexes in the previous period, but none of these led to data that was included in any publications. However, once we publish papers about the complexes, even if the final datasets do not come from the ESRF, we will acknowledge use of the facility.

5. Katan group
The group was regularly using beamtime for most of the sessions

The Katan group is focusing on inhibitors for the full-length phospholipase C (PLC) γ. The aim is to understand the mechanism of activation and autoinhibition of PLCγs and illustrate the effect of mutations on the functioning of the enzyme. We recently established the crystallisation of the full length PLCγ and we are currently screening for analogues that can inhibit the function of the protein. So far, we have collected several data sets of the full-length

native protein reaching a maximum resolution of 2.2 Å. We are working on the reproducibility of crystal growth using two different conditions while in parallel we are screening for structures with bound ligands and inhibitors. So far the initial set of inhibitors didn't reveal any bound ligand in the crystal structure. We are further working by improving the protocols through co-crystallisation and soaking of crystals as also by screening new inhibitors. Once we have fully established crystallisation conditions we shall seek for a more detailed fragment screening.

6. Keep group

Used the beamlines once, collecting a data set of a new protein complex from Mycobacterium tuberculosis

Weakly diffracting crystals of ScoAB were obtained. The best crystals approached 3 Å diffraction in spacegroup C2 on ID23-2, but with high Rmeas in low resolution shells. As yet Molecular replacement has been unsuccessful.

7. Kozielski lab additional report

Human Kinesins

Eg5 is a kinesin with inhibitors in phase III clinical trials against refractory multiple myeloma.
MCAK is a microtubule (MT) depolymerising kinesin, which is involved in error correction of MT-kinetochore attachment during the pre-metaphase stage of mitosis. Therefore, its key role in mitosis makes it a potential target for cancer chemotherapy and we will employ structure-based drug design to develop inhibitors targeting MCAK.

Viral proteins

Dengue and Zika virus

RdRp_D (DENV-3) is a part of the flavivirus RNA-dependent RNA polymerase (RdRp), serotype 3, located at the C-terminus of non-structural protein 5 (NS5) of Dengue virus. Fragment-based screening and structure-based drug design will be employed to identify and develop hits against this target.

MTase_D Dengue methyltransferase (MTase) is a viral protein essential for sequential methylation of the N7 and 2'-O positions of the viral RNA cap, using S-adenosyl-L-methionine (SAM) as a methyl donor therefore playing an important role in viral RNA replication. It is located at the N-terminus of DENV NS5. Fragment-based screening and structure-based drug design approaches will be undertaken for finding inhibitors targeting DENV MTase.

Antibacterial targets

FtsZ from *Mycobacterium tuberculosis* and *Staphylococcus aureus* is a bacterial protein that assembles into a ring-line structure (called the Z-ring) at the future site of the septum of bacterial cell division. It is a prokaryotic homologue of the eukaryotic protein tubulin. We will employ fragment-based screening and structure-based drug design to identify and develop inhibitors against this potential bacterial target.

SARS-CoV-2

Nsp1: Non-structural protein 1 (nsp1), a unique viral and conserved leader protein, has been shown to be a crucial virulence factor causing host mRNA degradation, suppressing interferon (IFN) expression and host antiviral signaling pathways. Our project aims at identifying fragment hits as a starting point for the development of more potent analogues by fragment-based screening of SARS-CoV-2 nsp1 via x-ray crystallography. Subsequently, ligands

binding sites will be identified and characterised, followed by chemical optimisation of hits using structure-based drug design. This proect is in collaboration with Dr. Matthew Bowler.
Nsp5: Mpro is a papain-like cysteine protease involved in most maturation cleavage events within the precursor polyprotein. Therefore, Mpro plays a vital role in the life cycle of CoVs. This project is expected to identify potential inhibitors of the essential Main Protease via crystallography methods to interrupt the replication and invasion of new CoVs.
Nsp10: nsp10 is a stimulator and scaffolding protein for two important CoV proteins, nsp14 and nsp16, forming complexes with both proteins. Whereas nsp14 contains N7-methyltransferase (MTase) and exoribonuclease (ExoN) activities, nsp16 shows 2'O-MTase activity. These activities are crucial for viral replication making all three proteins potential targets for antiviral therapy. We are conducting fragment-based screening via x-ray crystallography against nsp10.

Summary of active projects in the Kozielski laboratory.

Sample Name	Description	Source [category risk]	Expression host [category risk]	Number of datasets	Type of data
Eg5-inhibitor	Eg5 in complex with inhibitor analogues	Human [1]	*E. coli* [1]	1-3 per complex	Native
Eg5-fragment	Eg5 in complex with fragment hits	Human [1]	*E. coli* [1]	1-3 per complex	Native
MCAK-fragment	MCAK in complex with fragment hits	Human [1]	*E. coli* [1]	1-3 per complex	Native
RdRp-fragment	DENV-3 fragment in complex with small molecule hits	Viral [1]	*E. coli* [1]	1-3 per complex	Native
MTase_D-inhibitor	DMTase in complex with small molecule hits	Viral [1]	*E. coli* [1]	1-3 per complex	Native
MTase_Z-inhibitor	ZMTase in complex with small molecule hits	Viral [1]	*E. coli* [1]	1-3 per complex	Native
Nsp1-fragment	SARS-CoV-2 nsp1 in complex with fragment hits or analogues	Viral [1]	*E. coli* [1]	1-2 per complex	Native
Nsp5-inhibitors	SARS-CoV-2 nsp5 in complex with inhibitors or analogues	Viral [1]	*E. coli* [1]	1-2 per complex	Native

Nsp10-fragment	SARS-CoV-2 nsp10 in complex with fragment hits or analogues	Viral [1]	*E. coli* [1]	1-2 per complex	Native
FtsZ-inhibitor	FtsZ in complex with small molecule hits	Bacterial (*S. aureus* and *M. tuberculosis*) [1]	*E. coli* [1]	1-2 per complex	Native

List of planned PDB submissions as a direct result of ESRF time
Although we did not submit any coordinates to the PDB yet, we determined and refined a large number of proteins, protein-fragment, or protein-analogue complexes:
SARS-CoV-2 nsp1: ca. 640 datasets collected, so far 10 nsp1-fragment complexes identified and refined. 10 PDBs to be submitted.
SARS-CoV-2 nsp10: ca. 600 datasets collected, so far 7 nsp10-fragment complexes identified and refined. 7 PDBs to be submitted.
Dengue virus MTase: 4 PDBs to be submitted.
Dengue virus RdRp: 7 PDBs to be submitted.
***M. tuberculosis* FtsZ**: 6 PDBs to be submitted.
Human MCAK: 2 PDBs to be submitted.