Edexcel GCE

Biology

Advanced

Unit 5: Energy, Exercise and Coordination

January 2011

Scientific Article for use with Question 7

Paper Reference

6BI05/01

Do not return the Insert with the question paper.

Turn over ▶

P37491A

©2011 Edexcel Limited.

1/1/1/1/

P 3 7 4 9 1 A

edexcel

advancing learning, changing lives

Scientific article for use with Question 7

Generic master chefs' recipes for life

Imagine a bustling metropolis, sustained by myriad restaurants. At the heart of each is a busy kitchen with a table upon which lie volumes of a huge, ancient cookbook. Every restaurant relies on the same cookbook. Yet each uses a different selection of recipes – some churn out hamburgers exclusively, while others cook up exotic dishes such as moo-shoo pork and profiteroles. The various menus are created by mysterious behind-the-scenes chefs who specialise in particular cuisines.

Bizarre as it sounds, this kitchen scenario is a good analogy for what appears to be going on in the nuclei of living cells. The recipes are genes, the cookbook chromosomes and the chefs transcription factors – protein molecules that bind to DNA with the effect of switching genes on or off. Together, they hold the secret to one of the greatest puzzles in biology: how the virtually undifferentiated cells of the early embryo, all with the same genetic blueprints, develop into distinct tissues each dominated by the actions of distinct subsets of genes. How do liver cells, for example, manage to switch on genes for making detoxifying enzymes while leaving other genes dormant? How do skin cells 'know' that they need to draw on the genetic instructions for making fibrous keratin, but not, say, retinal photoreceptor proteins or the oxygen-carrying protein haemoglobin?

Over the past decade or so, biologists have been closing in on an answer, guided by an obvious strategy: get to know the chefs. As a result, transcription factors have become one of the hottest topics in molecular biology. Leading science journals vie for papers describing the latest advances in understanding how transcription factors latch on to DNA. Researchers compete to catalogue their molecular sequences, shapes and biochemical properties. International scientific conferences are convened to discuss the most recent results.

Now, with international efforts well under way to map and then sequence the human genome, biologists are anticipating an avalanche of new information. Much is already known about transcription factors from 'simple' organisms such as bacteria and yeast, while research into the proteins that switch genes on and off in the fruit fly and nematode worm – creatures with well-mapped molecular genetics – is progressing at a hectic pace. But with mammals, and in particular humans, it is still early days. Scientists don't yet know enough about the molecular genetics of mammals to embark on a full-scale assault on transcription factors. The human genome project could help to change that.

Yet even at this stage, the importance of transcription factors is clear. Without them, sun-drenched skin cells could not produce the protective pigment melanin, the hormone oestrogen could not trigger ovulation nor testosterone deepen the voices of teenage boys, and toxic chemicals such as dioxin could not wreak havoc on cells. As Robert Tijan of the University of California at Berkeley, a leading expert on transcription factors, puts it: 'Everything in biology comes down to how the genes are turned on and off.'

But there is much to learn. How many different transcription factors are there in the human body? Which ones influence which genes, and how? Do these proteins act alone or in teams? What part do they play in human diseases? Could scientists exploit their growing knowledge of transcription factors to combat diseases such as cancer and AIDS?

Cancer can result when genes involved in cell growth are turned on inappropriately. AIDS develops as the human immunodeficiency virus (HIV) invades cells, unleashes its own transcription factors, and commandeers gene expression. And heart disease is at least partly caused by deleterious genes that are expressed in the unlucky carriers through the normal action of their transcription factors. Many scientists hope to gain control of gene expression as a strategy for treating such diseases. Already, research is under way to develop antiviral agents through inactivation of viral transcription factors.

Gene control

Stephen Burley at the Howard Hughes Medical Institute and Rockefeller University in New York sees this approach as working hand-in-hand with gene therapy. This is the idea of inserting undamaged genes into cells to override the effects of damaged genetic material. Burley's vision is that one day, scientists will be able to use transcription factors to control the expression of these healthy replacement genes. 'Like having brakes and a gas pedal,' says Burley, 'you could rev up the transcription level or turn it off if it turns out to be harmful to the patient.'

But for the time being, most researchers are still looking at the basic science of transcription factors. One of the biggest surprises so far is how many transcription factors there are. 'Regulatory factors were supposed to be enormously scarce,' says molecular biologist Michael Levine of the University of California, San Diego, referring to conclusions drawn from older studies on bacteria. But it is now clear that at least 1000, and perhaps as many as 10 000, of the estimated 100 000 genes that make up the genome of a typical mammal may encode transcription factors. Moreover, says Levine, the transcription factors can occur in great concentrations, sometimes as many as 20 000 molecules in a cell.

Another surprising discovery is that different transcription factors can collaborate to produce a particular effect on a gene, much as groups of chefs with various specialities can cooperate to create many different menus. The number of possible combinations of transcription factors in mammals could be vast. This may explain why our genes can be 'interpreted' by cells in so many different ways. Molecular biologists are now going hell for leather to find out how various transcription factors combine in cells to turn on, or 'express', different sets of genes.

Since the 1960s it has been clear that genes become expressed through a two-step mechanism. First, the DNA code of the gene is copied, or transcribed, by the synthesis of a strand of messenger RNA which repeats the code found on the gene. This process requires the help of an enzyme called RNA polymerase, which must first bind to a region of the gene called the 'promoter' before passing along the DNA, assembling the messenger RNA. Next, enzymes in the cell use these messenger RNA molecules as templates for building proteins, each of whose chemical structures is specified by the original DNA codes. But, this mechanism doesn't explain how each cell in our body manages to select and express only certain genes.

At first researchers postulated that transcription was regulated by special protein subunits of RNA polymerase, dubbed 'sigma factors', which had to join up with the main part of the RNA polymerase, before transcription could occur. It was in looking for these sigma factors that the true answer began to emerge. In the early 1980s, Tijan discovered a protein called SP1 in human cells, and Rockefeller's Robert Roeder found a protein called TFIIIA in frogs' eggs. Both these proteins helped RNA polymerase but they could also activate specific genes. Now there was a plausible model for gene-specific activation.

Since then, researchers have identified several hundred transcription factors in yeast cells and mammals, including humans, that function similarly. The defining characteristic of these proteins is that they all have structures called 'DNA-binding domains'. These allow proteins to 'recognise' a particular stretch of DNA and dock into the groove between its two nucleotide strands like a key in a lock. Scientists have made great strides in uncovering the shapes of these molecules using techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy. They fall into families with names that reflect the shape of the protein, such as leucine zippers, helix-loop-helix proteins and zinc fingers. Each type of protein locks onto DNA grooves with its own distinctive type of molecular key.

The leucine zipper proteins have generated much excitement because they shed light on how transcription factors can collaborate to produce a range of different effects on genes. The best studied are proteins called *fos* and *jun*, which stimulate cells to divide and have been implicated in the chain of events leading to cancer. Each of these molecules functions like one half of a zip – the teeth being a string of leucine amino acids – which can be completed in three different ways: with one molecule of fos linked to another fos, one jun molecule linked to another jun, or a fos linked to a jun. These three pairings produce leucine zippers that bind to DNA in three different ways to produce a variety of responses.

As the research progresses, new families of transcription factors with quite different DNA-binding domains are sure to emerge. But the story is unlikely to end there. 'For years we've thought that if we understood the DNA-binding sequence, that would tell us about transcription regulation,' says Tijan. Now he and others are coming to appreciate that this is not enough. An equally important component of these molecules is a part that enables them to interact with each other, a region called the 'activation domain'.

Activators and repressors

The shapes and chemical properties of these domains are not yet known, but they have the all-important job of deciding which transcription factors can fit together, like pieces in a jigsaw, to form a molecular complex. Evidence is fast accumulating that it is the overall shape and size of the complex, not those of its protein pieces, that decides which gene it switches on. Indeed, biologists now have a general outline of what must happen before a gene can be switched on in a cell. First, the cell must produce each of the required transcription factors. Then these proteins must slot together to form a complex. Finally, this complex must have just the right shape to plug into an intricate array of grooves on the DNA. Only after that can an RNA polymerase enzyme begin the copying of DNA into RNA.

The detailed picture is a little more complicated. For a start, some transcription factors act in a negative sense, to prevent molecular complexes from activating their target genes. 'Repressor' proteins first came to light two decades ago, when scientists at Harvard University discovered a protein in phages that could block gene expression in infected bacteria. They act by binding to DNA and preventing RNA polymerase from moving along it. Now it is clear that repressor proteins can work in other ways, by blocking construction of the DNA-binding complex or interfering with its binding to DNA. 'Activators and repressors probably duke it out with one another,' says Levine, whose work seems to confirm this view.

Levine has spent nine years studying how genes are turned on and off during the development of fruit fly embryos. He noticed that variations in the levels of a transcription factor called 'dorsal' act to shape the development of muscles, internal organs, skin and nervous tissues. In other words, different levels of this transcription factor seem to switch on different sets of genes, some characteristic of muscles, others of skin and so on. Levine puzzled over this observation until further research produced the key: dorsal's versatility was all down to the fact that it could react with two other transcription factors. Various combinations of the three transcription factors were coming together at three binding sites on DNA. Whether dorsal switched nearby genes on or off depended critically on the spacing between the sites and the presence of the other two transcription factors. 'In certain promoter regions, dorsal finds itself next-door to bad company, co-repressors, which block its ability to act as an activator and convert it into a potent repressor,' says Levine.

Research by other biologists has shown that this is just one of the ways transcription factors can flip, Jekyll-and-Hyde like, between conflicting roles. Some transcription factors change character in response to signals from hormones. In effect, they behave as receptors for many of the body's hormones, such as thyroid. In test tube and cell culture studies, for instance, the thyroid receptor keeps certain growth-related genes dormant. But when thyroid hormone comes on the scene it binds to the receptor, binding it into a gene activator.

Some proteins perform the duties of a transcription factor on a strictly part-time basis. For example, the usual job of NF-k B, a protein important in inflammation and immunity, is to sit tight in the cell fluid, tethered to a second protein. But for reasons not yet clear, the bond between these two proteins is occasionally broken, allowing NF-k B to enter the cell nucleus and act as a transcription factor for certain genes.

Working together

Researchers also suspect that the way genes are packaged influences transcription. In cell nuclei, DNA is not only double-stranded and helical, it is also wrapped around large, globular proteins called nucleosomes. Numerous other proteins also help stabilise the DNA and build up the chromosome. Biologists are just beginning to look at whether interactions with these proteins could make some genes more accessible than others to transcription. Evidence already exists that the tagging of a gene with methyl chemical groups can shield it against transcription and result in 'genomic imprinting'.

Meanwhile, researchers continue to find new proteins involved in transcription. In all, as many as 60 proteins may be working together to turn on a single gene. Besides the gene-specific transcription factors, there are proteins that appear to foster chemical communication between the complex of transcription factors, RNA polymerase and DNA. Still other proteins may help direct the construction of the complex. Putting the whole story together will be difficult.

Fortunately, we do not need to understand the entire mechanism before medical applications can be explored, says Steven McKnight, discoverer of the leucine zipper and now research director of Tularik, in San Francisco. He is hoping that the use or misuse of one or several key transcription factors will be at the root of most diseases. McKnight contends that each transcription factor may use different mechanisms to turn some genes on and others off, but that all of its target genes will code for proteins involved in a single biological response, such as inflammation or cell growth. That means it should be possible to treat some diseases by interfering with a single transcription factor.

The challenge now is to figure out which transcription factors have the greatest impact on diseases, such as cancer, AIDS and arthritis. 'We're just beginning to know enough to say this factor controls this process, and we only know in a few isolated examples. But over the next 10 or 15 years, people will sort out the circuitry that controls gene expression in humans,' McKnight predicts. 'It's a lot of work, but in the end, we'll have a better chance at picking our targets for controlling gene expression.'

Along the way, other intriguing intellectual questions may well be answered. For example, have transcription factors become increasingly important with the evolution of more complex creatures? 'It will be interesting,' says Levine, 'to see if the percentage of transcription factors in higher organisms is more than in lower organisms – that is, whether the computer is the same, but the software gets more complicated.'

Already, developmental biologists have found that organisms as diverse as corn, yeast and humans contain very similar stretches of DNA called homeoboxes. Discovered in the 1980s in genes that control construction of the basic body plan of fruit flies, homeoboxes consist of sequences of 180 base-pairs that code for DNA-binding domains. Researchers now know that the genes that contain them encode a very large and important class of transcription factors which coordinate the activity of arrays of other nearby genes. Indeed, these so-called homeodomain transcription factors regulate clusters of genes that are arranged along the chromosome in a sequence that corresponds to the order – from head to tail – of the body parts whose development they control.

Insights into human biology and medicine are now beginning to emerge. Some genetic diseases – certain types of leukaemia and syndromes that affect nervous system development – have been linked to defects in homeodomain transcription factors. Much more shall be learnt from studies of laboratory mice, which share with humans a similar pattern of four rows of homeobox-containing genes, each one on a separate chromosome. 'In a general sense, what we find in mice, we'll find in humans,' says Matthew Scott, a developmental biologist at the Howard Hughes Medical Institute and Stanford University in California and a discoverer of homeoboxes. 'But at what level will we start to see differences?' That is not yet clear.

For those interested in discerning common patterns within nature, as well as appreciating the beauty of individuality, the study of gene expression provides rich rewards. 'Since every gene is different, you're looking at a new animal every time you look at a new gene,' explains Albert Baldwin, from the University of North Carolina at Chapel Hill. 'That's what makes transcription so exciting – it's like snowflakes or human beings – every gene will be regulated somewhat uniquely.

How genes evolve

ONCE, we could only marvel at the wonder of life. Like movie audiences not so long ago, we had little idea of what went on behind the scenes.

How times have changed. As the genomes of more and more species are sequenced, geneticists are piecing together an extraordinarily detailed "Making of..." documentary. Nowadays, we can not only trace how the bodies of animals have evolved, we can even identify the genetic mutations behind these changes.

Most intriguing of all, we can now see how genes – which are the recipes for making proteins, the building blocks of life – arise in the first place. And the story is not unfolding quite as expected.

The most obvious way for a new gene to evolve is through the gradual accumulation of small, beneficial mutations. Less obvious is how an existing gene that already does something important can evolve into a different gene. The scope for such a gene to change tack without capsizing the organism that carries it is very limited. However, as biologists realised a century ago, this constraint no longer applies when mutations produce an entire extra copy of a gene.

Trillions of copies

According to the textbooks, the process by which new genes form starts with gene duplication. In the vast majority of cases one of the copies will acquire harmful mutations and will be lost. Just occasionally, though, a mutation will allow a duplicate gene to do something novel. This copy will become specialised for its new role, while the original gene carries on performing the same task as before.

Surprisingly, gene duplication has turned out to be nearly as common as mutations that change a single "letter" of DNA code. During the exchange of material between chromosomes prior to sexual reproduction, mistakes can create extra copies of long DNA sequences containing anything from one gene to hundreds. Entire chromosomes can be duplicated, as happens in Down's syndrome, and sometimes even entire genomes.

Since duplication can throw up trillions of copies for evolution to work with, it is not surprising that over hundreds of millions of years, a single original gene can give rise to many hundreds of new ones. We humans have around 400 genes for smell receptors alone, all of which derive from just two in a fish that lived around 450 million years ago.

Not the whole story

This classical view of gene evolution is far from the whole story, however. A decade ago, Michael Lynch at Indiana University in Bloomington and a colleague outlined an alternative scenario. Genes often have more than one function, and Lynch considered what might happen after such a gene is duplicated. If a mutation knocks out one of the two functions in one of the copies, an organism can cope fine because the other copy is still intact. Even if another mutation in this other copy knocks out the second function, the organism can carry on as normal. Instead of having one gene with two functions, the organism will now have two genes with one function each – a mechanism Lynch dubbed subfunctionalism. This process can provide the raw material for further evolution. "A gene preserved by subfunctionalisation can later pick up a new function," Lynch says.

Some theoretical biologists think gene copies can also be preserved by other, more subtle, mechanisms, but the real challenge to the classical model comes from actual studies of new genes in various organisms. Earlier this year, in the most comprehensive study of its kind yet, a team led by Wen Wang of Kunming Institute of Zoology in Yunnan, China, looked at several closely related species of fruit fly. By comparing their genomes, Wang was able to identify new genes that have evolved in the 13 million years or so since these species split from a common ancestor.

One of Wang's surprise discoveries was that around 10 per cent of the new genes had arisen through a process called retroposition. This occurs when messenger RNA copies of genes – the blueprints sent to a cell's protein-making factories are turned back into DNA that is then inserted somewhere else in the genome. Many viruses and genetic parasites copy themselves through retroposition, and the enzymes they produce sometimes accidentally retropose the RNA of their host cells.

Dead on arrival?

The gene copies created by retroposition are not the same as the original, as genes consist of more than just the sequence coding for a protein. There are also "promoter" regions in the front of the coding part, to which other proteins bind, and this determines when and in which tissues the gene is turned on. Since retroposed gene copies lose their promoters, which are not transcribed into RNA, it used to be assumed that these partial copies were never expressed and gradually disappeared as mutations accumulated. Retroposed gene copies were dismissed as "dead on arrival", says Henrik Kaessmann of the University of Lausanne, Switzerland.

However, it has become apparent that a retroposed copy can sometimes get inserted in the genome near an existing promoter, making it active. Crucially, though, with a different promoter, it will be turned on at different times or in different tissues or both. In this way a retroposed gene can immediately acquire a new function.

This process may have created many of the recently evolved genes in us apes. A burst of retroposition in our ancestors, peaking around 45 million years ago, gave rise to many thousands of gene duplicates, of which at least 60 or 70 evolved into new genes, according to a 2005 study led by Kaessmann. The burst was probably due to a new genetic parasite invading our genome.

Brainy genes

Kaessmann's team is now studying some of these genes in more detail. Their work suggests that at least two, called *CDC14Bretro* and *GLUD2*, could be related to apes' increased cognitive abilities.

The evolution of new genes often involves even more drastic changes. In his fruit fly survey, Wang found that a third of new genes were significantly different from their parent genes, having lost parts of their sequences or acquired new stretches of DNA.

Where do these extra sequences come from? In complex cells, the DNA coding for a protein is broken into several parts, separated by non-coding sequences. After an RNA copy of the entire gene is made, the non-coding bits – the introns – are cut out and the coding parts – called exons – are spliced together. This edited RNA copy is then sent to a cell's protein-making factory. The modular form of genes greatly increases the chances of mutations reshuffling existing genes and generating novel proteins. There are all sorts of ways in which it can happen: exons within a gene can be lost, duplicated or even combined with exons from different genes to create a new, chimeric gene.

Variations on a theme

For instance, most monkeys produce a protein called TRIM5, which protects them from infection by retroviruses. In one macaque in Asia around 10 million years ago, an inactive copy of a gene called *CypA*, produced by retroposition, was inserted near the *TRIM5* gene. Further mutation resulted in cells producing a chimeric protein that was part TRIM5, part CypA. This protein provides better protection against some viruses. Although it might seem an unlikely series of events, in fact the *TRIM5-CypA* gene has evolved not once but twice – much the same thing happened in owl monkeys in South America.

Given enough time – or rather enough mutations – gene duplication and reshuffling can produce new genes that are very different from the ancestral ones. But are all new genes variations on a theme, or can evolution throw up new genes unlike any that already exist?

A couple of decades ago, it was suggested that unique genes could arise from what is called a frameshift mutation. Each amino acid in a protein is specified by three DNA "letters", or bases – the triplet codon. If a mutation shifts the starting point for reading codons – the "reading frame" – by one base, or by two, the resulting protein sequence will be completely different. Since DNA is double-stranded, any given piece can be "read" in six different ways.

Gibberish

The vast majority of mutations that alter the reading frame of a gene produce nonsense, usually dangerous nonsense. Many genetic diseases are the result of frameshift mutations wrecking proteins. It's a bit like swapping every letter for the next one along in the alphabet: the result is usually gibberish.

But not always. In 2006, Stephen Scherer of the University of Toronto in Canada and his colleagues searched the human genome for new genes that had evolved by duplication followed by frameshift mutations affecting at least part of the original gene. They found 470 examples, suggesting that the process is surprisingly common.

Another source of unique new genes could be the "junk" DNA littering most genomes. An early hint this might be so came a decade ago when a team at the University of Illinois revealed the genesis of the antifreeze protein produced by one Antarctic fish. The gene involved originally coded for a digestive enzyme. Then, around 10 million years ago, as the world's climate cooled, part of one of the introns – a piece of junk DNA, in other words – got turned into an exon and subsequently duplicated many times, generating the characteristic repetitive structure of antifreeze proteins. From a random bit of DNA evolved a gene vital to the fish's survival.

From scratch

Still, the antifreeze gene evolved from a pre-existing gene. What are the chances of mutations in junk DNA generating an entire new gene from scratch? Practically zero, most biologists thought until very recently. As Lynch points out, it takes a whole set of unlikely conditions for a piece of random DNA to evolve into a gene. First, some of the DNA must act as a promoter, telling the cell to make RNA copies of the rest. Next, these RNA copies must have a sequence that can be edited into a viable messenger RNA blueprint for the protein-making factories. What's more, this messenger RNA must encode a relatively long protein – the average length is 300 amino acids – which is unlikely because in a random stretch of DNA, on average 1 in 20 every codons will be a "stop" codon. Finally, of course, the new protein must do something useful. The obstacles seemed insurmountable.

Then, in 2006, David Begun of the University of California and his colleagues identified several new genes in fruit flies with sequences unlike any of the older genes. They suggested that these genes, which code for relatively small proteins, have evolved from junk DNA in the past few million years. Begun quotes Sherlock Holmes: "When you have eliminated the impossible, whatever remains, however improbable, must be the truth."

Junk DNA

This year, during his hunt for new genes in fruit flies, Wang found another nine genes that appear to have evolved from scratch out of junk DNA. For eight of the nine, Wang has identified the non-coding sequences from which the genes evolved in related species, ruling out the possibility that these genes were somehow acquired ready-made from another organism.

Altogether, an astonishing 12 per cent of recently evolved genes in fruit flies appear to have evolved from scratch. And Wang suspects this rate is low compared with other animals. "My gut feeling is it may be higher in vertebrates because they have more junk DNA," he says.

It looks as if Wang could be right. A team at Trinity College Dublin in Ireland has found evidence that at least six new human genes have arisen from non-coding DNA in the 6 million years or so since humans and chimps diverged. The work is ongoing but the preliminary findings were presented at a meeting in Barcelona, Spain, in June. "We're very excited," says team leader Aoife McLysaght.

First hurdle

How can the number be so high when the likelihood of a gene evolving in this way is so rare? Part of the answer could be the recent discovery that even though at least half of our genome is junk, as much as 90 per cent of it can be accidentally transcribed into RNA on occasion. "The first hurdle has already been overcome," McLysaght says.

This means it might not be that uncommon for random bits of junk DNA to get translated into a protein. Since most random proteins will probably be harmful, natural selection will eliminate these DNA sequences, but just occasionally one will strike it lucky, Begun says. A sequence that does something beneficial will spread through a population and rapidly evolve into a new gene, becoming optimised for whatever role it plays.

It will be many years yet before we have a clear picture of the relative importance of the various mechanisms by which genes can evolve. What is certain, though, is that the classical view of how they evolve is far from complete. Evolution isn't fussy – it'll take new genes wherever it can get them. "Natural selection is aggressively opportunistic," says Begun. "The source of the raw material is irrelevant."

And as sequence data continues to pour in, biologists are well on their way to working out how every one of our 20,000 or so genes evolved. Better stock up on the popcorn and make sure your sofa is comfy: this is going to be one epic "Making of..." documentary.

Acknowledgements

Meet genetics master chefs: All cells contain the same genetic issue New Scientist Issue 1922 © Copyright Reed Business Information Ltd.
Recipes for life: How genes evolve, Michael Le Page, New Scientist Issue 2683 © Copyright Reed Business Information Ltd.

BLANK PAGE

BLANK PAGE