The digital experiment

Microsoft applies its computing skills to biology

It is well known that the development of a new medicine is a long and costly process. What may be less well known is that Microsoft Corp, the founder of software that made the personal computer universal, is actively working to apply its digital skills to make laboratory practice in the life sciences more rational. The expectation is that drug development will become less costly and more predictable in the future.

A project known as Station B aims to replace trial-and-error experimentation with a system that is governed by programming languages and machine learning, enabling the compilation of biological knowledge that scientists can use to inform their decision making.

It is being led by Andrew Phillips, head of the Biological Computation Group at Microsoft Research, a subsidiary of US Microsoft Corp. Microsoft Research is located in Cambridge, UK. Dr Phillips holds a PhD in computer science from Imperial College London and joined the Cambridge group in 2005 in order to conduct research at the intersection of computer science and biology. Together with colleagues, he has been working for years on methods for understanding and programming what happens in biological systems, including inside human cells.

Station B was launched on 12 March after more than a decade of research. It is the first real world test of whether the marrying of computer technology and biology will work. The project's first partners are Princeton University in the US, the UK gene therapy company Oxford Biomedica Plc and the technology concern Synthace Ltd.

> ***"A key goal of the project is to understand how biological systems work so that we can programme them more effectively."***
>
> *Andrew Phillips*

"A key goal of the project is to understand how biological systems work so that we can programme them more effectively. From an engineering perspective this means understanding how biological systems, including living cells, perform information processing by 'reverse engineering' these systems," Dr Phillips told *MedNous* in an email correspondence.

"A second challenge is finding a way of performing biological experiments more systematically and reliably. A major problem in the life sciences right now is that most biological experiments are not even reproducible. Programming biology today is done by trial and error. The goal of Station B is to develop a platform that can be used by partners and collaborators to programme biology in a way that is more systematic and predictable," he said.

To test the project's value to the healthcare system and to patients, Microsoft has chosen to apply its theories to the global problem of antibiotic resistance and the emerging technology of gene therapy. The new partnership with Princeton is aimed at understanding the formation of biofilms – an association of microorganisms on living or non-living surfaces. Bacterial biofilm is infectious and can result in serious hospital infections. The partnership involves a collaboration with Bonnie Bassler, chair of Princeton's Department of Molecular Biology. Professor Bassler's research focus is on cell-to-cell communication in bacteria. The Station B platform will be used in the Bassler lab to construct and test different versions of proteins that play a role in biofilm formation, and try and understand how bacteria develop resistance to antibiotics. This means programming microbial systems, disturbing these systems, and then measuring the effects of these disturbances in order to try and reverse engineer how bacteria communicate.

The second partner, Oxford Biomedica, plays an important role in bringing the next generation of cell based gene therapies to the market through a manufacturing partnership with Novartis. Novartis, drawing on research from the University of Pennsylvania, developed Kymriah, the first cell based gene therapy to be approved by the US Food and Drug Administration. This approval was awarded in 2017 based on a clinical study in paediatric and young adult patients with acute lymphoblastic leukaemia. It showed an overall remission rate of 83%. Since its market launch, Kymriah has also been approved to treat large B cell lymphoma.

The therapy is effective, but it is also expensive with a list price of £282,000 in the UK. This arises, at least in part, from the cost of development which includes a complex manufacturing process. Kymriah is an autologous gene therapy based on the engineering of a patient's own T cells. The T cells are extracted from a patient and transduced with a lentiviral vector that encodes a chimeric antigen receptor (CAR) directed against the CD19 protein on cancer cells.

Oxford Biomedica is responsible for producing the lentiviral vector for Kymriah. Lenviral vectors are derived from the human immunodeficiency virus which has been disabled. The vector is grown in a human cell line followed by a purification process to yield the vector substance. This substance undergoes sterile filtration, concentration and filling to obtain the vector product[1].

In an interview, Oxford BioMedica's chief business officer Jason Slingsby, explained the procedure this way. "We use a cell line that is derived from human kidney cells, which is the default cell in the industry. You need to programme that cell to make something that it doesn't normally make,

in this case, a highly minimal, highly reduced viral vector which is a one-way delivery vehicle. You do that by putting stretches of DNA into that cell and that DNA makes proteins and those proteins assemble into a lentiviral vector particle which is much smaller than a cell."

Quality is an issue throughout the manufacturing process. "We are looking for lots of these vector particles but low levels of other kinds of DNA or other kinds of proteins. We want to have as many healthy and active particles [as possible]. These are complex things, much more complex than an antibody-based drug. They have membranes around them rather like a cell so you want to make sure that they are handled gently and are active and able to do their job," he said.

Under the Station B project, Oxford Biomedica will use software developed by Synthace to create large datasets that can be transferred to Microsoft for analysis. The goal will be to come up with computer models and algorithms to advance the next generation of gene therapy delivery technology. These datasets will come from actual laboratory experiments but will be too large for any one scientist or team of scientists to handle.

Strategic resource

Mr Slingsby said Oxford Biomedica sees the analysis of the large datasets as a strategic resource. "The goal ultimately as a business is to use these more rational insights from data analysis to allow us to improve the efficiency with which we can make lentiviral vectors," he commented. In addition to supplying Novartis with lentiviral vector for Kymriah, Oxford Biomedica has a gene therapy partnership with Sanofi SA for haemophilia A and B and with three UK universities and Boehringer Ingelheim GmbH to produce a gene therapy for cystic fibrosis. Both of these projects are at any early stage, but will require large amounts of vector product when they eventually go into clinical testing. Improving the efficiency of production should lower costs and make the future products more accessible to patients, the executive commented.

The third partner, Synthace, is a London-based technology company which has developed software that can be used in life science laboratories. It is the glue that holds the Station B project together because the product, Antha, can be adapted to many laboratory situations. Microsoft has been using Antha since 2016 as part of a strategic partnership with Synthace. Separately, Oxford Biomedica has been using Antha since 2017 as part of a four-way collaboration co-funded by the UK government to improve cell and gene therapy manufacturing capabilities in the UK. Besides Oxford Biomedica, the partners of this public sector initiative are Synthace, the Cell and Gene Therapy Catapult and Stratophase Ltd, a biotech services company.

Meanwhile, Microsoft has deployed Antha, plus some of its own software, to the Bassler laboratory at Princeton to help with the antibiotic project.

Antha is a tool in the whole process, essentially helping laboratory scientists programme robots to conduct complex experiments. It operates on open-source language that describes working practices that can then be translated into a form of automated command. Importantly the software can be used on hardware from many different vendors.

In an interview, Markus Gershater, chief scientific officer of Synthace, explained the overall role of Antha in the Station B project. "For artificial intelligence to have something to work on it needs data. And a lot of biological data at the moment is unstructured; it is poorly documented. Before we can really unleash the full potential of AI in a biological setting we need a way of generating large structured datasets. This is what the different people in the collaboration are using Antha for," he said. Once these datasets are generated and structured, then Microsoft can use machine learning to help interpret the results of the experiments.

Human serendipity?

How does this affect human serendipity which has been a source of many of the great scientific advances in the past? Dr Gershater said that the software is enabling people to test many more hypotheses than would otherwise have been the case. Automation also covers a gap in many laboratories when a researcher may be away or otherwise occupied and may miss some critical observations. "I think that for many, the jury is still out as to whether we should be aiming for full automation or not," he said.

He added: "I think hybrid automated-manual workflows seem very powerful, certainly for the stage that we are at the moment. What this means is that you get the robot to do the most complex parts, or the most tedious parts, or the things that are likely to be error prone. But if you were to visit our labs you would still see a lot of people working manually as well."

In a podcast issued by Microsoft Research on 13 March, Dr Phillips outlined what he sees as the trajectory for biological computing. He cited the production of insulin in the 1970s as the first example of programming a microorganism to make a medicine. This was followed by the sequencing of the human genome and most recently by genome editing and the creation of tools such as CRISPR-Cas9.

"We have this underlying technology that's allowing us to manipulate DNA, read, write and edit it. And that's also underpinned this technological growth in our ability to programme biology," he said.

There are many global health problems to which this technology could apply such as the resistance of bacteria to current antibiotics. Technology can play a role in helping scientists understand and overcome resistance with new treatments. A start, is making the drug development process more systematic and predictable, he noted.

Reference:

1. European assessment report, Kymriah, EMA/485563/2018, Committee for Medicinal Products for Human Use, 28 June 2018.

This article was written by the *MedNous* editor on the basis of a literature review and interviews with key participants in the Station B project.

