Industrial CASE Studentship 2021
(3-year Studentship)

Academic Supervisor: **Prof Andrew C R Martin**
Title: ***Bioinformatics tools for T-cell receptor based drugs***
Partner organization: **Immunocore, Ltd.**

1. Proposed Project details

Aims and Objectives We have developed a range of software and databases for analysis of antibody sequence and structure. In particular, abYsis is widely used online and is successfully sold to companies for in-house use. Antibodies have shown a huge degree of success as novel therapeutics and the related T-Cell Receptors (TCRs) are now being exploited. We wish to repurpose methods we have developed for antibodies for use with TCRs as well as developing new approaches and performing new analyses. Software will be developed to apply standard numbering to TCRs, such that sequence and structure features can be analyzed. Tools will be developed to store and analyze public sequence and structure data with a view to producing software that can be used in the development of TCR-based drugs.

Crystal structures will be analyzed to look at domain packing which can be exploited to improve TCR modelling. Huge nextgen TCR sequence datasets will then be analyzed taking a 'big data' approach storing sequence data, analyzing residue distributions and adding annotations, triaging sequences at each step to reduce unnecessary analysis. Triage will include screening out very unusual sequences and those with likely immunogenicity or developability issues. A basic web interface will be developed to allow the data to be explored and new sequences to be analyzed.

Fit to MRC Remit (MRC strategy document quotations are shown in italics) 1. The MRC *supports methodology development* and *promotes Bio and health informatics and computational modelling.* We will develop methods for analysis of TCR sequence and structure for application to the development of new biologic drugs. 2. The MRC is promoting *Precision medicine […] to better target interventions to an individual*. Like other targeted therapies, TCR drugs are being developed for use in specific cohorts of patients (e.g. with particular mutations or over-expressing particular proteins in cancer). 3. The MRC recognizes the importance of the *Antimicrobial resistance […] threat and opportunities for intervention*. TCR drugs are being applied to infection. 4. The MRC is investing in *Discovery science […] to push the frontiers of knowledge […] at different scales […] inspired by clear health problems, knowledge gaps or opportunities for interventions […] supporting a strong pipeline of academic translation*. The proposed analyses will address frontiers of knowledge and novel software will lead to academic translation through licensing the software. 5. The MRC recognizes the importance of *Data science vision and strategy.* The huge volumes of data that come from NGS analysis of TCR repertoires require specialist databases and approaches to fast data processing. 6. The MRC's *ultimate aim is to build a sustainable capability in health and biomedical data science*. The student will be trained in the biology of TCRs, the nature of NGS sequence data and in big-data databases and software development.

Rationale and importance Immunocore has recognized the application of TCR-based therapies in three hugely important health areas: oncology, infection, and autoimmune/inflammatory disease. As with other targeted therapies, the rationale is to reduce systemic effects while improving therapeutic effects. This project would perform analyses and introduce new software tools, based on very successful and popular analyses and tools used for antibodies, to improve the TCR drug development pipeline.

2. Project plan

Year 1 Our 'abnum' software for applying standard numbering schemes to antibodies [Abhinandan & Martin (2008) Molecular Immunology 45:3832-3839] will be modified for use with TCRs. The software will be tested using pre-numbered TCR sequences and refined as necessary. Publication expected. A 3-month visit to Immunocore to learn about TCR-drugs, problems during TCR-drug development, and work with methods used at Immuncore for development of display libraries and nextgen library sequencing and generate datasets.

Year 2 A basic pipeline will be developed to populate a NoSQL database (e.g. MongoDB) with numbered sequence and structure data. Analysis of domain packing using ~50 PDB structures using the approach developed for antibodies [Abhinandan & Martin (2010) PEDS **23:**689-697]; publication expected. Based on information gained from the placement, triage pipeline development will start.

Year 3 Continued development of the triage pipeline which will (1) remove very unusual sequences missing key features, (2) remove sequences with selected post-translational modifications, (3) identify and remove sequences with unusual length CDR-3s, (4) apply numbering, (5) map sequences to a template structure and remove those with surface clusters of unusual residues, (6) remove sequences with surface clusters of unusual residues, (7) implement any other triage steps proposed by Immunocore. Populate with Immuncore and/or public datasets. Develop a simple web interface. Publication expected**.**

3. Data sharing

Andrew Martin's antibody web site (www.bioinf.org.uk/abs/) and abYsis pages (www.abysis.org) are very widely used with some 2.5 million hits in the last year. Access to software developed for analysis of TCRs will be made available through these pages. It is expected that at least 3 papers will be published in appropriate journals (TCR numbering tool; analysis of domain packing; the overall pipeline). Where appropriate, software will be made open source and made available through www.bioinf.org.uk and GitHub (@ACRMGroup). It is expected that other software will be made available using the same model as in place for abYsis – free access via the web and commercially available licences (through UCL Business) for in-house use.

4. Academic research environment, training and support (UCL)

SMB at UCL provides a leading-edge research environment, superb access to technical resources, and world-recognized research expertise in areas such as bacterial secretion, aging, signal transduction, protein folding, drug metabolism and cellular protein trafficking as well as bioinformatics. Recent graduates have become researchers at leading UK and international universities and research centres as well as taking senior roles in industry with companies such as Boehringer Ingelheim, Novartis, Celltech and Genzyme.

Day-to-day supervision is provided by the primary supervisor who generally meets with the student on a weekly basis, as well as group meetings, and is available whenever required to answer questions and assist with problems.

We place a strong emphasis on high-quality interdisciplinary student training, and development of transferable skills. Students attend weekly seminars given by leading researchers and meet these external speakers over lunch. They attend weekly 'Friday Wraps', presented by students or post-docs; they give presentations to the department and produce posters. Attendance at conferences is expected and presentations (orally or of posters) are strongly encouraged. In addition, the ISMB (of which SMB is a part) organizes research retreats and symposia providing other opportunities for networking for development of their future research and career. All these foster presentation and communication skills.

Students are required to gain at least an average of 20 training points per year. Seminars, conferences and demonstrating for practicals contribute to these. In addition many training courses are available within the division and through the UCL Doctoral School including statistics and transferable skills ranging from generic research techniques and ethics, through document preparation with LaTeX, to presentation, writing skills and programming. UCL actively participates in 'Software Carpentry' workshops and students are able to attend the Bioinformatics MSc lectures at Birkbeck. Students have ready access to a full suite of electronic support facilities including the library and e-journals.

UCL has exceptional computing facilities including group and departmental servers and university-wide super-computing (approaching 10,000 cores as well as GPUs). Within AM's group, as well as desktop workstations, facilities include main file- and compute-servers and a web-server (total 72 cores) with ~60TB of disk space with daily local and weekly remote backups.

5. Non-academic partner organisation research environment, training and support

Immunocore is a well-established Oxfordshire company already actively involved in supporting multiple iCASE studentships with Oxford University, Southampton University and Birmingham University. The PhD students involved in these projects regularly visit Immunocore for placements and have implemented an Immunocore PhD network providing a supportive and interactive community for PhD students. Immunocore aims to develop soluble T cell receptors to treat human disease including cancer, autoimmunity and infectious diseases and actively seeks out academic partners to explore biologic systems relevant to their projects.

Immunocore will provide training and mentorship to the student in the areas of TCR discovery, TCR structure and developability of TCRs as soluble therapeutic agents. We expect that the student will spend at least 3 months at Immunocore to generate TCR sequence datasets, and to better understand the developability needs of our soluble TCR platform.

Immunocore can provide training in molecular cloning, protein expression, library preparation and mutagenesis, screening for binding, BIAcore, and structural studies as and when it is appropriate for the project. These practical expertise are particularly valuable as they will be useful in a number of related settings. Moreover, the student will be involved in weekly team meetings and will have the opportunity to present and disseminate their findings under the mentorship of experienced leaders at Immunocore. Lastly, the student will be exposed to strategic decision making from the point of view of a company that makes decisions not only based on the science, but also on the business case and unmet need of patients. This experience will provide a unique opportunity for personal development and the acquisition of translational skills beyond what is accessible in a purely academic setting.

The team at Immunocore have specific expertise in these areas which is hard to reproduce in most other environments, including in purely academic settings. Finally, the development of soluble TCRs, which this project will contribute towards, could be used downstream for cancer therapy, which is a very exciting opportunity for the student. Thus, the training is not only in individual techniques but also in the development of a pipeline from basic science to translation.

6. Management and monitoring

In SMB, the lead role in monitoring and managing progress on a daily basis is taken by the primary supervisor. In addition, students' progress is monitored via a thesis committee consisting of the primary and secondary supervisors and an independent 'thesis chair'. Meetings occur at regular intervals and there are prescribed activities (oral presentations and/or written work) for each meeting. The primary supervisor is often asked to step out of meetings for a few minutes such that the thesis chair can receive verbal feedback from the student regarding supervision and research progress. The thesis chair writes a brief report which is agreed by the supervisors and the student. Meetings and reports are recorded in the student's online research log. Members of the thesis committee cannot act as an examiner for the student, but the second supervisor and thesis chair are responsible for examining the student for their upgrade to PhD status at around 1 year. The departmental graduate tutor has an overall monitoring role and will intervene to liaise with students and supervisors to try to resolve any problems with a student's progress, or with supervision. The graduate research administrator provides another point of contact for administrative issues and sends out reminders for meetings and reports, as necessary.

7. The collaboration

The project would not be possible without the collaboration since it will be led very clearly by the needs of Immunocore in developing TCR-based drugs. The project is relevant to Immunocore and the rest of the sector because it will open up existing software, widely used in antibody drug development, to the TCR field. In antibody drug development, licencees have estimated that abYsis saves them between 10 and 50% of their time in developing new candidate drugs.

During the placement the student will identify inhouse datasets (TCR-pHLA structures and sequencing data) that can be used to build the bioinformatic tools detailed in the project and will work towards understanding the needs of Immunocore in terms of developability of soluble TCR-based drugs.

This placement will offer added value to the student and the project because it will enable the student to familiarise themselves with the TCR technology, deep-dive into the practical issues that this project may address, and learn new techniques and strategic thinking. The placement will offer added value to Immunocore by enabling strong links to be developed with the student and the academic supervisory team, will provide opportunities and emphasis for new thinking about the technology, and the transfer of bioinformatic skills between the student and Immunocore staff. Additionally, publications are of key importance to Immunocore's strategy.

Immunocore has a track record in TCR drug development with 3 molecules currently in the clinic (one in phase 3) and several other programmes in late stages of development. Immunocore is current supporting 3 other iCASE studentships and has an internal PhD programme that currently supports 3 PhD students. They are committed to providing expertise and guidance throughout the project and facilities during the placement.

The broad project goals have been discussed elsewhere in this application. IP in the software will remain with UCL, but a free licence will be provided to Immunocore. All IP relating to specific examples and drugs that may be developed during the project (or using software developed during the project) will remain with Immunocore. A detailed agreement will be put in place later.

The collaboration is key to guiding the requirements for the project and the software and in teaching the student about TCR-based drug discovery. It will benefit hugely from the expertise and datasets available at Immunocore. The placement is planned for the first year in cases circumstances should change and public datasets can be used for development.