

HOW DO CELLS CONTROL PROTEIN LEVELS?

THE GENETIC INFORMATION THAT DEFINES EACH OF US IS STORED IN THE DNA WITHIN OUR CELLS. DR GEORG KUSTATSCHER, FROM THE UNIVERSITY OF EDINBURGH IN THE UK, IS RESEARCHING HOW THIS INFORMATION IS TRANSLATED INTO THE PROTEINS THAT ARE ESSENTIAL FOR LIFE. HE WANTS TO DISCOVER HOW OUR CELLS REGULATE THE AMOUNT OF PROTEINS PRODUCED, AND WHAT HAPPENS WHEN THIS PROCESS GOES WRONG. HOPEFULLY, THE ANSWERS WILL INCREASE OUR UNDERSTANDING OF CANCER AND HELP TO DEVELOP NEW TREATMENTS FOR CANCER PATIENTS

TALK LIKE A PROTEOMICS RESEARCHER

DNA – the molecule that stores the genetic information in our cells

GENE – a section of DNA containing genetic information and instructions to produce proteins

MACHINE LEARNING – computer programmes which learn to perform complex tasks without being explicitly programmed to do so

mRNA – the molecule that copies the genetic information when it is used to make proteins

PEPTIDE – a section of protein produced, for example, when proteins are digested by enzymes

PROTEINS – biological molecules that contribute to the activities in our cells

PROTEOME – the complete set of all the thousands of proteins that make up our body

PROTEOMICS – the large-scale study of proteins

The cells in your body only have a diameter of ~0.01 mm, and yet every one contains two metres of DNA! These two-metre-strands of DNA are tightly coiled inside the nucleus of every cell and contain all the genetic information that your cells need to make you who you are.

Genes are small stretches of DNA that encode for different functions in the body. Every cell should have two copies of each gene, one inherited from each biological parent, and humans have over 20,000 different genes in total. These genes also carry all the information which the cell needs to create different proteins, the biomolecules that carry out the actual biological activities in a cell. These proteins range from enzymes that break down nutrients to antibodies that fight bacteria to regulator molecules that control cell division.

But sometimes, the biological processes required to create proteins from DNA go wrong. And this is what Dr Georg Kustatscher is investigating. Georg is a proteomics researcher at the University of Edinburgh, where he is studying how our cells regulate their protein levels.


HOW ARE PROTEINS PRODUCED?

The process of turning a gene into a protein

is called gene expression and consists of two stages. In the first stage, transcription, genes are copied into molecules called 'messenger RNA' (mRNA). mRNA is similar to DNA, but it only contains the single gene that was copied, rather than the whole two-metre DNA sequence. mRNA travels from the cell nucleus, where the DNA is stored, to the cytoplasmic part of the cell, where most biological activities take place. In the second stage, translation, the mRNA is converted into proteins by molecular machines known as ribosomes.

"This basic relationship of how DNA turns into mRNA and then into proteins has been known for 60 years," says Georg, "but where it really gets interesting is how these processes are controlled."

The genetic information stored in your DNA is the same in every cell in your body. So a liver cell and a neuron contain exactly the same genes, yet they have completely different functions in the body and so need different proteins to operate. About half of our 20,000 genes are 'always on', because they produce proteins that all cells need at all times. The remaining genes produce proteins that are only needed by certain types of cells, or in certain situations, such as to repair a cell when it gets damaged.



This means cells must control which proteins they produce, and in what amounts, depending on their function in the body and their situation. This can be done either by alternating the amount of mRNA transcribed from DNA, or by altering the amount of protein translated from mRNA.

WHAT GOES WRONG IN CANCER CELLS?

“In a living cell, the proteome is very dynamic,” says Georg. “Proteins are constantly being made and broken down. The balance of these processes is essential for healthy cells, and proteome imbalance is a feature of many diseases, including cancer.”

In cancer cells, the relationship between DNA, mRNA and proteins can be out of balance. A common problem is gene-copy number alterations. Normal cells have two copies of each gene, but some cells lose one of those copies, or they acquire additional copies of the same gene. Gene additions or deletions should result in more or less mRNA being produced, and therefore an excess or shortage of certain proteins. However, there is an unknown mechanism that normally protects cells against the negative consequences of this. This mechanism ensures that the protein levels of deleted or amplified genes remain unaffected. “But for some genes in some situations, this compensation mechanism doesn’t work,” says Georg. He thinks that when this happens, these genetic mutations cause cells to become cancerous. Georg’s team is researching how this compensation mechanism works in the hope that they can find ways to activate or modify it to help cure cancer.

HOW DOES GEORG STUDY PROTEINS?

To investigate how proteins are produced, Georg’s team need to measure the amounts and types of different proteins in cells. First, they extract the proteins by dissolving the entire cell in a strong detergent. Next, the proteins are digested into smaller chunks (known as peptides) using enzymes that are naturally found in the intestine. These peptides

are then analysed using a mass spectrometer which measures the amount of each peptide and its mass. It then splits each peptide into even smaller fragments and determines their masses as well, creating a ‘fingerprint’ which is unique for each peptide. As each protein produces a unique combination of peptides, the data from the mass spectrometer can be used to identify which proteins were present in the cells, and in what amounts.

THE IMPORTANCE OF MACHINE LEARNING

In the past, researchers had to identify proteins by hand. Today, mass spectrometers can collect data on thousands of proteins in one single sample, so examining each result by hand would be far too time consuming. To help analyse the big data sets that proteomics experiments produce, researchers are using ‘machine learning’. To identify differences between proteomics samples from healthy individuals and patients with a certain disease, researchers label some example cases. The computer uses these examples to ‘learn’ the differences between the samples and identify which of the thousands of proteins are critical for distinguishing the two sample types. The trained algorithm can then assign unknown samples to one of the two classes, helping to diagnose potentially ill people.

SO HOW WILL GEORG DETERMINE HOW CELLS REGULATE PROTEIN LEVELS?

“Our first important question is about the nature of the regulatory processes,” says Georg. “If a cell needs to increase the concentration of a protein, will it increase the rate of transcription (mRNA synthesis) or that of translation (protein synthesis)?”

To answer this question, Georg will measure the levels of proteins in the cell, as well as the amount of mRNA and DNA. If high levels of mRNA correlate with high levels of the protein that it produces, this would suggest that the cells are controlling the protein level through transcription, by altering the amount of



DR GEORG KUSTATSCHER

Institute for Quantitative Biology,
Biochemistry and Biotechnology,
University of Edinburgh, UK

FIELD OF RESEARCH

Proteomics, Cancer Biology

RESEARCH PROJECT

Investigating how cells regulate protein levels and fail to do so in cancer

FUNDER

Medical Research Council

This work was supported by the MRC, under grant number MR/T03050X/1. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the MRC.

mRNA produced. Conversely, if high protein concentrations do not correlate with high mRNA levels, this would suggest that the cells control the protein level through translation, by altering the amount of proteins generated.

Mass spectrometry and machine learning will allow Georg to measure thousands of different proteins and mRNA at the same time, so he can work out which proteins are controlled by which mechanism. Comparing the mechanisms in healthy cells and cancer cells will help scientists to understand how protein production is going wrong in cells with cancer. Hopefully, they will then be able to develop new treatments to correct these processes in cancer cells.

Georg’s research shows how fundamental research into biological processes allows us to think about new approaches to treating diseases such as cancer. With knowledge of how cells control protein levels, and how this goes wrong in cancer cells, future scientists may be able to correct the causes of cancer.

ABOUT PROTEOMICS

Proteomics is the study of proteins on a large scale. Our body contains many thousands of different proteins; proteomics involves investigating how these proteins are produced, used and broken down inside cells. The field combines ideas from areas such as medicine, biology, chemistry, engineering and computer science, so it is a very diverse and rapidly growing research area.

Georg tells us more about the field of proteomics:

WHY IS IT IMPORTANT THAT BIOLOGISTS CONSIDER THE WHOLE CELL AS A COMPLEX SYSTEM, RATHER THAN STUDYING INDIVIDUAL PROTEINS IN ISOLATION?

Until recently, biologists believed that a complex system like our cells or body is simply the sum of its parts. If we were to study and understand each gene and each protein on

its own, we would eventually understand how entire cells work. But this is the same as trying to capture a symphony by listening to each instrument on its own rather than the whole orchestra. Now, biologists realise that we must understand how genes and proteins interact with each other to build the complex system that is a cell. Proteomics allows us to study many proteins at the same time to understand these interactions.

HOW IS MACHINE LEARNING USED IN THIS FIELD?

Proteomics produces huge amounts of data on the many thousands of different proteins in a single cell. To process and understand these data, we need to use computational techniques such as machine learning. A network of proteins in a cell is not unlike a social network, and many of the computational tools that we use are similar to the algorithms Facebook is using. For example, we try to understand which proteins form communities and what those

proteins have in common. We have also used machine learning to predict the function of unknown proteins. By teaching the computer how well-known proteins behave, it has learnt to identify which features of a new protein will likely correspond to which functions in the body.

WHAT IS THE FUTURE OF PROTEOMICS?

The next frontier for proteomics is the clinic. Over the past 10-20 years, we have seen how genomics (the large-scale study of DNA and genes) is transforming medicine and health care, and the potential of proteomics for medical research and diagnosis is even greater. I think proteomics has a great future and will be one of the key technologies for biomedical research and clinical diagnosis over the next decades. This is the most rewarding thing about working in this field.

EXPLORE A CAREER IN PROTEOMICS

- Many countries have proteomics societies, such as the British Society for Proteome Research (www.bspr.org) or the US Human Proteome Organization (www.hupo.org). Find the proteomics society in your country at www.hupo.org/National-and-Regional-Societies
- Many academics and research institutions run outreach activities on cell biology, which can be a good route into proteomics. For example, The Wellcome Centre for Cell Biology at the University of Edinburgh runs events such as lab visits for secondary school students (public-engage.bio.ed.ac.uk)
- After completing a PhD, a proteomics researcher can expect a starting salary of around £34,300 which will increase with experience.

PATHWAY FROM SCHOOL TO PROTEOMICS RESEARCHER

At school, study STEM subjects such as biology, chemistry, maths and computer science.

Proteomics can be entered through a wide range of undergraduate and postgraduate degrees, including biochemistry, molecular biology and bioinformatics. As Georg explains, "Proteomics is a very interdisciplinary research area. We need biologists, chemists, physicists, engineers, computer scientists and medical doctors."

Most proteomics projects struggle to find researchers with the required computational analysis skills. "So, whatever you do, make sure you learn some computer programming," advises Georg.

GEORG'S TOP TIPS

- 01 Follow the ideas that genuinely interest you rather than second-guessing what other people might like to read about.
- 02 Recognise opportunities when they arise and be prepared to take advantage of them.
- 03 Get some hands-on lab experience early on.



HOW DID GEORG BECOME A PROTEOMICS RESEARCHER?

I didn't really consider becoming a scientist until my last year in high school, even though I almost failed biology that year. I would have loved to become a musician (I spent most of my free time playing in a band) or a writer, but I did always like the idea of performing experiments.

My earliest experiment that I can remember was when I tried to create a rainbow paint! I ground up all my pencils and mixed them together in some water, stirring and cooking for days. Needless to say, that never worked! But then again, most experiments fail, at least the first time round.

I don't believe that anyone is born as a scientist, or as anything else for that matter.

If you enjoy doing something then you are going to be good at doing it, sooner or later. Many of the most successful scientists in the world have had very unusual career paths, and that probably goes for all walks of life.

I had been an active researcher for 15 years before I received the funding to start my own research group. I worked in several fields over those years, including cell biology, mass spectrometry and bioinformatics. Some people may consider that a lack of focus or poor career planning, but it has put me in a position to pursue a unique type of interdisciplinary research now.

Outside of work, I enjoy spending time with my family and being outdoors, playing

football or visiting the Scottish Highlands.

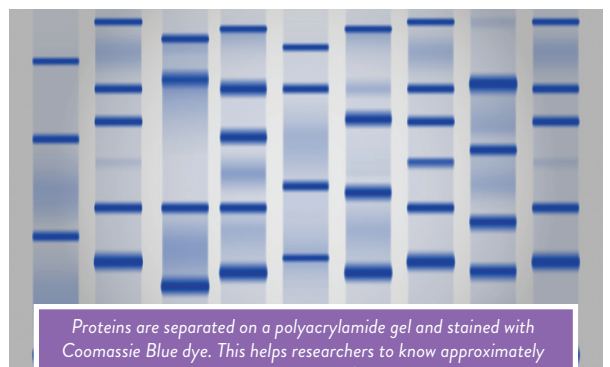
But fair warning: academic research is not really a 9 to 5 job, so you don't have too much free time as a scientist.

My biggest career achievement was obtaining a group leader position in Edinburgh.

Running an independent research group has always been a dream for me, but has at many times felt unachievable, especially when trying to balance work and being a single parent, which I was for 5 years or so. I'm also proud that some of my research insights were added to a new edition of a famous biology text book, so students all over the world will learn about them!



The result of a proteomics experiment where the concentration of proteins was measured in many mutant cells. A machine learning algorithm then grouped proteins together that were up- or down-regulated in similar ways across these mutants, suggesting that they are functionally related.



Proteins are separated on a polyacrylamide gel and stained with Coomassie Blue dye. This helps researchers to know approximately how many proteins are in a sample, before they inject it into the mass spectrometer.