

# MAPPING EVERY PROTEIN IN THE HUMAN BODY

ALMOST TWENTY YEARS AGO, THE HUMAN GENOME PROJECT COMPILED THE CODE OF EVERY GENE IN THE HUMAN BODY. NOW, PROFESSOR NEIL KELLEHER, OF NORTHWESTERN UNIVERSITY IN ILLINOIS IN THE US, WANTS TO DO THE SAME THING FOR EVERY PROTEIN IN EVERY DIFFERENT KIND OF CELL. IT IS A MAMMOTH TASK, BUT HE BELIEVES THAT WITH ENOUGH AMBITION AND COLLABORATION, IT CAN BE DONE – AND IT PROMISES TO BRING HUGE BENEFITS FOR HUMAN HEALTH

## TALK LIKE A BIOSCIENTIST

**DNA** – a nucleic acid that carries the genetic information of living organisms

**GENE** – a DNA sequence that carries the code for a particular protein

**GENOME** – the complete set of genes of any organism

**PROTEIN** – molecules within the body made up of chains of amino acids, coded for by DNA

**PROTEOFORM** – a specific form of a particular protein, that may differ in shape or structure

**PROTEOME** – the complete set of proteins of any organism

**PROTEOMICS** – the study of proteomes and how they function

Proteins have an endless array of possible shapes and structures, which means they can fulfil a huge diversity of functions. They give the body structure, are integral to the functioning and replication of cells, and are ruthlessly efficient at protecting us from disease. Understanding how they work, therefore, can bring huge benefits for medicine and related fields.

Professor Neil Kelleher, of Northwestern University in Illinois in the US, has his sights set on bringing about a major leap forward in

our understanding of proteins, by cataloguing every single protein in the human body. The implications of such a project are huge. “The Human Proteoform Project will allow us to know ourselves, and to understand our underlying biology with unprecedented precision,” explains Neil.

## THE COMPLEX WORLD OF PROTEINS AND PROTEOFORMS

We already have the human genome mapped out thanks to the Human Genome Project, a massive

undertaking completed in 2003 that mapped the DNA sequence of every gene within the human body. Since genes code for proteins, shouldn't we be able to extrapolate the human proteome from this? In reality, of course, it is not that simple.

The base sequence of a gene tells us the amino acid sequence of the protein that it codes for, but that is only the start of the story. During and after the processes that create a protein from a gene, there can be any number of additional changes made to the protein. For instance, certain conditions within the cell may lead to the addition of extra features such as carbohydrate groups, or even change the entire three-dimensional structure of the protein. This means that proteins with the same underlying amino acid sequence can have drastically different functions, which cannot be predicted from the gene alone. These differing varieties are known as proteoforms.

Uncovering these differences is a main aim of the Human Proteoform Project. But this is no easy task – Neil estimates there are about one billion different proteoforms within the human body. It will take a monumental collaborative effort, not to mention considerable funding and the latest research technology, to pull it off.



## PROFESSOR NEIL KELLEHER

Chemistry, Molecular Biosciences and  
Medicine, Northwestern University,  
Illinois, USA

.....

## FIELD OF RESEARCH

Chemistry, Biochemistry and  
Measurement Science

.....

## RESEARCH PROJECT

Mapping every form of every  
protein within the human body

.....

## FUNDERS

National Institutes of Health,  
select private foundations and  
private companies

## HOW IT WORKS

The first step for the project is to map out the body's 4,000-odd different types of cell. These can be recognised and catalogued according to the unique array of proteins found on each cell surface. Categorising these is important because different conditions within each type of cell can lead to different proteoforms. Once that is complete, the second step involves cataloguing the proteins themselves. This will be achieved by finding their molecular weight – not just a simple weighing-scales measurement, but rather uncovering the identity of every atom within the protein molecule and how they fit together.

The project will use a technique called top-down proteomics to carry out this second step. "Imagine a big pile of gold coins," says Neil. "One of the coins is plastic and has a lighter mass to the others. How do you find it? Do you weigh every coin? Or do you split the pile in half and see which pile is lighter? If you keep splitting the pile like this, you will find the plastic coin faster – this is the 'top down' strategy." Top-down proteomics uses this logic to pinpoint and characterise individual proteoforms in a mixture.

This technique differs from bottom-up proteomics, which involves digesting the proteins first and sorting through the resultant fragments. This can be good for identifying particular proteins, but cannot distinguish between proteoforms because the overall shape and structure of the protein has been lost – so this technique is not suitable for Neil's project.

## NOT EXACTLY SMALL FRY

"This project is very ambitious, and poses a significant career risk for me," says Neil. "It will be fantastic if it is completed, but if it does

not generate sufficient enthusiasm within the scientific community, it could never materialise." In scale, it can be compared to the Human Genome Project. This became economically feasible when the mapping technology became efficient enough to map the genome at a cost of about \$1 per DNA base. Neil believes working on a similar economic model, with a cost of about \$1 per proteoform, should mean that the project could be completed within ten to fifteen years.

This is a fairly long timescale and requires a lot of forward thinking from researchers and funders alike. "We have a worldwide consortium of 400 members, with increased funding coming in too," says Neil. "This makes me hopeful that the Human Proteoform Project may come to pass within this decade, and not the next one!"

## CHANGING LIVES

Given this project will be so massive and expensive, what benefits could it bring to society? Neil believes that a major application is within medicine. A huge range of diseases, from cancer to heart disease, is caused by faulty proteins – in essence, bad proteoforms. By having an atlas of all these unwanted proteoforms, it will become dramatically easier to detect and combat disease. "This project could accelerate our progress towards next-generation technology, leading to significant improvements in human health," says Neil.

It is widely known that the earlier a disease is detected, the easier it is to treat, and this certainly holds true for cancer – yet detecting it is often not easy. For instance, currently, the most common test for prostate cancer involves screening for a protein called prostate-specific antigen (PSA). However, PSA can exist as

many different proteoforms, a number of which are not screened for, so it can commonly be missed. In the future, knowing exactly what these proteoforms are should make the screening test much more sophisticated and accurate.

As well as detection, the project could also help with developing cures. Protein-based drugs are a rapidly expanding form of treatment. They function by detecting and attacking certain proteins – for instance, the proteoforms that exist on the surface of cancer cells. The more proteoforms that are known, the easier it becomes to develop highly specific treatments.

Despite massive advances in medicine in recent decades, a lot of detection and treatment methods are still very generalist. For instance, chemotherapy treatment does not just target cancer cells, and can take a severe toll on the rest of the body too. The Human Proteoform Project has the potential to change this for good, and could revolutionise medicine forever.

Neil's vision reminds us that unwavering ambition and a willingness to take risks can lead people to great things. We wish Neil the best of luck – the future is exciting!



# ABOUT PROTEOMICS

Proteomics is the large-scale study of proteins. It concerns the proteome, which is the name given to the entire set of proteins within an organism. Neil explains what it is like to research in this field.

## WHAT DO YOU FIND MOST CHALLENGING ABOUT PROTEOMICS?

Unlike genomics, we cannot amplify the molecules that we study. This means we are often working with very tiny amounts of protein, which raises difficulties. This is a core challenge of proteomics, but proteins are massively important for biology, so we make it work.

## WHAT HAVE BEEN SOME HIGHLIGHTS OF YOUR WORK IN PROTEOMICS?

Going public about the Human Proteoform Project was a major moment for me. Additionally, I get a lot of pleasure from discovering the precise traits of proteins in the human body. This is a 'positive feedback loop' that has sustained my dedication over decades, despite obstacles along the way.

## WHAT TYPE OF PEOPLE DO YOU COLLABORATE WITH?

I collaborate with a very wide range of people. Doctors, clinician-scientists, biochemists, computer scientists, business leaders and

research directors all take an interest in my work. Pooling our skills and networks is a fantastic way to make progress.

## WHAT WILL BE THE BIGGEST FOCUS FOR THE NEXT GENERATION OF SCIENTISTS WITHIN PROTEOMICS?

Once all the proteoforms are mapped, the next big challenge is deciphering the exact function of each one and how they can be modified.

## EXPLORE A CAREER IN BIOSCIENCE

- Disciplines such as biochemistry, molecular biology and biomedicine have the potential to lead to a career in proteomics.
- Many universities offer public outreach schemes that help young people get a 'taster' of working in the lab. For instance, Northwestern University, where Neil works, sees many school students visit the campus and engage in summer research.
- According to Payscale.com, the average salary for a molecular biologist is \$59k.

## PATHWAY FROM SCHOOL TO BIOSCIENCE

Neil recommends taking subjects like chemistry and biology, and also computer programming and statistics. Physics and areas of mathematics can also be useful. Once within university, Neil recommends taking courses in cell biology, organic chemistry and genetics.



*Jeannie Camarillo has been a researcher at the Kelleher Research Group for two years and is set to become a leader in the proteomics field.*



*Phil Compton is a staff member with the Kelleher Research Group. His specialty is in construction and repair of the mass spectrometry devices used to perform top down proteomics.*



*Dr Ben Garcia is Head of Biochemistry and Molecular Biophysics and leads the Garcia Lab at Washington University in St. Louis. Ben conducted postdoctoral research with Neil and now specialises in developing novel mass spectrometry methods.*





# HOW DID NEIL BECOME A BIOSCIENTIST?

## WHAT WERE YOUR INTERESTS AS A CHILD?

On the academic side, I enjoyed science and woodworking. I liked working with my hands, which helped lead me to a career in lab research. On the sports side, I enjoyed wrestling and golf, which helped me learn fortitude and mental toughness.

## WHAT INSPIRED YOU TO BECOME A SCIENTIST?

At sixteen, I took an internship at the chemistry lab of the Weyerhaeuser paper company. It was there I learned that chemists

have the solutions to real-world problems – and I even helped solve a small one myself!

## WHAT ATTRIBUTES HAVE LED YOU TOWARDS SUCCESS?

Persistence, self-belief and an ability to embrace risk.

## HOW DO YOU OVERCOME OBSTACLES IN YOUR WORK?

I find it useful to accelerate towards obstacles – I even revel in the struggle. That has helped me overcome substantial barriers and paved the way to success.

## WHAT IS THE PROUDEST MOMENT OF YOUR CAREER?

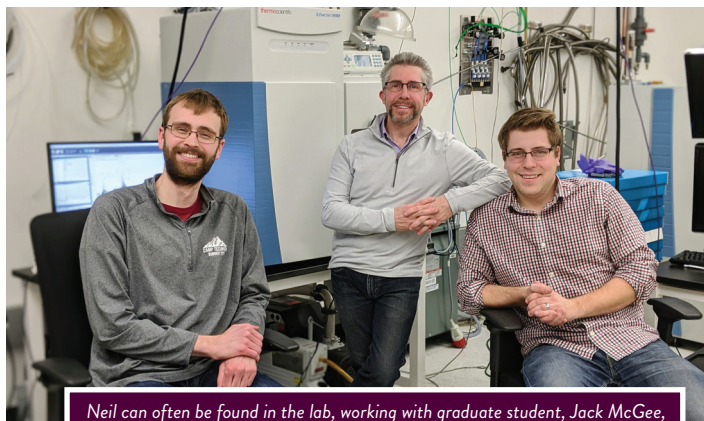
I believe the highlight of my career is still to come! Having said that, the publications of some of my best research papers have been great moments. I also place great value on training the next generation of scientists.

## HOW DO YOU SWITCH OFF FROM WORK?

I take pleasure in exercise, especially swimming, golf and jogging.

## NEIL'S TOP TIPS

- 01** Don't follow directly in the footsteps of others. Forging your own path is harder, takes a while – but positions you well in Life!
- 02** Do find good role models and learn how to learn from them (#networking).
- 03** Embrace risk – ambition feeds on it!



Neil can often be found in the lab, working with graduate student, Jack McGee, and staff member, Jared Kafader.