Summer Internship 2021 Report

Qilong Tang (Leilei)
Pembroke College

Location of Internship: The Rosalind Franklin Institute
Research Group: The Structural Biology Group
Lead Academic: Professor James Naismith
Internship Dates: July 12 – September 3, 2021 (8 weeks)

1. Institute information

   The Rosalind Franklin Institute is a new national institute, funded by the UK government through UK Research and Innovation, dedicated to bringing about transformative changes in life science through interdisciplinary research and technology development.

   Their work focuses on five complementary themes (Artificial Intelligence and Informatics, Biological Mass Spectrometry, Correlated Imaging, Next Generation Chemistry, and Structural Biology), which will produce technologies that allow us to see the biological world in new ways, from single molecules to entire systems. This insight will speed up drug design and development, and push forward our understanding of human health and disease.

   Website: https://www.rfi.ac.uk/

2. Overview of internship experience

   After several lectures and tutorials on biophysical methods, I find structural biology very captivating. Therefore, I hoped to work in a structural biology lab to gain some insights.

   In this internship, I worked on nanobodies against the receptor-binding domain (RBD) of the spike protein of SARS-CoV-2. There are two major tasks of my internship. The first one is data analysis, and the second one is structure-solving. The data analysis aims to find a 'good' nanobody binding zone on the RBD in which mutations tend not to occur naturally at high frequencies. The goal of the second task is to solve the structure of a new nanobody, H6.

   For me, the most important thing in this internship is to know what 'doing research' feels like – analysing data, trying to see if there could be any pattern, trying to use biochemistry knowledge to give possible answers etc. Getting to know some practical skills comes second. (Side note: doing a structural biology internship really helps me with my paper I and VI)

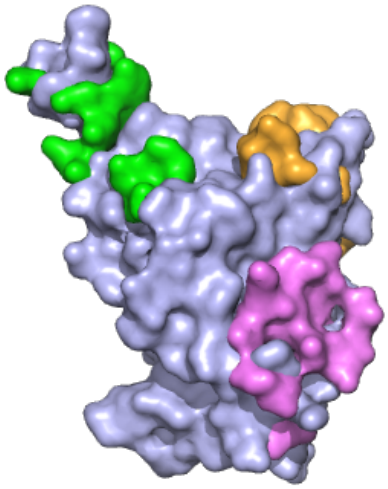
3. Details of internship experience

Data analysis

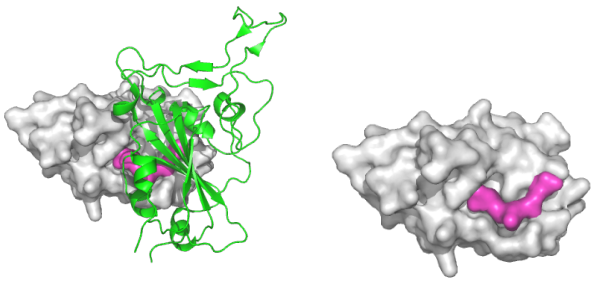
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the seventh coronavirus that infects humans[1]. SARS-CoV, MERS-CoV and SARS-CoV-2 can all cause severe diseases[1]. Although vaccines are being developed, it is still essential to develop a treatment to alleviate the symptoms of COVID-19. In previous studies by Schoof et al.[2], monoclonal antibodies that bind to the spike protein of SARS-CoV-2 showed therapeutic promise but must be produced in mammalian cells and needed to be delivered intravenously. In comparison, nanobodies which are single-domain antibodies can be produced more easily and cheaply in bacteria or yeast, and their stability may enable aerosol delivery.
SARS-CoV-2 binds to the angiotensin-converting enzyme 2 (ACE2) using its receptor-binding domain (RBD) at the cell surface hence invade the cell and cause disease. Antibodies (including nanobodies) against SARS-CoV-2 neutralise the virus by binding to the RBD and preventing the virus from binding to ACE2. However, the SARS-CoV-2 is constantly mutating, and nanobodies can lose their binding to the RBD when specific mutations occur in the RBD. It is, therefore, important to analyse the binding patterns of nanobodies to RBD and find out whether there could be a region on the RBD that is relatively constant. The nanobodies that bind to this constant region should be less susceptible to mutations on the RBD. These nanobodies are therefore more potent to be developed as a treatment for SARS-CoV-2 and its variants.

To find the constant region (the 'good' zone), I found the structures of 30 nanobodies on the PDB (protein data bank). I analysed the interactions between the RBD and the nanobodies using PISA-EBI (PDBePISA). According to the results analysed by PISA-EBI, the interactions between the RBD and nanobody include hydrogen bonds and salt bridges. While hydrogen bonds are found in all the nanobody-RBD structures, salt bridges are found in some nanobody-RBD structures. The nanobodies were then grouped into different zones according to the residues on RBD involved in the hydrogen bonds they made. The nanobodies were subdivided into three 'modes' based on the orientation of their bindings to the RBD. The zone with the least naturally occurring mutations thus is found.

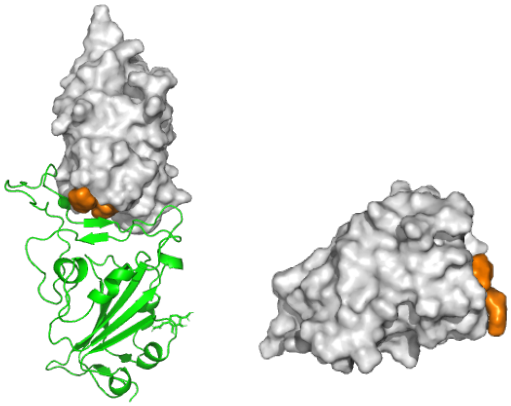
Figure.1a RBD with zone 1, zone 2, zone 3 highlighted in green, pink, orange, respectively.



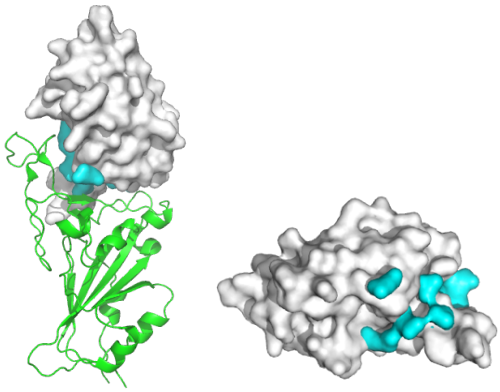
1b



1c



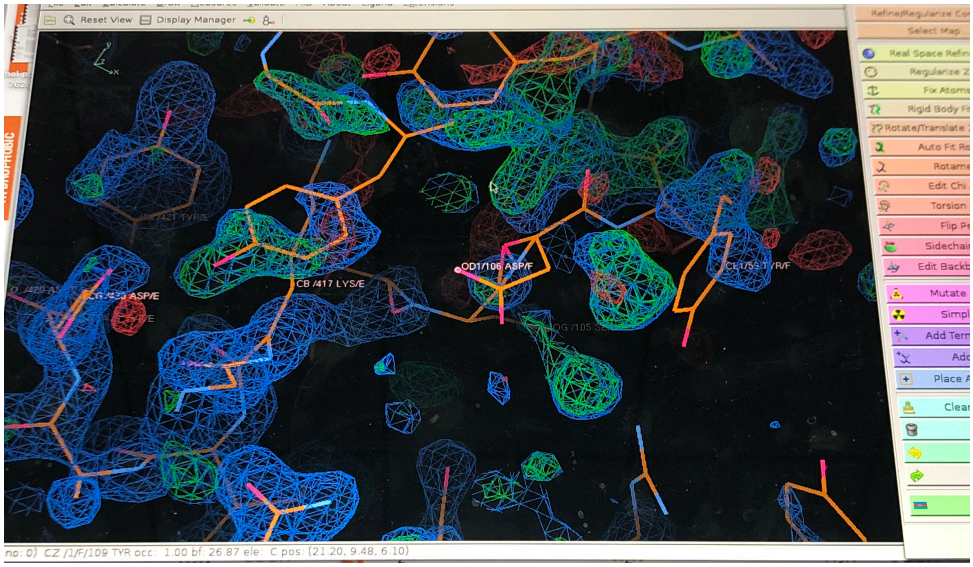
1d



Structure-solving

I learned to solve the structure of the new nanobody using CCP4i and COOT (software).
CCP4i: molecular replacement, Refmac

COOT: manually refine the structure of the protein



After several rounds of structure refinement, one needs to add water molecules into the model and obtain the TSL parameters and weight to get a better model.

TLS motion determination
Website: http://skuld.bmsc.washington.edu/~tlsmd/

The best R-free value I managed to achieve is 0.2003

Other experiments
Data Collection for X-ray crystallography (essential structural biology skills))
I shadowed on a postdoc of how she obtained the electron density map from X-ray crystallography on a nanobody complex. It was a unique experience to see modern data collecting, where everything was finished remotely.

Protein purification
Immobilized Metal Affinity Chromatography, IMAC (Ni^+ column)
Size Exclusion Chromatography (gel filtration chromatography)
I purified the RBD of the delta variant of SARS-CoV-2 expressed by mammalian cells using nickel column followed by gel filtration chromatography.

Having done the protein purification in a real lab, I better understood the considerations I need to take while designing a protein to purify as well as in the protein purification process.

ITC (isothermal titration calorimetry)
Isothermal Titration Calorimetry (ITC) is a technique used in quantitative studies of a wide variety of biomolecular interactions.

The device I used to carry out ITC was MicroCal PEAQ-ITC, and I used it to measure the binding affinity (dissociation constant, K_D) of a new nanobody to the RBD, as well as the binding affinity between ForT (The enzyme ForT catalyses C–C bond formation between 5′-phosphoribosyl-1′-pyrophosphate (PRPP) and 4-amino-1H-pyrazole-3,5-dicarboxylate to make a key intermediate in the biosynthesis of formycin A 5′-phosphate by Streptomyces kaniharaensis.) and its substrates.

This part helps me better understand the principles, applications, advantages and disadvantages of ITC, which is a biophysical technique often appear on Paper I in biochemistry part I exam.

LC-MS (liquid chromatography-mass spectrometry)

4. References

1. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nature Medicine* **26**, (2020).
2. Schoof, M. *et al.* An ultrapotent synthetic nanobody neutralises SARS-CoV-2 by stabilising inactive Spike. *Science* (2020) doi:10.1126/science.abe3255.