Report for MX-2488
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
We report the annual progress of the MX-2488. We continued our work focusing on viral proteins mainly from SARS-CoV-2 and HIV-M. A new group has joined our BAG and we report 3 new publications. All groups participated in the current proposal have made significant progress in their projects and we are expecting more publications in the near future.

Below we are providing 2 highlights of our research as also a more detailed summary of the BAG activities.

Research Highlight I: Emerging variants of SARS-CoV-2 NSP10 highlight strong functional conservation of its binding to two non-structural proteins, NSP14 and NSP16

The coronavirus SARS-CoV-2 protects its RNA from being recognized by host immune responses by methylation of its 5' end, also known as capping. This process is carried out by two enzymes, non-structural protein 16 (NSP16) containing 2'-O-methyltransferase and NSP14 through its N7 methyltransferase activity, which are essential for the replication of the viral genome as well as evading the host's innate immunity. NSP10 acts as a crucial cofactor and stimulator of NSP14 and NSP16. To further understand the role of NSP10, we carried out a comprehensive analysis of >13 million globally collected whole-genome sequences (WGS) of SARS-CoV-2 obtained from the Global Initiative Sharing All Influenza Data (GISAID) and compared it with the reference genome Wuhan/WIV04/2019 to identify all currently known variants in NSP10. T12I, T102I, and A104V in NSP10 have been identified as the three most frequent variants and characterized using X-ray crystallography, biophysical assays and enhanced sampling simulations. In contrast to other proteins such as spike and NSP6, NSP10 is significantly less prone to mutation due to its crucial role in replication. The functional effects of the variants were examined for their impact on the binding affinity and stability of both NSP14-NSP10 and NSP16-NSP10 complexes. These results highlight the limited changes induced by variant evolution in NSP10 and reflect on the critical roles NSP10 plays during the SARS-CoV-2 life cycle. Data presented in both highlights have been collected at ESRF beamline Massif-1 and were essential for both projects.

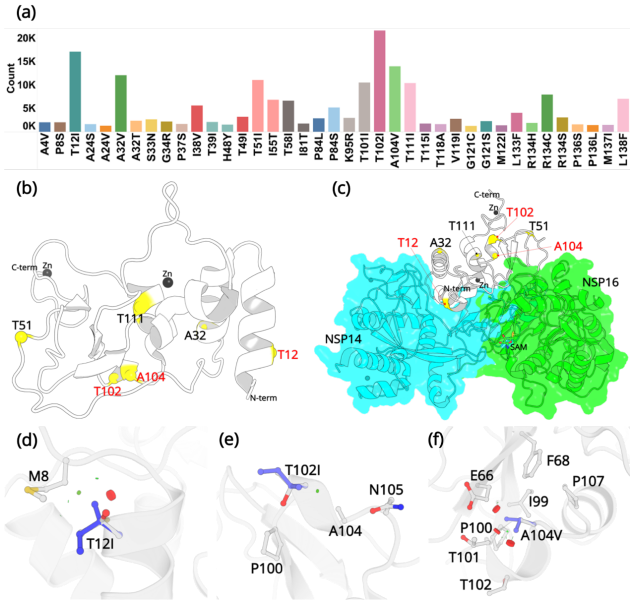

Figure 1. (a) Mutation count of the top 39 mutations in NSP10 extracted from >13 million WGS. (b) Locations of top six mutations on the NSP10 structure. Mutation positions labeled in red are the most frequently occurring and were used in crystallographic studies. (c) The spatial position of the mutations relative to NSP14 and NSP16 structures. Local structural environment of (d) T12I, (e) T102I and (f) A104V mutations (blue sticks) superimposed on the wild-type structure (white).

Publication: Wang, H., Rizvi, S.R.A., **Dong, D., Lou, J., Wang, Q., Sopipong,** W., Su, Y., Najar, F., Agarwal, P.K., **Kozielski, F**. and Shozeb Haider (**2023**). Emerging variants of SARS-CoV-2 NSP10 highlight strong functional conservation of its binding to two non-structural proteins, NSP14 and NSP16. *Elife, accepted.*

Research highlight II: High-Confidence Placement of Fragments into Electron Density Using Anomalous Diffraction—A Case Study Using Hits Targeting SARS-CoV-2 Non-Structural Protein 1

The identification of multiple simultaneous orientations of small molecule inhibitors binding to a protein target is a common challenge. It has recently been reported that the conformational heterogeneity of ligands is widely underreported in the Protein Data Bank, which is likely to impede optimal exploitation to improve affinity of these ligands. Significantly less is even known about multiple binding orientations for fragments (<300 Da), although this information would be essential for subsequent fragment optimization using growing, linking or merging and rational structure-based design. Here, we use recently reported fragment hits for the SARS-CoV-2 non-structural protein 1 (nsp1) N-terminal domain to propose a general procedure for unambiguously identifying binding orientations of 2-dimensional fragments containing either sulphur or chloro substituents within the wavelength range of most tunable beamlines. By measuring datasets at two energies, using a tunable beamline operating in vacuum and optimised for data collection at very low X-ray energies, we show that the anomalous signal can be used to identify multiple orientations in small fragments containing sulphur and/or chloro substituents or to verify recently reported conformations. Although in this specific case we identified the positions of sulphur and chlorine in fragments bound to their protein target, we are confident that this work can be further expanded to additional atoms or ions which often occur in fragments. Finally, our improvements in the understanding of binding orientations will also serve to improve the rational optimization of SARS-CoV-2 nsp1 fragment hits.

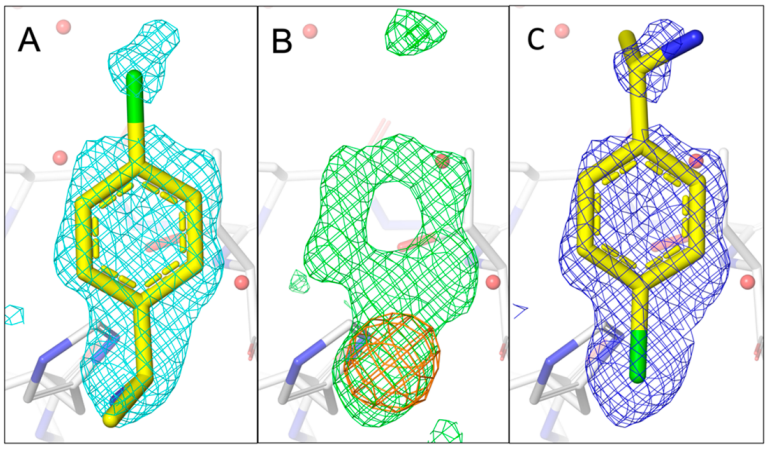

Figure 1. Comparison of the fragment binding site of previously reported conformation of **7H2** with that obtained using a chlorine anomalous difference Fourier map. **A)** The published configuration of **7H2** with the $2mF_o$-DF_c map around the fragment is shown in cyan. **B)** Anomalous difference Fourier maps calculated from data collected at 4.5 keV (orange) with the mF_o-DF_c map (green) in the fragment binding region. **C)** Refined $2mF_o$-DF_c of **7H2** with the chlorine located in the centre of the anomalous peak. The electron density mostly covers **7H2**.

Publication: Ma, S.; Mykhaylyk, V.; Bowler, M.W.; Pinotsis, N.; Kozielski, F. (2023). High-Confidence Placement of Fragments into Electron Density Using Anomalous Diffraction—A Case Study Using Hits Targeting SARS-CoV-2 Non-Structural Protein 1. *Int. J. Mol. Sci.* **24**, 11197.

Additional reports from usage of beamlines:

1. Irving group (UCL)

(1) We have identified a mutant of α-1-antitrypsin (AAT) in a patient population that we have found to be functionally inactive as a protease inhibitor. We have crystallised the variant in the native form and also in the protease-cleaved form. Data collection of the protease-cleaved form was undertaken at ID23-1, and a dataset obtained at 2.4 Å which was solved with two molecules in the asymmetric unit. This revealed the unfavourable burial of a polar aspartic acid side chain into a hydrophobic pocket.

(2) We have performed a ligand-viewed NMR-based screen which has identified compounds that show evidence of interaction with the polymer. Our data suggests the cleaved form of the protein is a suitable structural surrogate for the subunit of a polymer. We have performed crystal soaks of the cleaved form of AAT with these compounds. Overall, the 5-30 minute soaks resulted in degradation of the diffraction quality of the crystals to >3 Å; datasets of multiple crystals with three compounds were obtained at ID23-1. Unfortunately no positive density was visible for these compounds. Longer soaks over a two week period did not show any diffraction (at ID30A-3).

(3) We have developed a toolkit of monoclonal antibodies that are able to recognise AAT and a homologous serpin, neuroserpin, 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. We have previously obtained a 3.9 Å dataset of the complex between Fab9C5 and AAT and extensive attempts to obtain an improved resolution have failed. In this period we successfully obtained a much higher resolution dataset of Fab9C5 alone at ID23-1 which was solved to 2.2 Å. This will facilitate 3D single particle reconstructions from cryo-EM data and also help to refine the 3.9 Å dataset. We also obtained a 1.3 Å dataset of Fab1A10 which is reactive to neuroserpin and is being used for cryo-EM studies of neuroserpin polymers.

2. Frank Kozielski (School of Pharmacy, UCL)
Group's additional report details
SARS-CoV-2.

We conducted fragment-based screening via X-ray crystallography for SARS-CoV-2 **nsp10**. We managed to collect ca. 600 datasets for nsp10 crystals soaked with individual fragments and identified and verified ten fragment hits. We employed an MST assay for the ten fragment hits and determined their K_d values. We identified and described three ligand-binding pockets in nsp10. For SAR-by-catalogue, we identified and purchased 80 analogues in total and screened them by X-ray crystallography. 19 of those were confirmed as hits. We then determined their K_d values using MST and compared with their corresponding fragments, eight of them were significant improved by 5 - 20-fold. Based on these results we obtained an MRC-DPFS grant, which will commence December 1st.

SARS-CoV-2 non-structural protein 1 (**nsp1**) is regarded as a potential drug target for its crucial role in viral replication and invasion. We conducted fragment-based screening using X-ray crystallography to identify fragment hits, which serve as a starting point for lead compound design and development. Through four rounds of fragment screening at beamline Massif-1, we identified 12 fragment hits that bind to two distinct binding pockets in the N-terminal nsp1 domain. We subsequently purchased 164 analogues of the 12 fragment hits and prepared approximately 200 crystals of nsp1 soaked with analogues for screening at Massif-1. More than

40 analogues binding to nsp1 have been identified by PanDDA analysis and manual inspection. Orthogonal biophysical assays of the fragment analogues are in progress.

We collected various datasets for SARS-CoV-2 **nsp14** crystals soaked with various fragments. Among the 33 fragment hits previously identified from MST assays, we obtained 6 nsp14-fragment complexes. Although most fragments bound to known pockets, a change in space group was observed in one of the complexes and we identified a new fragment binding site. These nsp14 crystals routinely diffract to 1.9-2.4 Å and can be used to test fragment analogues to improve the initial fragment hits.

We conducted fragment-based screening of **dengue virus** and **zika virus methyltransferase (MTase)** at DLS collecting diffraction data of around 700 MTase-fragment soaks. The data allowed us to identify a range of novel ligand binding sites on MTase. As the crystals showed twinning and contained eight molecules in the asymmetric unit, we searched for novel crystallisation conditions. We were able to grow a new crystal form, space group $C222_1$, diffracting routinely to better than 2.0 Å resolution. We therefore switched to this untwinned crystal form collecting native diffraction data and three MTase-fragment analogue complexes. These structures will allow structure-based drug design and will serve as a basis to develop MTase-targeting inhibitors with higher potency and efficacy. Based on these results, a grant application will be submitted to the MRC September 7th.

We also collected data at Massif beamtime for **dengue RdRp** in complex with fragments (about 200 data sets). Data have been processed using the PanDDA software and seven fragments binding to DENV RdRp have been identified. A manuscript on RdRp fragment hits is in progress.

Human MCAK structures in the absence (nucleotide-free) and presence of various nucleotides (Mg^{2+}ADP, ATP-like states represented by AMP-PNP and ATPγS) have been determined and the structures are being prepared for submission. The manuscript is in preparation.

We have collected data on a large variety of crystals of ***M. tuberculosis*** **FtsZ** with fragments. These complex structures reveal novel inhibitor binding sites and will allow conducting structure-based drug design (in progress). These complex structures will serve as a basis to develop inhibitors with higher potency and efficacy. A grant application based on these results will be submitted to the MRC in November. A new PhD student has taken over the project from the previous PhD student.

List of publications attributable to ESRF time in the last 12 months

Wang, H., Rizvi, S.R.A., **Dong, D., Lou, J., Wang, Q., Sopipong,** W., Su, Y., Najar, F., Agarwal, P.K., **Kozielski, F**. and Shozeb Haider (**2023**). Emerging variants of SARS-CoV-2 NSP10 highlight strong functional conservation of its binding to two non-structural proteins, NSP14 and NSP16. *Elife, accepted.*

Ma, S.; Mykhaylyk, V.; Bowler, M.W.; Pinotsis, N.; Kozielski, F. (**2023**). High-Confidence Placement of Fragments into Electron Density Using Anomalous Diffraction—A Case Study Using Hits Targeting SARS-CoV-2 Non-Structural Protein 1. *Int. J. Mol. Sci.* **24**, 11197.

Ma, S., Damfo, S., Lou, J., Pinotsis, N., Bowler, M.W., Haider, S., and **Kozielski, F.** (**2022**). Two ligand-binding sites on SARS-CoV-2 non-structural protein 1 revealed by fragment-based x-ray screening. *Int. J. Mol. Sci.* **23**:12448.

Borsatto, A., Akkad, O., Galdadas, I., **Ma, S., Damfo, S.**, Haider, S., **Kozielski, F.,** Estarellas, C. and Gervasio, F.L. (**2022**). Revealing druggable cryptic pockets in the Nsp-1 of SARS-CoV-2 and other β-coronaviruses by simulations and crystallography. *Elife* **11**:e81167.

3. Larsen/Pinotsis Team (Towers/Selwood groups UCL)

The work of the Larsen team (continued by N. Pinotsis) focused on new inhibitors and interactions for the HIV capsids.

We screened crystals of the HIV-1(M) capsid protein (CA) soaked with a revised peptide derived from the known co-factors Nup153. This time the peptide was present in the crystal structure after collected a full data set at about 2.6 Å max resolution. This structure together with previous CA structures with inhibitors and peptides will be used for a manuscript describing novel inhibitors of the HIV virus.

4. Gouge Team (UCL/Birkbeck)

Cells are constantly subjected to oxidants that in turn can damage all their constituents. They therefore rely on complex systems to sense the redox balance and activate the detoxification machineries. Failing at maintaining the redox balance is a hallmark of human diseases, including cancer and neurodegenerative diseases. We are interested in the major regulator of the oxidative stress response in eukaryotes, the Nrf2 pathway. Nrf2 is a transcription that is activated under oxidative conditions. In absence of stress, Nrf2 promoters are repressed by the heterodimer Bach1-sMAFG. Upon activation, Nrf2 competes with Bach1 to bind sMAFG and the DNA.

We obtained crystals of the repressed (Bach1-sMAFG-DNA) and active form (Nrf2-sMAFG-DNA) of the promoters. Unfortunately, the crystal structure of Nrf2-sMAFG-DNA has been published by another group. However, over the last months, we have focused our attention on the repressed form of the promoters. We screened about 60 crystals with different DNA lengths. We managed to collect several datasets with a maximal resolution of 4.2 Å. However, the diffraction suffers from severe anisotropy. Future approaches will focus on screening more oligonucleotides, microseeding matrix screening, cryo-protection optimisations, and crystal dehydration.

5. Wallace group (David Hollingworth, Birkbeck College)

The majority of our beamtime for this period has focused on looking for the sites of drug interactions with the voltage gated sodium channel, NavMs. The drugs we choose for crystallography are ones that have been shown to interact with our channel by other methods (usually NMR) and we use crystallography in an attempt to identify and characterize the specific binding sites for the drugs within the channel. To aid in further study (especially by electrophysiology) we also used our beamtime to study identified drug binding site mutants with the drugs of interest. Crystal to crystal variability in channel-ligand interactions means that getting a ‘high quality’ dataset can require collection of many co-crystal datasets before a usable one is found. However, we have had a few successes during this session. Publications, and the linked submission of PDBs, require our collaborators to work with our findings to produce in-cell results which is a bottleneck for us. We predict, however, that some of the findings from this last beamline period should be published within a year from now.

6. Parkinson Group (School of Pharmacy, UCL)

Parkinson and Waller research groups use structure-based drug-discovery approaches as a key element to provide orthogonal structural data to support their biophysical data, help identify new molecular targets and validate new molecular entities as potential therapeutic agents centered on diverse families of nucleic acid secondary structures.

We have recently determined the crystal structure of the human insulin gene-linked polymorphic region (ILPR). This work represented the first structure elucidation by X-ray methods of an intramolecular i-motif and will also provide information on insulin regulation. The ILPR structure was deposited to the PDB with PDB ID **8AYG (on hold)**. We have also made significant progress in revealing the i-motif structure of the regulatory region of nitric oxide synthase (NOS) from *Paracoccus Denitrificans*. The NOS determination will provide an understanding of the control of the climate-damaging gas, N_2O, in soil bacteria.

Native NOS crystals were grown and also NOS crystals in the presence of a ligand were generated. Native ILPR and NOS crystals were soaked into a variety of i-motif binding ligands to investigate their potential interactions. Also, crystals of an ILPR variant (ACA) were tested to see the effect of minor modifications on the ILPR structure. Data collections were conducted on at the ID23-1 and the ID30A-3 beamlines (MX-2488) on the 5th of May and 23rd July 2023, respectively. Good quality data was collected and for NOS and ILPR soaks and we were able to confirm the presence/absence of ligand. Also, good quality data was collected for the S100P sequences in the presence of the QN-302 and CMO3 ligands. We are currently analysing the data from the collections with the aim to resolve the NOS, S100P-CMO3 and S100P-QN-302 structures. We are also doing crystallizations to obtain more crystals of other i-motifs and G-quadruplexes and see their interactions with the QN-302 ligand.

7. Lamber group (Kingston University), Pinotsis (Birkbeck College)

Our research is focusing on the myomesin protein family which is an essential protein for striated muscle contraction. Myomesin proteins bind to the thick filaments and keep them align around M-band which is the centre of the sarcomeric unit.

Last year we determined the structure of the C-terminal domains of myomesin-2, which is overexpressed in heart and fast twitch muscles. Currently we are working on the C-terminal of myomesin-3, which is expressed in medium speed twitch muscles. Initial crystals provided a data set of around 5 Å max resolution and we are currently optimising the crystallisation conditions and testing more crystals to achieve higher resolution.

8. Djordjevic group (UCL)

In March 2023 we tested 2 crystals of a synthetic, custom-designed small protein crystals on ID23.1. The crystals generated diffraction pattern that was consistent with a protein, but the resolution was poor, 8-10 Å and the images also showed features of a powder diffraction. Protein purification and crystallisation will be optimised for future experiments.

In June 2023 we obtained 4 data sets on ERSF ID23-2 for the crystals of neuropilin-2 in a complex with the specific small molecule ligand. The crystals were obtained from 4 different crystallisation conditions. Data were compromised by icing on the crystals. The best data set included diffractions to 2.4 Å but it had overall Rmrg of 29%. The crystals represented a new crystal form, and we were able to solve the structure. The electron density map clearly showed the presence of the ligand within the expected binding site, but the quality of the map

was poor, and we did not proceed with the refinement. This is the first data set for the neuropilin-2 antagonist.

9. Mark van Brügel (Queen Mary University of London)

SAS-6 is an evolutionary highly conserved and essential centriole duplication protein. We and others have demonstrated that SAS-6 establishes the symmetry and diameter of centrioles through self-assembly into rotationally 9-fold symmetric rings. Intriguingly, SAS-6 homologues can also be identified in a few eukaryotic lineages that do not possess centrioles, raising the question what assemblies are formed by these homologues, what cellular roles they play and whether (in case of pathogenic eukaryotes) they could provide potential drug targets. To gain insight into the structural basis of SAS-6 self-assembly in centriole-less eukaryotes, we have crystallised a SAS-6 construct from one of these organisms. From the crystals, we collected a native dataset at ESRF (MX-2488 ID30A-3, 23/07/2023) to a resolution of 3.5 Å ($P4_1$ 146.20, 146.20, 213.58, 90.00, 90.00, 90.00). Although the dataset was of high quality and showed no "pathologies", no molecular replacement solution could be found, probably due to the large unit cell with (most likely) 20-25 copies of SAS-6 in the asymmetric unit. Thus, we will next experimentally phase this crystal form and also screen for other crystal forms from different constructs (that might be easier to solve). To this end, we need more synchrotron time at ESRF. The obtained data will help us to understand the role and importance of SAS-6 homologues in eukaryotes that do not possess centrioles and might lead to novel insights into organelle evolution.

10. Katan group (UCL) , Pinotsis (Birkbeck College)

The Katan group is focusing on inhibitors for the full-length phospholipase C (PLC) γ from rat. 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. So far, we have collected several data sets of the full-length native protein reaching a maximum resolution of 2.2 Å and in the presence of IP3. We are now screening for more binding ligands and inhibitors. We have determined the structure of the PLC with IP1, while additional density was apparent when ADP was added. While the PLC-IP1 complex is clear and can be refined, no other soaking (with ADP or inhibitors) revealed any clear binding. We are planning to continue our inhibitor screening with modified analogues.